DESIDOC Journal of Library & Information Technology, Vol. 32, No. 3, May 2012, pp. 277-284

Establishment of Institutional Mechanism for Building National Repository in Health Sciences

Surinder Kumar

National Informatics Centre, A-Block, CGO Complex, New Delhi-110 003
E-mail:suri@nic.in

ABSTRACT

National consolidation of published and unpublished literature in the field of biomedical sciences can play a major role in scholarly communication to help the end users in providing research published in the country. Institutional repositories are a good approach for a cost-effective publishing with a cooperation and participation of each institution for capturing, preserving, managing, and nurturing the discussion. In turn, metadata can be harvested centrally to access the digital information of common interest whereas individual libraries should able to preserve digital assets. Institutionalisation mode has been recommended for building national digital repository system for the country. The public funding should be provided to apex body so as to formulate the requisite policies for the spread of open access movement in the country and also formulate a long term sustainable model for building national level system in the country.

Keywords: Institutional repositories, health sciences, intellectual property rights

1. INTRODUCTION

There has been scholarly communication, for centuries, collecting and providing access to physical copies of texts. However, in the first decade of 21st century it has been moved to digital access to texts and various other media. The changing economics of purchasing, serving, and storing has resulted in a complex landscape, where secondary storage facilities are needed to accommodate the growing volume of digital information[1]. The national consolidation of published and un-published literature, particularity in biomedical field is very important to sharing the cases studies, case conference, research reports submitted to funding agencies and a typical discussion to help the socio-biological conditions of respective countries including the teaching, research, and scholarship. Institutional repositories (IRs) are a good approach for a cost-effective publishing with a cooperation and participation of each institution for capturing, preserving, managing, and nurturing the discussion. In turn, it can be used centrally to access the digital information of common interest whereas individual libraries should able to preserve digital assets. It requires efforts from the various stake holders such as faculty, administration, scholarly societies and publishers to address the common problems. Thus, the economically sustainable development, and maintenance of high-quality and readily accessible research collections should become possible.

2. CURRENT SCENARIO OF INSTITUTIONAL REPOSITORIES

In current scenario of IRs and National service providers in the country, there is a need to look at more robust and sustainable model for providing seamless access to large amount information available in 62 IRs in the country. There are number of initiatives in development of building National level system such as Cross Archive Search Services of Indian Repositories (CASSIR) (http://casin.ncsi.ernet.in) of Indian Institute of Science (IISc) is one of the largest service providers harvesting data from 22 IRs in the country. However, the development of CASSIR is project-based not serving the national requirement as it does not have any defined policy for harvesting data from the repositories. There are other service providers which are catering the need of subject-based harvester and yet to reach the critical mass in terms of articles harvested. The present models except Open J Gate (http://www.jgate.com), which is running on a purely advertisement and company's and license fee model, and OSSD repository is a collaborative and institutional model in the development of Mycobacterium Tuberculosis (MBT) repository, evolved as a national level

Received 17 October 2011, revised 09 December 2011, accepted on 02 January 2012, online published 18 May 2012

system from public-funded system. The public-funded model is suitable for India, which would slowly become self-sustainable and become a robust and sustainable model for providing seamless access to large amount information on particular focused topic (like information bank) for information available in health sciences institutes in the country.

The present study has evolved a model with an institutionalisation called national digital repository system (NDRS) for intellectual capital through cooperative collection, storage and management initiatives with public fund at the initial phase and later as participatory model that can shape and support national consolidation in the country.

3. INSTITUTIONALISING OF NDRS

Institutionalisation requires in driving the organisational and financial infrastructure that supports and sustains its present activities and operations. The organisational, governance, and financial systems have some traits in common, but depend on the relationships between the IRs and NDRS[2]. It should have the following functions:

- It should be developed and run by an independent advisory or governing board to be managed by professional bodies. It would oversee general policy matters, such as the kinds of materials accepted, allocations of infrastructure as a facility, and the apportionment of individual IRs as share of operating costs. Such board also guides general investment and budgetary strategy and future development of the national level repository.
- It should also look after the management operation such as scheduling, workflow, logistics, and production of NDRS.
- It should be funded by national apex bodies for the initial establishment, then managed and operated by a single institution/university so as to provide services to end users on a fee or no fee basis. The financial arrangements underlying creation and maintenance of the individual repositories vary depending upon the financial strength of the institutes.
- It should follow publishing trend particularly in the open access initiatives to achieve greater efficiencies in storage, service and preservation.

It is proposed to maintain cooperatively managed repositories as a central facility backed by institution repositories to serve the community at large as archival repositories, and it would ensure preservation as the key to safeguarding the intellectual and cultural heritage of the country. The NDRS can adopt any one of the following models[3]:

3.1 Consortium Repository Model

The consortium model (or shared model) is one in which a group of institutions, generally without individual pre-existing repositories, come together to develop a single instance of a repository to house content from all participating institutions. The example used for this model is Washington Research Library Consortium (WLRC) (http://www.wrlc.org/ in the USA with its online repository 'ALADIN Research Commons (ALADINRC)' and White Rose Consortium in UK with its online repository 'White Rose Research Online (WRRO). It provides access to each institution's scholarly output to all member institutions. It is a voluntary where anyone can join, based on the policy of the consortia. In consortia model, usually successful, governance is through an independent organisation (not-for-profit organisations) can be formed by including the participating IRs for policy and programmatic direction from the consortia's governing. Representation of the institution administration level is also likely to shape the repositories' programmes around the larger agendas of the participating institutions. The operations of consortia repositories are to be placed under one of the consortium member libraries. In arrangements where one consortium plays a greater role in operation of the repository than others do, some might fear that the interests of that institution would prevail over those of the others or of the consortium at large. In practice, however, strong, formalised governance and financial policies and procedures can level the playing field. Administrative mechanisms require that policy and programme decisions be unanimous, strengthening this assurance[1].

3.2 Platform Model

Brown[3] defined platform model as one in which a group of institutions, again usually without pre-existing repositories, collaboratively develop a customised, customisable version of a software platform which is then used as the basis for repositories at each participant institution or hosted centrally. The example used in this model is Shodhganga (http://shodhganga.inflibnet.ac.in) developed by INFLIBNET for e-thesis which is hosted centrally. The funding is providing by UGC to each participating university and software used is DSpace. It has kept the provision of harvesting data from the associative OAI-PMH compliant repositories. It provides a single point access to all the resources of all the participating organisations. It has minimum development cost at the part of organisations. However, it requires greater coordination among partners.

3.3 Proprietary Repository Model

The proprietary repository model (PRM) depends on one institution (host), where it takes full responsibility of

establishment of facility and some of the heads of the participating institution will be part of the governance and operate independently by the host organisation. Host institution provides the fund as part of the central institutional budget. The facility is part of the institutions effort to coordinate and traditionally act independently. For the benefit of having good IRs, host can have coordinating committee for services and functions selecting representative from all of the participating IRs. For example, Open Repository is a hosted solution from BioMed Central that builds and maintains customised DSpace repositories on behalf of institutions and organisations. This allows an institution to preserve and showcase its intellectual output to a larger and unrestricted audience[1].

Capital and operating fund, in proprietary model, can be supplemented by the fee derived from various repository-related activities, as a set annual payment, based on the amount of material stored and transaction-based fees for collection-related services. The IR may store materials at the repository on cost-recovery basis. Such materials include institute records, artwork, special materials, scholarly publication, and other learning/ course materials. Direct costs are recovered through charges for accessioning, retrieval storage and administration[3].

The proprietary model usually provides service-based fee-for-service basis for its storage and affiliated services are more akin to a landlord-tenant relationship than like relationship model or between repository and consortium members. The depositing organisations are normally not eligible to participate in governance of the facilities, and they do not have any significant investment in the welfare of the repository. Under such arrangements the 'tenants' are not likely to be strongly motivated to support necessary capital improvements or other measures that advance the facility's broader, long-term goals.

3.4 Aggregation Model or Participatory Institution Model

In aggregation model or participatory institution model, the institutions come together to create a search aggregator to harvest content from their repositories and present it via a single search interface to maximise the impact of the harvesting system. It is mandatory for the member institutions to have their repositories to be part of a harvesting system. The CARL metadata harvester and search engine[4] and the Dutch Digital Academic Repositories (DARE) (http://www.darenet.org) are the example of PIM. The PIM model will have its National Advisory Board (NAB) for policy and programmatic direction, which comprises representation from participatory IRs, experts in the subject, technologists, finance representative from the funding apex body and NDRS administrator as Member-secretary. The NDRS board will have input from the institutional level committee (ILC) to enable cooperative-management, collection-sharing initiatives and ensure the collective interest of the IR and to maintain governance and budgetary control. The operation of the IRs is delegated to individual institutions. The NDRS exercises a more direct measure of control over the facilities than it does over the individual IR. An apex body keeps direct accountability because the facilities represent significant capital investments that benefit more than one IRs within the system. At the IR level, there will be an administrative body, within the organisation hierarchy, ILC, that lies beneath the management of the institution and individual IR level. The purpose of the committee is to bring the interests of the individual participating IRs and other stakeholders to bear on shaping the programmes of the repositories. The ILC is composed of the IR administrators, a representative of the academic activities of the institutions and a representative from the NDRS for better co-ordination, where ILC develops policy and programmes[2]. Funds for capital expenses are from the apex bodies, and most of the operating funding come from the participating respective institutional repositories of the respective subject groups, for example (medical subject) funding is from Indian Council of Medical Research (ICMR), Director-General of Health Service (DGHS), where National Informatics Centre (NIC) serves as technology partner. As the intellectual materials are considered as public good, it is suggested that the total project should be fully-funded by the Government of India, and at institutional level to be supported by respective institutions wherever mass repositories are shifted to central facilities through their respective institutions.

4. SUITABLE INSTITUTIONALISATION MODEL FOR NDRS

The present study recommends PIM, which will have its own participatory policy and programmatic direction, as it provides opportunity for representation from participatory IRs, scholarly contributors in the subject, technologists, finance (representative from the funding apex body, and NDRS administrator as Member-secretary. It would enable cooperative-management, and collection-sharing initiatives, ensure the collective interest of the IR, and maintain governance and budgetary control over them. The operation of the IRs is delegated to individual institutions. The NDRS exercises a more direct measure of control over the facilities than over the individual IRs. An apex body (as an initial funding agency) keeps direct accountability as the facilities represent significant capital investments that benefit more than one IRs within the system. It also has control and policy input from the IRs and involve a representative of the academic activities of the institutions for better co-ordination. Funds for capital expenses are from the apex bodies, and most of the operating funding come from the participating IRs of

the respective subject groups. Also, it will be part of technology partners with the respective institutions.

4.1 Objectives of Proposed NDRS

The proposed NDRS should have the following objectives:

- It should keep track of the current demands, and publishing trends (particularly the OAI) to achieve greater efficiencies of storage, service, and preservation. It is proposed with intention to maintain cooperatively managed repositories, as a central facility, backed by IRs to serve the community at large as archival repositories, it would ensure preservation as the key to safeguarding the intellectual and cultural heritage of the country, that may be of value for scholarship.
- It may serve as secondary storage, as an absolute need, in keeping the requirement of digital space and expertise, where it can optimise for preservation. It should also have defined the kind of materials that should go into these repositories; how they are to be organised, governed, and sustained; and how collaborations among IRs, between IRs and scholars should be nurtured. In fact, the scholars are looking at repositories to serve as an aggregated materials based on subject or discipline, simultaneously and over time, creating repositories of archival collections that collect on behalf of an entire nation.
- It should hold new collaborative solutions, as web-based digital library collaboration, for new forms of cooperation. Nonetheless, it has to build sustainable repository collections to fulfill the crucial societal roles or heritage collections. The study proposed that there should be an agreement on core attributes of repositories, protocols for governing, management and preservation, and how to make such information widely accessible. It should also serve as an archival back-up collection that will be used for disaster recovery.
- It is about creating a network of IRs, not like libraries that ends in general agreement and backed away from depositing copies. It should require a 'strong inter-institutional culture' on a long-term commitment that can achieve scales of economy and improved stewardship. Repositories should go beyond the mere sharing of storage space to the sharing of management and access. It should be achieved through building trust and transparency among member repositories and in providing its operation, access and preservation[5].
- It is largely the products of inter-institutional efforts to accommodate digital-based materials and cost-effective solutions to collections storage; pool resources and offer a shared space; subjected to common standards and managed by a single organisation. It should support a certain degree of interdependence and cooperation among the participating repositories with respect to the preservation at system level and at aggregated holdings and shared corpus of research materials. It should work by systematising and coordinating responsibilities seek rationalisation and achieve economies by aggregating from among the participating institutions individual repository[6].
- It should also explore the extent to which the repositories represent an emerging architecture, whereby the participating IRs might move beyond serving their communities and participate in a national network for cooperative preservation.

4.2 Policies for Selection of Materials

It is the institutions' policy (at IR level) to decide the selection of certain categories of materials and should adopt the following guidelines[7]:

- Repositories also provide an economical and practical means of storing problematic bodies of material, such as materials having access restrictions and have collection-level control.
- The NDRS actively manage the intake of materials to achieve goals that go beyond merely providing place storage towards proper access providing expertise.
- Rationalisation will be undertaken by involving coordinated collection responsibilities, negotiating collectively electronic-journal licensing, and assembling shared collections of record.

4.3 Withdrawal Items from NDRS

The NDRS to consider relatively stable and encourage the idea that materials moved to the facilities are intended for permanent storage, with the principle that minimum maintenance and traffic promote cost-effective operations. Policies governing the removal of materials from storage by IRs may vary, some repositories maintain 'one-way door' policies; however, under certain circumstances, permit materials to be removed and reintegrated. The NDRS keep consideration to change on par with the technological changes whenever it needs to reconsider policies for adapting technologies

4.4 Implications for Collective Management

The NDRS has to realise economies in the care and administration as common facilities, subjecting the collections to many of the same procedures and conditions of service, so as to achieve a high degree of coordination among IRs. The responsibilities lies in managing the collections as a single entity with respect

to access and control, even though IRs may retain ownership of them.

4.5 Collection Ownership

There are important differences between common management of collections and shared ownership. For example, IRs retain the right to withdraw their materials from the repository. Ownership of collectively managed materials nonetheless continues to be a volatile issue, but within the context of the repositories, the practical distinction between shared management of a body of materials and actual ownership can become difficult to make.

4.6 Intellectual Property Rights and NDRS

Long term preservation helps authors more confidence in the future accessibility and more incentives to deposit the content and is one the challenge for IRs, keeping in view of changing hardware and software/file format. It is difficult to determine the cost of preservation. Hence the NDRS expects institution repositories that it must have policy in terms of content management, rights management, long term preservation policy, etc.

The access to materials on the IRs is cost free at the users level and hence the understanding intellectual property is important for creating and disseminating content on NDRS. As OA repositories of scholarly resources are available to all free at the point of use therefore intellectual property issues are to be carefully followed so as to provide the service for a longer period of time. Digital preservation requires new workflow, new skills, and close cooperation across different professions ranging from traditional preservation management to computing science.

Joint Information System Committee (JISC) (http://www.jisc.ac.uk), RLG-OCLC (http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf) and OAIS model provides framework to implement preservation strategy and IRs software a mean to implement this preservation strategy. It is advisable to follow the standard guidelines developed from these organisations.

4.7 Architecture of Proposed NDRS Model

The architecture of NDRS consists of three layered architecture comprised of the data providers (content layer), service providers (aggregation layer) and interface layers. The content layer is a distributed institutional repository which is OAI-PMH compliant exposed metadata to be used in central metadata repository in aggregation layer[8]. It has been depicted in Fig.1.

The aggregation layer has main components of OAI-PMH harvester and the central metadata repository. The OAI-PMH harvester harvests and aggregates metadata records from OAI-PMH data providers. Harvested records are aggregated and stored in the metadata repository which may serve as a base on which various services can be developed subsequently. The interface layer provides

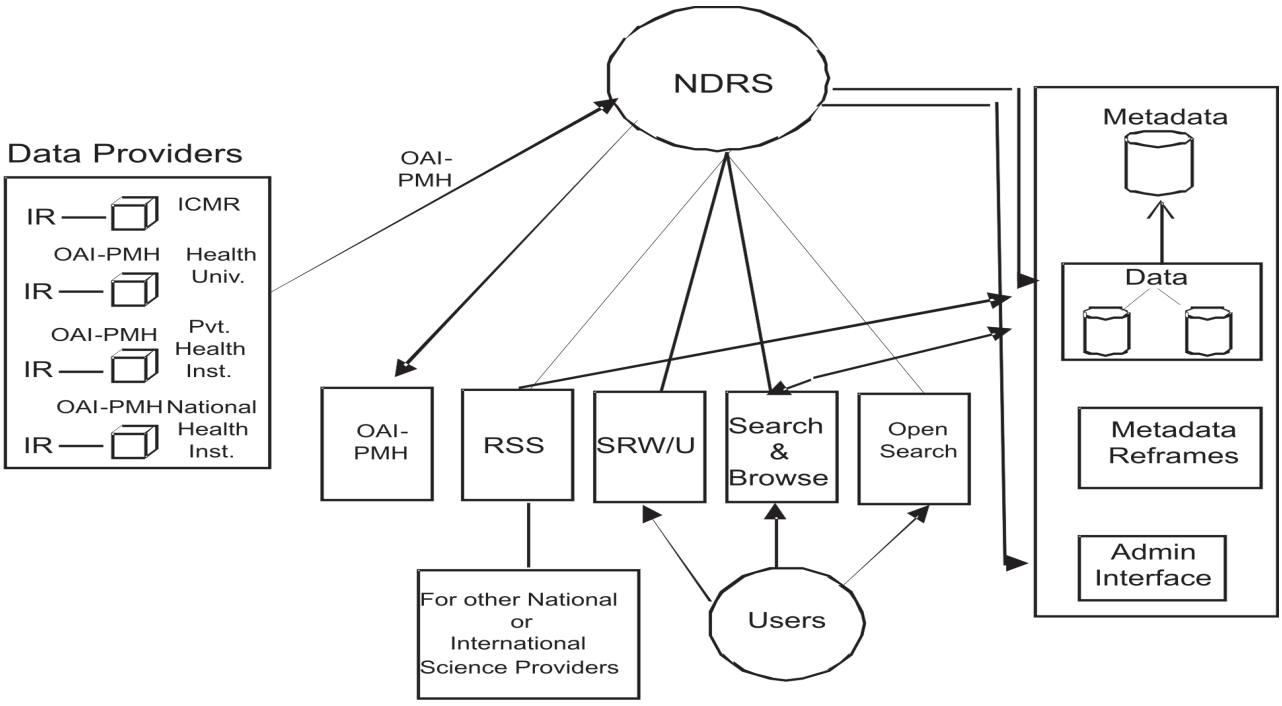

Figure 1. Proposed architecture of NDRS.

both user-oriented services as well application-oriented services. User services include conventional services such as browse and search services. Application-oriented services are based on interoperability interfaces which support machine-to-machine or application-to-application data exchange or service interactions. The RSS service interface will publish newly added records as RSS feeds to support RSS service; OAI-PMH interface is used to re-exposing harvested metadata records is used to re-exposing harvested metadata records to enable data sharing to other service providers globally. SRW/U interface is a standard API search engine used as an interface to repositories[9]. Further, it is suggestive that portal-in-a-browser model for NDRS may be adopted.

4.8 Proposed Services of NDRS

The major goal of NDRS is to link repositories in such a way that services can be built upon them that provide to value and utility to the user. Apart from browse and search facilities, following services are proposed to be made available to end users:

- It should integrate name authority, citation analysis and automated subject classification services into their application software. These should be incorporated so as to provide consistency and uniformity of data[11].
- Metadata encoding transmission services (METS) wrapper should be used with the descriptive metadata expressed using the DC Library Application Profile. It will help in digital preservation of these resources for a long time and METS should be adopted as national standard for preservation of digital objects in the country[12].
- Regard to the poor quality of metadata, it is possible to develop middleware programme to analyse the harvested records in number of ways. For example, for file format, NDRS should build a list of possible file formats used by various data providers and identify risks associated with by proprietary or obscure formats. Also in case of type field which is mostly interpreted in a wrong way, strategy should be evolved to get suitable data from data repositories[13].
- Persistent identifiers would be evolved either using open source like PURL or commercial handler so that each resource should be identified with persistent identifiers.
- Some features such as personalisation, annotation, alerting and linking to related documents in search results should be developed.
- In the beginning, it has been proposed that NDRS harvests only metadata from the member institutes. As bandwidth is increasing and storage costs have continued to diminish, it has become feasible to harvest metadata and the associated digital objects, both documents and ancillary data. This process has been facilitated by the emergence of a protocol for metadata and digital object harvesting related to the OAI-PMH. In future, full-text could also be harvested and used to support value-added services; therefore, NDRS can be also used as the basis for preservation. However, to provides rights over OAI-based full-text, authentication and authorisation services may also implemented over NDRS[14].
- The OA movement and the OAI are both rooted in cooperative and collaborative philosophies. Therefore cooperation and collaboration would be important keys to the development of future information environment, therefore, policy administrators and developers need to work together.

5. RECOMMENDATIONS

There are many technical, social, and expertise issues required for establishing IR such as selection of hardware, communication bandwidth, suitable software, international standards and documentation, including operational aspects like loading software, uploading data, set-up test server, manage process, maintenance and load material advocacy and preservation.

- The NDRS has to provide consultancy and training about all the aspects both at national as well as institution level, particularly for health science institutions in the country. Considering the number of health science institution, at various levels (research down to college level) in the country, it becomes necessary to establish a national body in the line and pattern with JISC (http://www.jisc.org) in UK to provide advisory as well as technical services to individual repositories.
- A few health sciences institutions in India have IRs and there is a need to motivate and guide majority of other health sciences institutions to establish an interoperable IR and encourage authors to deposit the full-text of their research output to IRs.
- There is a need to create awareness and educate the authors about the copyrights, so that they can withhold rights to submit their publication to IRs and the institutions should make it mandate to submit their publication to open access publications.
- Research councils/apex bodies/funding agencies (such as ICMR, DSIR) have to adopt policies on mandating OA in all institutions, under their funding purview like in National Institute of Health (NIH) and Howard Hughes Medical Institute in the USA, Wellcome Trust and the Research Councils, UK in the UK and the European Research Council.

- Health science academies (like Neurological Society of India and others) should play an active role in building open access culture, with active support of National Knowledge Commission.
- The unique chance of health sciences institutions is that there are many case discussion reports which trigger research to incorporate in IRs as an unpublished works.

6. CONCLUSIONS

The main principle of proposed NDRS model is to achieve more economical and robustness of the services including storage and retrieval. As per the study, majority of the existing IRs are suffering due to lack of requisite storage space, experts, and expertise. Therefore, NDRS should have responsibilities to provide technical supports like installation of IR software, training, guidelines in the use of application and customisation through formal and informal channel of communication. The goal of NDRS is to develop a linking model of OA repositories together effectively, and permit service providers to develop their offerings over the whole corpus of OA materials. As the scholarly communication is moving towards barrier free access to information, NDRS should able to support world-class teaching, research and scholarship in a cost-effective way.

In building intellectual capital, through cooperative collection, storage and management initiatives, it is required to institutionalise for proper governance and to mobilise required fund, manpower, experts and expertise in cooperative and collaborative mode. There are many model suggested in the study they are establishing like an institution such as shared model or consortium repository model (CRM), platform model, aggregation model or participatory institution model (PIM) and proprietary repository model (PRM).

The NDRS is proposed as a PIM, having its own participatory policy and programme direction, where it provides opportunity for representation from participatory IRs, experts/scholarly contributor in the subject, technologists, finance representative from the funding apex body and NDRS administrator as Member-secretary. Proposed NDRS should have its own objectives, policies for resource selection and management, withdrawal items from repositories, collection ownership, intellectual property rights, architecture, and type of services to be introduced and updated from time to time.

REFERENCES

1. Reilly, B.F. Developing print repositories: Models for shared preservation and access. CLIR, Washington. 2003. http://www.clir.org/pubs/reports/pub117/content.html (accessed on 15 September 2011).
2. Reilly, B.F. Underlying organisational and funding models. *In* Developing print repositories: Models for shared preservation and access. CLIR, Washington, 2003. http://www.clir.org/pubs/reports/pub117/contents.html (accessed on 15 September 2011).
3. Brown, J. Comparing consortial repositories: a model-driven analysis. http://eprints.ucl.ac.uk/18977/1/18977.pdf (accessed on 15 September, 2011).
4. Jordan, M. The CARL metadata harvester and search service. *Library Hi Tech*, **24**(2), 2006.
5. Reilly, B.F. Foreword. *In* Developing print repositories: Models for shared preservation and access. CLIR, Washington, 2003. http://www.clir.org/pubs/reports/pub117/contents.html (accessed on 15 September 2011)
6. Reilly, B.F. Introduction: Purpose of the study. *In* Developing print repositories: Models for shared preservation and access. CLIR, Washington, 2003. http://www.clir.org/pubs/reports/pub117/contents.html (accessed on 15 September 2011).
7. Reilly, B.F. Regional repository policies: Selection and management regimes. *In* Developing print repositories: Models for shared preservation and access. CLIR, Washington, 2003. http://www.clir.org/pubs/reports/pub117/contents.html (accessed on 15 September 2011).
8. Zhu, Z.; Ma, J.; Lu, L.; Liu, W. & Wu, D. Development of an institutional repositories network in Chinese Academy of Sciences. *In* 2nd IEEE International Conference on Information Management and Engineering (ICIME), Chengdu.
9. SRW/U. OCLC. http://www.oclc.org/research/activities/srw/default.htm (accessed on 15 September 2011).
10. Swan, A. & Awre C. Linking UK repositories: Technical and organisational models to support user-oriented services across institutional and their digital repositories: Scoping study report: Appendix. JISC. http://www.jisc.ac.uk/uploaded_documents/Linking_UK_repositories_appendix.pdf (accessed on 15 September 2011).
11. Jordan, M. The CARL metadata harvester and search service. *Library Hi Tech*, **24**(2), 2006
12. Swan, A.; Needham, P.; Probetes, S.; Muir, A.; Oppenheim, C.; O'Brien, A. & Rowland, F. Developing a model for e-prints and open access journal content in UK further and higher education. *Learned Publishing*, 2005, **18**, 25-40. http://eprints.ecs.soton .ac.uk/11000/1/Eprints_LP_paper.pdf (accessed on 15 September 2011).

13. Martin, R. ePrints UK: Developing a national e-prints archive. *Ariadne,* **35**. http://www.ariadne.ac.uk/issue35/martin/ (accessed on 15 September 2011).

14. Kumar, S. & Karisiddappa, C.R. METS: Metadata standards for digital preservation in the context of institutional repositories. *In* Digitisation and Metadata Standards, DESIDOC, 2009.

About the Author

Mr Surinder Kumar is working as Senior Technical Director at National Informatics Centre (NIC), Department of Information Technology, New Delhi. He joined NIC in 1989 and worked on development of National databases such as 'IndMED', medIND, UNcat, and Union Catalogue of Biomedical Serials in India. He is also involved in imparting training to medical librarians and professionals from 1989 onwards. He has conducted several workshops on: CDS/ISIS (with NISSAT), information retrieval (with ILA, SIS and DELNET), and DSpace (with TERI, DELNET, Delhi University and others). He has been working on developing repository of the official debates of Rajya Sabha using DSpace containing around 4.5 lakh records. He is working on number of digitisation projects to the Government of India. He has published 35 research papers in various National and International journals and proceedings.