

Institutional Repository at National Aerospace Laboratories: A Case Study

Poornima Narayana, Jayashree S and Indrani V

Information Centre for Aerospace Science and Technology,
National Aerospace Laboratories, Bangalore, Karnataka India
{poornima, jay, indrani}@css.nal.res.in

Abstract. The Information Centre of NAL with its state-of-the-art expertise, infrastructure and services initiated setting up of its own repository during 2003 using, the then most popular open source software Greenstone Digital Library (GSDL). The work progressed rapidly with the adoption of the open source software GNU EPrints for archiving and managing the digital collections. This paper in detail explains the working model of NAL's Institutional Repository. It discusses the technology employed and methodology adopted in building the same. The collection process of different data types, processing and depositing the same to IR is discussed in detail. It is summarized by providing the current status and the statistics on number of hits received.

Keywords: Institutional Repository, Open Access Initiative, National Aerospace Laboratories, ICAST

1 Introduction

The emergence of Internet has brought enormous opportunity to bring the results of research primarily to academicians, scientists and scholars through digital communication to anyone at any time, anywhere in the world. Although it is now possible to have free access to exhaustive information on the web, still significant amount of research is not available freely. While the delivery techniques for scientific publications have changed rapidly, the economic ramifications have hardly changed. During the 1990s several e-print archives as well as a few hundred peer-reviewed, electronic, scholarly journals emerged.

The journal crises lead to 'Open Access movement' advocating free access to scientific research output. Open access may be defined as "a philosophy to achieve the goal of accessing and making available the digital material free of charge, which may or may not be free from copyright and licensing restrictions".

2 Institutional Repository

Institutional Repositories represent an important OA-channel and are relatively new developments in the scholarly communication process compared to open journals and subject-specific repositories. The characteristics of IR include - Institution-based; Scholarly material in digital formats; Cumulative and perpetual; Open and Interoperable. Institutional Repositories are “digital archives of intellectual products created by the faculty, staff and students of an institution or group of institutions accessible to end users both within and outside the institution”. The IR may hold various types of publications, such as pre-prints and post-prints of journal articles, conference papers, research reports, theses, dissertations, seminar presentations, working papers and other scholarly items.

Another service that gained momentum after establishment of different institutional repositories is the metadata harvesting for indexing metadata from OAI-compliant archives or repositories and journals. These services are supported by a protocol known as OAI-PMH (Open Access Initiative - Protocol for Metadata Harvesting). Some of the well known Metadata harvesters are: OAIster set up at University of Michigan which harvests metadata from 670 institutions, ARC at Old Dominion University, ArXIV of physics, mathematics and computer science, UIUC registry of cultural heritage materials etc.

3 IR in Indian Context

A number of Indian scientific research institutions, universities and corporate R & Ds produce high quality research accompanied by innumerable scholarly communications published by national and international journals and conference proceedings. India has adopted the Open Access much ahead of other developing countries. A good number of high quality, peer-reviewed open access journals are being published by India covering a wide spectrum of subjects. More than 22 academic and research institutions have set up their Institutional Repositories as indicated by ROAR (Registry of Open Access Repositories) viz., IISc, IIMK, ISI, NCL, NIO, RRU, NAL, NIT and so on.

Some of the metadata harvesters in India are ‘Search Digital Libraries’ of DRTC/ISI, ‘Knowledge Harvester@INSA’, ‘SJPI Cross Journal Search Service’ from NCSI and ‘SEED’ from IIT, Delhi.

4 About NAL

NAL, a constituent of CSIR is India's pre-eminent civil R & D establishment in aeronautics and allied discipline and has made very significant contributions to all Indian aerospace programmes. It is spearheading the effort to design and develop small and medium size civil aircrafts in India. NAL is well equipped with modern and sophisticated national facilities like the Wind Tunnel Centre and the computerized full scale fatigue test facility. The laboratory has developed two major aircrafts, HANSA, the first composite trainer aircraft and the SARAS, a fourteen seater multi role light transport aircraft.

4.1 Information Centre at NAL (www.icast.org.in)

The Information Centre of NAL (ICAST), an ISO-9001:2000 Certified Centre along with its parent organisation has been recognised as the National Information Centre for Aerospace Science and Technology by the National Information System for Science and Technology (NISSAT/DSIR), UNESCO and AR&DB. The ICAST with its state-of-the-art expertise, infrastructure and services caters to the information requirements of the Indian aerospace community in particular and the engineering and technical personnel in general. The centre is well known for its aerospace collections of books, journals and specifically technical reports from NASA, DLR, ONERA, NLR, ARL and UTIAS. It offers Online, Web and CD-ROM based literature search services including access to e-journals, Newspaper Clippings, Online Public Access Catalogue and Union Catalogue of Serials. The centre has setup a portal 'AeroInfo' (www.aeroinfo.org.in), first of its kind in the country, which serves as one window information search facility for aerospace web sources in the world in general and India in particular.

4.2 Infrastructure

NAL has FDDI network of 1GB with 100mbps LAN and 2 mbps Internet connectivity. ICAST is well equipped with Application Servers, Terabyte storage and backup solution, high end scanners like Minolta book/microfiche scanners and good number of flatbed scanners with back to back and batch mode facility, online network printers, digital presenter with touch screen facility and LCD projector.

5 NAL-IR (www.nal-ir.nal.res.in)

The Institutional Repository at NAL is the digital archive of the research output of the scientists. Since the inception of NAL during 1959, till date the R&D staffs have published more than 20,000 research publications in various forms.

ICAST initiated setting up of its own repository during 2003 using, the then most popular open source software Greenstone Digital Library (GSDL), developed at University of Waikato, NZ. More than 300 papers at abstract level along with few full text contributed by the scientists at NAL were uploaded, but were made accessible only through NAL's Intranet. During 2004, the work progressed rapidly with the adoption of another open source software GNU EPrints for archiving and managing the digital collections. The knowledge base of NAL-IR covers Journal Articles, Conference Papers, Technical Reports, Presentation/Lectures, Preprints, Project documents, Patents, Thesis and Images.

'Metadata' is the key for content management, content organization, and bibliographic control. NAL-IR has adopted the qualified Dublin core metadata schema to describe the bibliographic content of the knowledge base. Two metadata qualifiers such as 'NASA Subject headings' for indexing the documents, because of its exhaustive coverage in the Aerospace discipline and the major 20 divisions at NAL have been incorporate. The divisions existing earlier but of late either merged with other divisions or do not exist have also been incorporated.

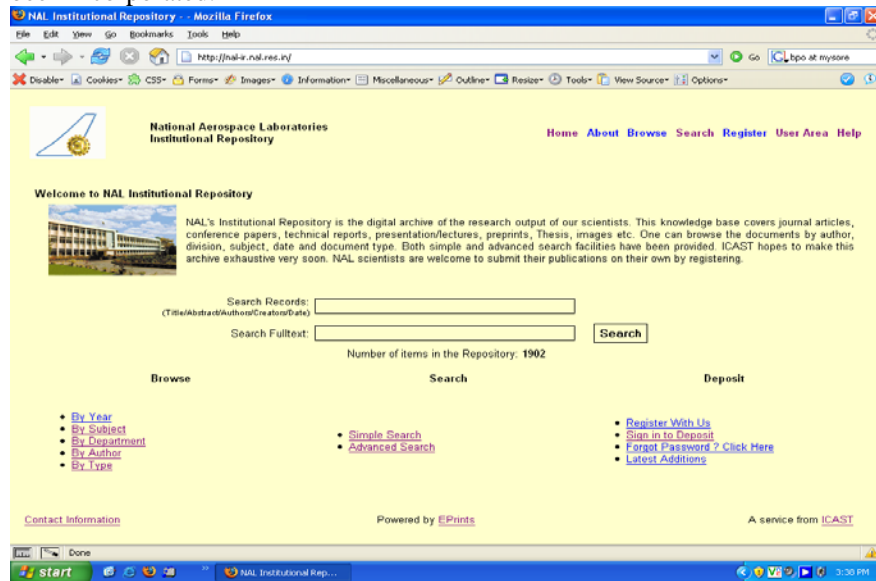
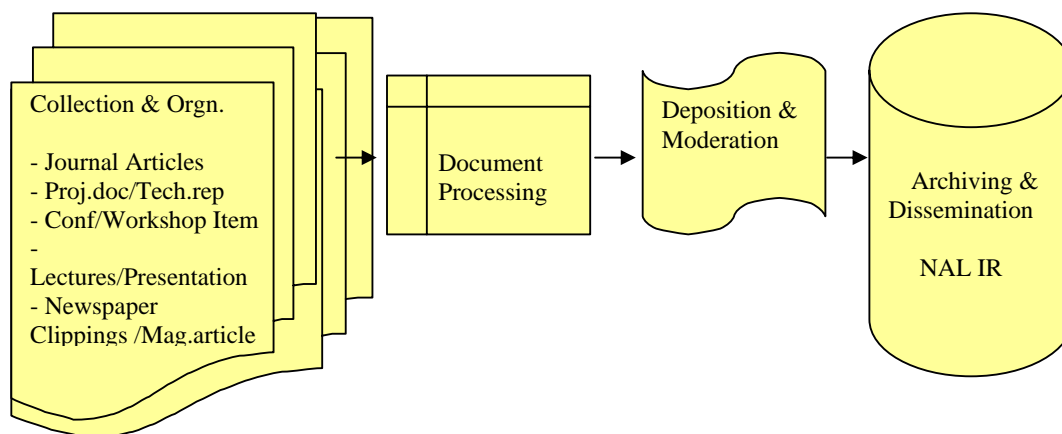


Figure1. Institutional Repository of NAL

6 Technologies Employed and Supported

NAL-IR runs on GNU EPrints Open Archive Software (revision EPrints 2.3.6), developed by the University of Southampton, England and freely distributable archive system available from *software.eprints.org*. This IR is powered by MySQL for database, Apache Web server, scripting language PERL, XML standard for retrieving, the Document Object Model (DOM) platform- and language-neutral interface that allow programs and scripts to dynamically access and update the content, structure and style of documents, ParaCite for both reference parsing and location. It supports OAI-PMH for effective interoperability, VLit for electronic locations (URLs), Valid XHTML and Valid CSS for checking Cascading Style Sheets for display.

7 Methodology Adopted



7.1 Collection & Organisation

While the scientific staffs are trained to contribute their publications and other documents directly to IR, but it is mostly the staffs at the information centre are responsible for collection of the documents available both in soft copy and hard copy form. and upload them to IR.

7.1.1 Data Collection work-flow @ the Information Centre:

- Collection of documents from Annual Report – Recent publications

The Annual Report published at NAL covers research output of NAL scientists, which is one of the Master list for

collection building. Authors are contacted either for hard copy or soft copy preferably of their research reports.

- Interaction with scientists personally – Archival Collection
Scientists are contacted personally for their respective publications for archival collection
- Searching through Aerospace Database
CD-ROM/online version is used for searching the research publications of NAL scientists, especially for technical reports and journal articles.
- E-Journals

Searching is available from 4000 e-journals of 20 major publisher's viz., Science Direct, Oxford University Press, Cambridge University Press, etc. for bibliographic details as well as full-text papers.

7.2 Document Processing

The scientists provide their articles either in hard copy or soft copy. The hard copies of the research papers are digitized by using the high end scanners and Adobe Acrobat 5.0 with default setting of 300dpi and are converted into editable pdf format. The Electronic versions of the documents received are converted into editable pdf format.

7.3 Deposition process

Because of the restricted nature of the ongoing projects at NAL, policy decision regarding the level of accessibility, the publications to be archived has been defined by the management at NAL including the IR managers. With advertisement, promos and training, scientists are encouraged to upload their papers to IR, though information centre staffs upload majority of the papers. The publisher's copyright & archival policy is checked (<http://www.sherpa.ac.uk/romeo.php?all=yes>) before depositing the full-papers to the IR.

Once the user registers, the deposition process of EPrints has the following fields...

- EPrint type (Journal Article/Conference Papers/Patents)
- Core bibliographic Information (Title/Author/Dept)
- Publication information (Publisher name/Place/DOI/Size of the document)
- Status of item (Published/unpublished/inpress)
- Abstract & References
- Subjects (Aeronautics/Astronautics/Engineering/Chemistry)
- Additional Information (Copyright issues)

- Documents attaching (Uploading full-text paper)
- Moving/Editing document into eprint (Complete metadata is listed)

Once the document is deposited, it is stored in the buffer. Each document has to be reviewed before it is archived. The reviewer has to check the management policy, publisher's copyright policy and errors if any, before final submission to the main archive. In case of any missing elements, the document bounces back to the depositor's buffer for correction. Once the reviewer bounces the document, the depositor gets an email alert and the corrected document again goes back to the reviewer for the necessary submission.

8 Browse and Search Features

The following are the browsing features:

By Year– Allows browsing from 1962 to present.

By Subject – Allows browsing by 99 NASA Aerospace subject categories.

By Department – Allows browsing through more than 20 departments.

By Author – Author browse displays author names followed by number of articles archived.

By Type – This covers browse by journal articles, conf papers, tech reports, patents etc.

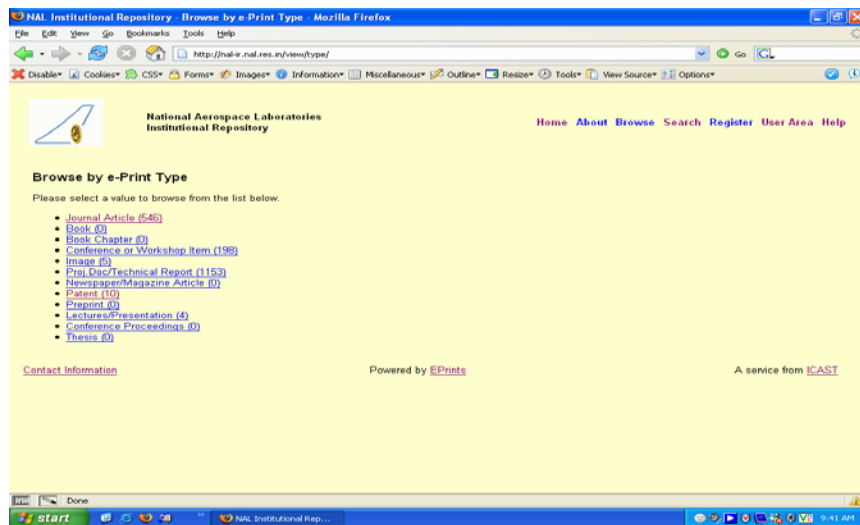


Figure 3. Browse by Document Type

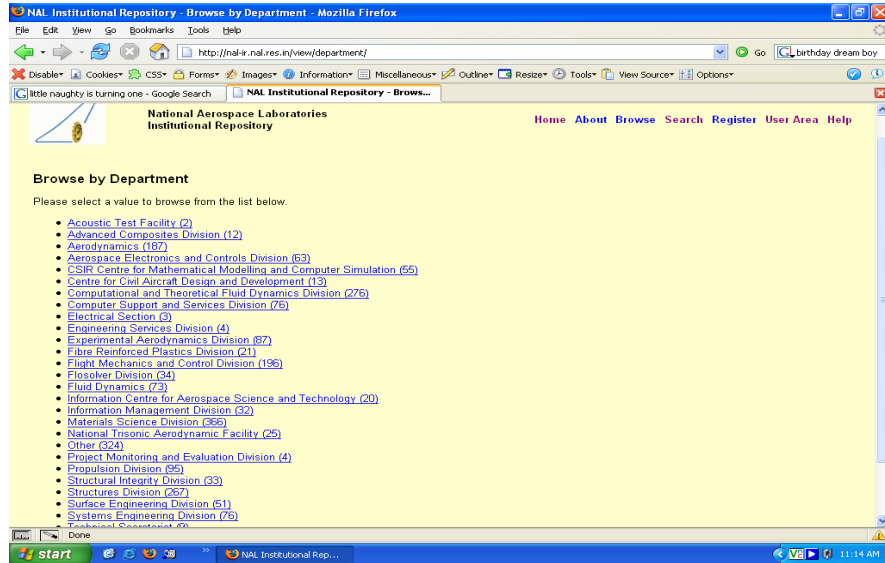


Figure 4. Browse by Department

NAL-IR provides both Simple and Advance search features. Advance search allows field level searching of nearly a dozen fields for ex: title, author, subject, department etc.,

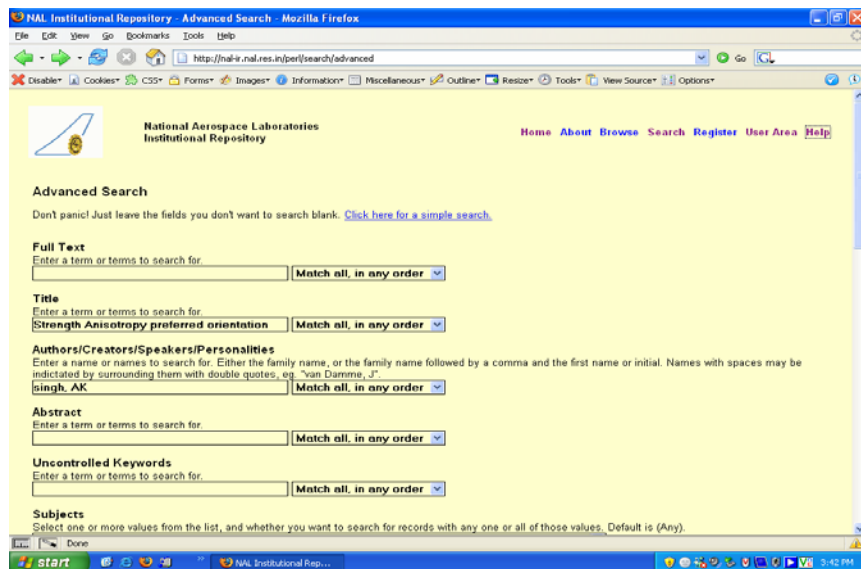


Figure 5. Advanced Search options

9 Status and coverage

At present there are more than 2000 records in the repository. The majority are the Project Documents/Technical Reports of NAL which is around 1500, followed by journal articles (560), Conference/Workshop papers (200) and Patents (10). The other document types, like, Theses, Book Chapters and Newspaper articles are yet to be covered.

10 Statistics

10.1 Country-wise Downloads

More than 13,500 downloads of the full texts of NAL-IR has been recorded between Jan. to 25th June 2006 (Table1). This statistics does not include downloads within NAL. The number of documents downloaded varied depending upon the date of uploading and thereby this statistics is basically an indicator of download trend of IR documents. Notably, downloads from United States is highest with 33.6% followed by India 23.61%, United Kingdom 9.48%, Canada 3.63%, Japan 3.2%, China 1.65%, Germany 1.37%, Netherlands 1.21% and so on.

10.2 Document-wise Downloads

Table – 2 gives the range of statistics of downloads of the documents held by NAL-IR. It is satisfying to note that >40% of the documents of the archive have been downloaded, although the number of downloads vary from document to document. Three documents have been downloaded more than 200 times followed by 8 documents from 100-200; 13 documents were downloaded from 50-100 and so on.

Sl. No.	Country	Total Access	Percentage
1	United States	4550	33.63
2	India	3195	23.61
3	United Kingdom	1283	9.48
4	Canada	491	3.63
5	Japan	433	3.2
6	China	224	1.65
7	Germany	186	1.37
8	Netherlands	164	1.21
9	Sweden	146	1.08
10	France	138	1.02

Table1. Country-wise Full Text Downloads (Jan-June 2006)

Range of downloads	No. of Documents
200 - 379	3
100 - 199	8
50 - 99	13
25 - 49	23
20 - 24	21
15 - 19	18
> 15	175

Table2. Document-wise Downloads (Jan-June 2006)

11 Metadata Harvesting Service

NAL has initiated harvesting OAI compliant IRs of CSIR Laboratories in India. Currently, this service has harvested metadata from the archives of 3 sister laboratories, NAL, Bangalore, NCL, Pune and NIO, Goa using PKP Harvester developed by Public Knowledge Project UBC, Canada.

12 Immediate and Future Plan

The Centre plans to adopt DSpace Open Source Software for archiving and maintaining the NAL-IR. Once the job concerned to usual document types like journal articles, conference papers and technical reports is completed, other document types like presentations, images, patents and theses would be taken up. The metadata harvesting facility is still in experimental stage. Shortly the same would be taken up on a regular basis covering IRs of all 40 CSIR chains of laboratories and also those of aerospace institutions in the country. The centre is in the process of conducting workshops for training staff of both CSIR and aerospace set ups.

References

- [1] Arunachalam, Subbaiah (2004). India's March Towards Open Access. *SciDev.Net* 5 march 2004.
- [2] Bo-Christer Bjork (2004). Open Access to Scientific Publications: An Analysis of the Barriers to Change?. *Information Research*, 9(2).
- [3] Branin, J, Smith, M, Staples T, Van de Velde. (2002). Institutional Repositories: Part1: Overview and Organizational Issues. Pt 2: Functionality and technical issues.
<http://www.cni.org/tfms/2002b.fall/projects.html>
- [4] Coleman, Anita, Roback, Joseph. (2005). Open Access Federation for Library and Information Science. *D Lib magazine*, 11(12)
- [5] Das, Anup Kumar; Sen, B K; Dutta, Chaitali. (2005). Collection Developments in Digital Repositories in India.
http://mit.gov.in/apr_2005.htm
- [6] David Prosser. (2003). Information revolution: Can Institutional Repositories and Open Access Transform Scholarly Communication. *ELS gazette*. 15.
- [7] Lynch, Clifford A (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. In: ARL Bimonthly Report 226, February 2003.
- [8] Rajashekar T B. (2003). Improving the Visibility of Indian Research: An Institutional, Open Access Publishing Model. *Indo-US Workshop on Open Digital Libraries and Interoperability*, June 23-25, 2003.
- [9] Suber, Peter (2002): Open Access to Scientific Journal Literature. *Journal of Biology*, 1(1), 1-3.
- [10] Suber, Peter, Arunachalam, Subbaiah (2006). Open Access to Science in the Developing World: Open Access, Unrestricted Access to Published Research. *93rd Indian Science Congress*, 3-7, January, Hyderabad, India.
- [11] Yeats, Robin. (2003). Institutional Repositories. *Vine*, 33 (2), p. 96-101.