DIGITAL COLLECTION BUILDING: A CASE STUDY

Ashalatha Laxminarsaiah and Iqbalahmad U Rajgoli

Indian Space Research Organisation Headquarters
Antariksh Bhavan, New BEL Road, Bangalore, India
(asha, iqbal786)@isro.gov.in

The Digital Library plays an important role in creating, preserving, indexing, and retrieving the various forms of digital assets. The ISRO HQ Digital Library provides cost effective facility for systematic archiving of research results and other documents of the institution. This paper discusses about various E-collections at ISRO HQ Library and explains the infrastructure (Hardware, Software and Network) and other requirements for creating and managing E-collections. It provides web-based mechanism to deposit and access the same to the user community on to their desktop through ISRO SPACENET (Intranet). This paper illustrates how to set up Digital Library using DSpace software and how to create different communities and under each community, many collections and how to customize the metadata fields according to institutional requirements.

Keywords: Digital Library; E-collection; DSpace; ADRIN news tabloid and Digital library

1 Introduction

The library, being a social institution, organized and administered for the benefit of the members of the society depends on the financial and other material resources made available for providing satisfactory library services to the users. The greater the library resources (i.e. including e-resources) better the services to the users. Libraries should attempt for more and more services to the readers. Digital Libraries of today should be in a position to provide the readers about all the literature available on the subject of interest, from where it can be obtained and retrieved in appropriate time.

2 Space Science Centres

Governments carry a substantial part of expenditure on the space industry, as some parts of a country's space-related efforts (e.g. R&D) are of a public nature. Expenditures of low-income countries such as China and India have a higher purchasing power than similar expenditures in high-income countries like USA and UK. Space department budget estimates for the year 2006-2007 is Rs.3610.42 crores.

The secretariat of DOS and Headquarters of ISRO are located in Bangalore. The overall coordination of the space programme is carried out by programme offices of ISRO Headquarters in different areas like satellite communications, earth observation systems, launch vehicle programmes, space science, technology transfer and industry coordination, international cooperation, publications and public relations and budget and economic analysis.

The major establishments of DOS are as follows:

Antrix Corporation Limited
Development and Educational Communication Unit (DECU)
INSAT Master Control Facility (MCF)
ISRO Inertial Systems Unit
ISRO Satellite Centre (ISAC)
ISRO Telemetry, Tracking and Command Network (ISTRAC)
Laboratory for Electro Optic Systems (LEOS)
Liquid Propulsion Systems Center (LPSC)
National Atmospheric Research Laboratory (NARL)
National Remote Sensing Agency (NRSA)
Physical Research Laboratory
Regional Remote Sensing Service Centre (RRSSC)
SHAR center (SHAR)
Space Applications Center (SAC)
Vikram Sarabhai Space Centre (VSSC)

In addition to this Space programmes entails cooperation with a number of academic and research institutions, industries, the end users and space agencies in other countries. Indian Space Programme, since its inception, has pursued international cooperation very actively with countries like Brunei, Canada, China, Europe, France, Germany, Hungary, Indonesia, Japan, Mauritius, and USA, Sweden etc.

2.1 Space Science Libraries in India

ISRO/DOS libraries are of different sizes in collection and budget. They also function independently in each centre. There are four major libraries: VSSC, ISAC, SDSC, SAC, two medium size libraries: ISRO HQ, NRSA, nine smaller libraries are ISTRAC, MCF (Hassan), LPSC (Valaimala), LPSC (Mahendragiri), LPSC (Bangalore), NARL, PRL, IIRS and RRSSC five centres (Nagpur, Kharagpur, Jodhpur, Bangalore and Dehradun; libraries run by non-professional staff), which are located at different places in the country. In ISRO/DOS there is no nodal agency and have decentralized management of libraries. All libraries function independently and have access to resources to each other's resources with Inter Library Loan facility.

3 About ISRO HQ Library environment

The Library and information services at ISRO HQ is a specialized library among other ISRO centres libraries regarding its collection and services catering to the needs of more than 500 patrons most of them are scientists and top level management people and other institutions. ISRO HQ, Bangalore, internally generates enormous amount of information in the form of lectures/speeches delivered by eminent personalities of ISRO, documentary films, internal reports, reprints, preprints, conference proceedings, conference papers and etc. The library currently holds 20,000 books, 1500 Hindi books, 14,000 reports, 1,500 standards, 1,300 CD-ROMs, 1,800 non-book materials and subscribes to 150 periodicals.

3.1 Scope of the ISRO HQ Digital Library

ISRO HQ Library resources are of heterogeneous in nature apart from the format also. ISRO HQ Digital Library has played a major role in delivering the contents efficiently and effectively to the desktop of users through Intranet (SPACENET) of ISRO throughout India.

In order to keep users abreast with latest scientific and technical news of importance, initially it was decided that the Digital Library would include only news items (Newspaper clippings). The important news items are identified, and downloaded from online version of leading national dailies and few foreign websites, and are archived into server on the same day using ADRIN (Advanced Data

Processing Research Institute) News Tabloid in-house developed software. Current and retrospective search facility is provided on ISRO SPACENET (i.e. Intranet) for the benefit of ISRO community. Users have the option to view the full text of the news item by date, title, and keywords etc.

ISRO is an R & D institution, generates information in multi-disciplinary subjects like satellite applications, remote sensing, tele-medicine and tele-education and other allied topics in various forms like lectures/ speeches delivered, scientific and technical reports, papers published either in a journal or in conferences held at various ISRO centers, and in collaboration with other institutions etc. Every effort has been taken to acquire e-form of publications from the respective authors at the time of submission for digitalization. The same is digitized retrospectively. To digitize these documents best quality scanner is used. The scanned document is converted into searchable PDF document. In addition to this ISRO organizes seminars/conferences on various topics, and as far as possible the same publication received either as CD or DVD format, and contents of the same are uploaded to DSpace platform under various collections. The user community can access the content on Intranet by title, author, keyword, report number and year of publication.

ISRO HQ Library has a good collection of documentaries in the form of films of satellites, launch vehicles and etc. Recently films have been converted to video format and it has been uploaded to DSpace collection on Intranet and the user community can access the same.

3.2 Digital Library Initiative

ISRO HQ Library has established the digital information system in a multi-user environment using LINUX operating system. Access to information, search and retrieval from client nodes is made easy with Graphical User Interface (GUI) through web based browser. Digital Library of ISRO HQ is hosted on Intranet (SPACENET) along with OPAC and other services on library home page as shown in Figure.1

Figure 1: ISRO HQ Digital Library Portal

3.3 E-collection at ISRO HQ Library

E-collections at ISRO HQ consists of categories as shown in Figure 2.

Figure 2: E-Collections at ISRO HQ Library

3.3.1 Infrastructure requirements for creating and managing E-Collections

Basic requirement for E-collections and digital library is well established and planned network environment in an organization. At ISRO HQ nearly 600 terminals are having connectivity with 100 Mbps local area network with high-speed multiprocessor servers and powerful workstations for high data transfer rate over Intranet (SPACENET).

3.3.2 System requirements

Hardware requirements: ISRO HQ at present has Pentium IV processor with 80GB memory on which E-collections are hosted.

Software requirements: The server is running on Red Hat Linux 9.0. DSpace digital library software is installed on the server end and can be accessed through Intranet by any ISRO Centre all over India.

4 Software selection

When it was decided that the library would render the News alert service, the next question was selection of software. Advanced Data Processing Research Institute (ADRIN), a unit of Department of Space located at Secunderabad has developed an Automatic News Gathering software **"ADRIN NEWS TABLOID".** This in-house software has been developed using a set of programs based on Visual Basic 6.0, Internet Explorer 6.0 and Microsoft Access (2000 version), runs on windows 95/98/NT/2000 versions of operating systems as shown in Figure 3. This software was exclusively developed for gathering news items from daily newspapers. News gathering takes place in many ways i.e., category-based, website-based, keyword based and search engine-based.

The Automatic News Gathering program developed by ADRIN attempts to combine both website-based and Keyword-based Internet search techniques into one and offers features to search Internet both for Web content (Static web pages) and News (Dynamic web pages) as shown in Figure 4. Initially when the News alerts service was started, Keywords and Websites were defined. This program gathers News items from various web sites as shown in Figure.5.

This programme offers to organize results found in the search in a database and appropriate news items were posted on the Intranet website.

Figure 3: ADRIN News Tabloid (ANT) - Opening window

Figure 4: ADRIN News Tabloid (ANT) – Opening Window Settings

Figure 5: ADRIN News Tabloid (ANT) – News Display Module

Digital Library cannot contain only newspaper clippings alone. The amount of research output brought out in the form of research papers, research reports, standards etc., should also form part of this repository. When these points were taken into consideration, it was noted that the news gathering software doesn't have the facility to support different forms and formats of various documents to be hosted on Digital Library. Hence, the search for software, which can support all these formats, was mooted.

During the process of searching the appropriate software for building the ISRO HQ Institutional Digital Library, many types of software such as Greenstone, E-Prints etc., came on the scene, which are free and can be downloaded easily. Few well-known softwares are available on Internet and can be downloaded from the Internet. Most of the Universities and Research and Development organizations throughout the world are actively planning and implementing Institutional Digital Library. The software has to satisfy three criteria:

- Available as an open source license, i.e. they are available for free and can be downloaded, customized as per the institutional requirement and can be upgraded.
- Comply with the latest version of OAI metadata harvesting protocols - This helps in the implementation and can be participated in a global network.
- Currently released and publicly available software.

4.1 Some selected softwares which are used by different institutions, are:

- **Archimede:** Software developed by Laval University Library, Canada was designed for electronic preprints and post prints of institution faculty and staff Archimede was designed to support multilingual institutional implementations and it organizes the content submission process in intranet environment managed locally by research community. Archimede Institutional Repository system manages the university or institution electronic theses and dissertations. This was developed on a variety of Java open sources technologies, runs on many operating systems and can be

installed on an existing technical infrastructure of any institution (http://archimede.bibl.ulaval.ca).

- **CERN Document server software** (CDSware): This software was developed and maintained by CERN (The European Corporation for Nuclear Research) to support electronic preprint servers, online library catalog and web-based document repository systems. CDSware was developed to handle large repository of different format of documents (http://cdsware.cern.ch).
- **DSpace:** MIT (Massachusetts Institute of Technology) and HP **(**Hewlett-Packard) have created DSpace software as digital repository to manage intellectual output of multidisciplinary Research and Development organizations. DSpace support digital preservations, planning and managing institutional repository in a large institution. DSpace allows workflow and customization. It supports community/ collection based content and submission by different user community (http://www.dspace.org).
- **E-prints:** Is the largest and widely distributed, installed software developed by University of Southampton, with minimum technical expertise. The software can be installed by any institution world over. By its integrated advanced search, extended metadata and other features, the software can be customized to local requirements (http://www.eprints.org).
- **Greenstone:** This software is for building and distributing digital library collection. New Zealand Digital Library Project at the University of Waikato has developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open source software can handle multilingual documents, with search and browse facility under GNU General Public License (http://www.greenstone.org).

Each of the above software systems is designed to meet the original requirements of developing Digital Library. Archimede was designed to support multilingual institutional implementations. CERN can handle large Digital Libraries with different types of documents. DSpace supports community based content policies and

submission process and accommodates various kinds of digital document formats. Eprints is a useful Digital Library system with large user community.

But when there is a need for technical support and training in using the software, DSpace was found suitable. Though many libraries in the city are using Greenstone and E-Print but the majority of the libraries are using DSpace as it has got tremendous potential and can support numerous forms and formats. It was also noted that by using DSpace, there is a possibility of interacting with other libraries in the city for technical support. Moreover it is open source software and can be customized as per the institutional requirement.

5 Metadata customization

The digital documents of the Digital Library are organized into different communities. New communities will be created as and when they are required depending on the new and emerging areas and activities of ISRO HQ.

Metadata (data about data) describes each resource so it can be discovered and evaluated by the user. It also supplies elements that aid in the administration of the Digital Library, and maintains the structure of the digital library by providing linkages among related objects. The ISRO HQ Digital Library has developed a core set of metadata elements for the project-related documents and other documents to accommodate them in the digital library as shown in Figure 6.

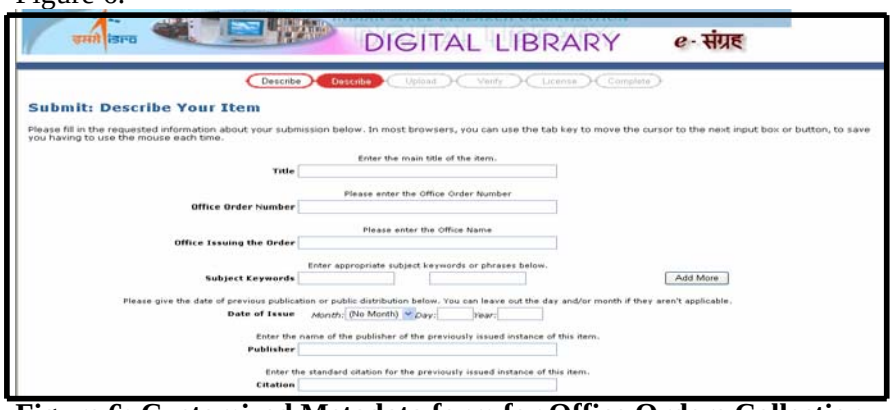

Figure 6: Customized Metadata form for Office Orders Collection

The summaries of the various ongoing projects and initiatives are prepared and sent to the Prime Minister, the Cabinet and the Cabinet Secretary of Government of India. Dr. Rajeev Lochan, Assistant Scientific Secretary and Director, INSES (Office of Information and Software Engineering services) and a team from INSES in ISRO HQ thought of putting these summaries into Digital Library, and make them available to the entire ISRO community through Intranet portal. For this a separate community called "Digital Library of Government Documents" was created. Under this community three collections are included. The back issues of these summaries are scanned and saved in PDF format and were accommodated in the Digital Library.

ISRO issues office orders pertaining to its policies and procedures from time to time. For the official orders issued from time to time, a separate collection called "Office orders" is created. Office orders, which were in hardcopy from retrospective date, were scanned using quality scanner with optical character recognition software, and the document generated in searchable PDF format is uploaded to DSpace collection.

6 Collaboration with other divisions of ISRO HQ

In an organization different sections/divisions play greater role in successfully carrying out any project or initiative. These divisions/sections may not be directly collaborating with each other but even opinions, suggestions and guidance from different divisions is valuable. In building the ISRO HQ Digital Library, library collaborated with INSES division (Office of Information and Software Engineering Services) at ISRO HQ and its main function is to help Chairman, ISRO, in evolving policies related to computers and software and work for then implementation in ISRO/DOS, software engineering, support for mission software validation, inter-centre and intra-centre connectivity through computer network and etc.

Now there are three communities in DSpace (i.e. library documents, government documents, and publications brought out by Publications and Public Relations Office). The government document collection and ISRO publications are scanned and uploaded by Assistant Scientific Secretary Office. Library staff takes care of the collection

pertaining to library. It was decided that INSES would take care of DSpace server maintenance, DSpace software maintenance, configuration, metadata customization and taking database backup regularly. The P&PRU section is mainly responsible for digitisation and uploading of annual reports and Space India (an in-house publication of ISRO).

7 Different formats for different collections

It is felt that maintaining uniformity among different collections is one of the areas of concern while building Digital Library. For ISRO HQ Digital Library, different formats were chosen for different collections. As mentioned earlier, library staff scans the online version of leading national dailies for news pertaining to ISRO, news of national importance and general news. For the daily news updates in the major area of research i.e., space sciences and other subjects of interest to ISRO, five websites have been identified (i.e.,www.spacedaily.com,www.spacetoday.net, www.spacenews.com, www.universetoday.com and www.thespacereview.com) and the important news items are gathered, organized and uploaded to the repository:

- The newspaper clippings are saved in HTML format. Even the font, font size and font colors are taken care of in order to maintain uniformity.
- For the Lectures/Speeches delivered by eminent personalities like Prof. U R Rao, Dr. K Kasturirangan, Dr. G Madhavan Nair and others, it was decided to upload them as MS Word documents.
- For the Conference/Seminar proceedings, which are video captured, MPEG format was found suitable.
- For the government documents and articles published by the ISRO HQ community and annual reports, PDF format is used.

8 Conclusion:

Traditional libraries are limited by storage space; digital libraries/ repositories have the potential to store much more scholarly information and require very little space to contain it. As such the cost of maintaining institutional digital library is much lower than that of a traditional library. Digital libraries can adopt innovations in electronic and audio-video book technology.

Considering the advantages like no physical boundary, round the clock availability of information, multiple access to information resources, faster information search and retrieval, preservation and conservation of exact copy of the original document and cost, ISRO HQ has set up Digital Library using open source software "DSpace". With the development of Digital Library, information access and retrieval has been increased at campus level and at ISRO/DOS level also. The aim is to deliver the right information to the right reader at the right time as enunciated by Dr. S R Ranganathan.

9 Acknowledgement:

Authors would like to thank Dr. Rajeev Lochan, Assistant Scientific Secretary and Director, INSES, ISRO HQ for his support in establishing Digital Library and guidance and valuable suggestions in bringing out this paper.