Chapter 27

DIGITAL ARCHIVES OF RAJYA SABHA DEBATES

Vivek Kr Agnihotri, P Gopalakrishnan, Mukul Pande & N S Arneja

I) OVERVIEW

The Parliamentary debate is the most important record of discussion held on the floor of the House. The debate is required to be preserved for future use not only by parliamentarians but also by public including researchers as it gives the reflection of the various government policies and stand taken by the Parliament on various issues at different times. The Parliamentary reporters are responsible for preparing the record of discussions held during the House proceedings. On a single day, a debate of 300-500 Pages is generated by the Parliamentary reporters. The version generated by the Reporters is called the uncorrected version. The Editorial and Translation Service prepares the Official Version of the Debate after incorporating the minor corrections and editing. This Official version of the House Proceedings, known as 'Official Debates' is printed. The 'Sales and Archives Section' of Rajya Sabha Secretariat keeps the copies of the Printed Debates. At present all the Official debates in Printed format are available since the inception of the Rajya Sabha in 1952. The storage of debates needs lot of space. Besides, with the passage of time, the condition of paper copies of the Official Debates is deteriorating.

Servicing of Members' request for the Debate: Many a times, Members of Parliament need previous debates on a particular topic or of a particular person. Manual searching from the physical debates is very tedious, cumbersome and time consuming process. It used to take several weeks to provide the requisite information to the Members by the branch. Thus there were two major reasons for taking up the Digitization of Debates namely, the safe archival and Quick and easy retrieval.

II) RESULT INDICATORS

1. Key Performance

a. Stakeholder services and benefits achieved through ICT interventions

The major stakeholders of the Parliamentary debates are:

- The Members of Parliament
- The Government Departments
- Journalists and Media Organizations

- Researchers and Research Organizations
- State Legislatures
- Other Countries Parliaments
- General Public at large

The Key benefits of the Online searchable Debates repository are that without visiting any Physical Library or requesting the Parliament Secretariat, one can easily locate the desired and relevant portion from the Parliamentary Debates as per his convenience by just visiting to the parliament Debate Portal of the Rajya Sabha (http://rsdebate.nic.in).

b. % of services covered as ICT interventions

The Rajya Sabha Secretariat has achieved a major milestone by creating the online Debates repository. It has saved the tremendous time of the Members of Parliament as well as any other user in locating the relevant debate and efforts of the Secretariat staff in keeping the Physical Debates as well as in searching the relevant debates on Members requests or for any internal use.

c. Geographical Spread in the State achieved

Since the Rajya Sabha debates are available online on the Internet at address (http://rsdebate.nic.in) and are available freely, They can be searched from anywhere in the globe, wherever the Internet facility is available.

2. Efficiency improvement

a. Time saving / improvements in the delivering the above set of services.

There is tremendous saving of time and efforts in publishing of Debates in the repository. Once electronic text is received, the dedicated manpower prepares the metadata and upload the debate in the server.

b. Cost savings for delivering above set of services.

The Solution is very cost effective as it saves lot of human efforts required in culling out the required debate text from huge volumes of Debates.

c. Time Saving for availing the services (reduction in cycle time)

There is a tremendous saving of time in searching the required debate. Now, all the customers of Debates can search the repository themselves as per their needs. The search Engine is quite fast and gives the instant results.

d. Cost Saving for availing these services

The Members of Parliament and other users of the Debates need not approach Parliament for their requirements. The Debate repository is available in the Cyberspace freely and anyone can browse the repository and search the required debate instantaneously, without spending the money on purchasing the paper copy of the debates.

III) ENABLER INDICATORS

1. Processes

a. Major front end process changes and implemented

The Debate Portal provides the front end for searching of debates. For processing of Debates, separate set of programs have been developed for preparing the meta-data of the Debates and dividing the debates Title/Subject wise. A dedicated team of Officials of E&T Service and hired manpower does this job. The Electronic copy of the current debates is received from the Printer for further processing for publishing on the website.

b. Major back end process changes and implemented

Earlier, only the Paper copy of the debate was being sent to the Printer, who used to re-input the whole debate and prepares the printed version of the Debate. Now, since typing of debate was being done in house on the computer, it was decided that, printer will be provided, soft copy of the debate duly audited by the Editorial & Translation service. It has saved a lot of time required in retyping of the debate by the printer and proof-reading of the same.

2. People and Resources

a. Project management & Monitoring – Full time team in place

The Digitization Project is being executed jointly by Editorial & Translation Service and a team of Officers from NIC and NICSI. The IT Section of the Rajya Sabha Secretariat is providing the administrative support as well as monitoring the project, as a whole.

b. Change management strategy defined and implemented

The debate Digitization was entirely a new project. It required forming new teams for various activities, providing them the requisite training and change in the functioning of the previous method of publishing debates on the website. Careful planning of all the processes was done, procedures were defined and changes were implemented along with the required training.

c. Leadership support (Political, Bureaucratic) and its visibility

The Debates Digitization Project was approved by the Vice President and Hon'ble Chairman, Rajya Sabha. The Project is being monitored at highest administrative level. The success of the Project is the support and monitoring at the highest level.

d. Financial Model (Funding pattern , Business model PPP etc) defined and implemented

In order to scan the old debate and prepare the PDF files and metadata of the debates, open tender was floated by NIC on the behalf of Rajya Sabha Secretariat. The scanning job was awarded and the complete scanning and metadata preparation job is likely to be completed by August 2010. Simultaneously, the data received from the Vendor is being checked by the in-house staff and after making necessary correction is uploaded in the server.

3. Technology

a. Disaster Recovery & business continuity plan defined & implemented

The Debate application has been hosted on the server installed in the NIC's Data Centre at CGO Complex. The IDC has the well defined backup, recovery mechanism as well as Disaster recovery Site at NIC, Hyderabad.

The DSPACE solution was found to be the most cost effective and technically feasible solution which provides all the desired features. Since it was available in open source minimum efforts were required in tailoring it as per the needs of the Rajya Sabha Secretariat. The NIC has the requisite resources for maintaining the Debate application.

b. Security and confidentiality standards defined and implemented

The Debate application has been security audited for any vulnerabilities and hackers attack. The DSapce itself has the inbuilt security features and it is a robust application being used world wide for Digital Libraries.

Creation of metadata in DSpace

Depositors for each item submitted also provide metadata about that item. dspace@official debates of Rajya Sabha holds three types of metadata about deposited item:

Descriptive Metadata: DSpace uses a set of elements and qualifiers from 'Dublin Core' based on Library Application profile set of elements and qualifiers. It consists of title, author, creation date, subject, format etc. As the official debates of Rajya Sabha contain a large amount of data,

data input programs are written to convert data into DSpace format which are then uploaded using bulk import utility programme of DSpace.

Administrative Metadata: This includes preservation metadata, provenance and authorization policy data. Most of this is held within DSpace's relation DBMS scheme. Provenance metadata (prose) is stored in Dublin Core records. Additionally, some other administrative metadata (for example, bit stream byte sizes and MIME types) is replicated in Dublin Core records so that it is easily accessible outside of DSpace.

Structural Metadata: This includes information about how to present an item, or bit streams within an item, to an end user, and the relationships between constituent parts of the item. In addition to some basic technical metadata, bit streams also have a 'sequence ID' that uniquely identifies it within an item. Additional structural metadata can be stored in serialized bit streams, but DSpace does not currently understand this natively.

c. E-people

Many of DSpace's features such as document discovery and retrieval can be used anonymously, but users must be authenticated to perform functions such as submission, email notification ('subscriptions') or administration. Users are also grouped for easier administration. DSpace calls users e-people, to reflect that some users may be machines rather than actual people. DSpace holds the following information about each E-people:

- E-mail address
- First and last names
- Authentication information, such as an encrypted password
- A list of collections for which the e-person wishes to be notified of new items

d. Authorization

DSpace has a flexible authorization system. To perform an action on an object, the user must have permission; DSpace operates a 'default deny' policy.

e. Workflow Steps

A collection's workflow can have up to three steps as described in Table 1. Each collection may have an associated e-person group for performing each step; if no group is associated with a certain step, that step is skipped. If a collection has no e-person groups associated with any step, submissions to that collection are installed straight into the main archive. In other words, the collection receives a submission. If the

collection has a group assigned for workflow step 1, that step is invoked, and the group is notified. The task of performing that workflow step is put in the 'task pool' of the associated group. One member of that group takes the task from the pool, and it is then removed from the task pool, to avoid the situation where several people in the group may be performing the same task without realizing it. The member of the group who has taken the task from the pool may then perform one of three actions, shown in Table 1.

Otherwise, workflow step 1 is skipped. Likewise, workflow steps 2 and 3 are performed if and only if the collection has a group assigned to those steps. If a submission is rejected, the reason (entered by the workflow participant) is e-mailed to the submitter, and it is returned to the submitter's workspace. The submitter can then make any necessary modifications and resubmit, where upon the process starts again. If a submission is 'accepted', it is passed to the next step in the workflow. If there are no more workflow steps with associated groups, the submission is installed in the main archive.

Workflow Step	**Possible actions**
1	Can accept submission for inclusion, or reject submission
2	Can edit metadata provided by the user with the submission, but cannot change the submitted files. Can accept submission for inclusion, or reject submission.
3	Can edit metadata provided by the user with the submission, but cannot change the submitted files. Can accept submission for inclusion, or reject submission. Must then commit to archive; may not reject submission; may not reject submission.

Table 1: Possible Workflow Steps in DSpace

f. Subscriptions

As noted above, end-users (e-people) may 'subscribe' to collections through the Web user interface in order to be alerted when new items are added to those collections. Each day, end-users who are subscribed to one or more collections will receive an e-mail giving brief details of all new items that appeared in any of those collections the previous day. If no new items appeared in any of the subscribed collections, no e-mail is sent. Users can unsubscribe themselves at any time.

g. Search Interface

DSpace uses Lucene Search engine which is a part of Apache Jakarta Project. It provides fielded searching, stemming, and the ability to incrementally add new indexed content without regenerating the entire index. Its API allows indexing new content, regenerating the index, and performing searches on the entire corpus, a community, or collection. Another feature has been added in DSpace application by providing an option to search within results obtained. This has helped the users to refine the search and get the desired results much faster. The various parameters on which the Rajya Sabha Debates can be searched are:
Session Number, Debate Date (yyyy-mm-dd), Debate Type, Debate Title Subject, Title, Members Participated, Question Type, Question Number , Questioner Name, Minister's Name, Ministry etc.

Another important mechanism for document discovery in DSpace is the 'browse'. This is the process whereby the user views a particular index, such as the title index, and navigates around it in search of items. Indices that may be browsed are title, date of debate, members. Additionally, the browse can be limited to items within a particular collection or community. For Rajya Sabha Debates Browse facility has been provided as:

- Debate Title wise
- Members Participated Wise
- Debate Date wise
- Debate Title Subject wise

User has also been given the option of advanced search where user can search on specific field with option of connecting field with Boolean logic. The Rajya Sabha Debates provides Advanced Search facility. Here it searches in the text as well as in metadata.

(Vivek Kumar Agnihotri, Secretary General, Rajya Sabha, Rajya Sabha Secretariat, P Gopalakrishnan, Addtl. Secretary & FinancialAdvisor, Rajya Sabha Secretariat, Mukul Pande, Director, RS & N S Arneja, Sr. Technical Director, NIC, Parliament Division, email: arneja@nic.in)