

Parliament Digital Library

Content Creation, Metadata Standards and Guidelines for Scanning

**Parliament Library
Lok Sabha Secretariat
Parliament House, New Delhi.**

CONTENTS

S.No.	Description	Page No.
1.0	e-Parliament Digital Library – An introduction	1
2.0	List of Type of Documents to be Published in Parliament of Library.	2
3.0	Digitization-Technical Standards and Best Practices followed	3
3.1	Technical Standards - Scanning Mode	
3.2	File Formats	
3.3	Dublin Core Metadata Element Set(DCMES)	
3.4	File Naming Conventions	
3.5	Creation of Metadata and Uploading	
3.6	Digital Files	
4.0	Annexure 1. Metadata Tags 2. Guidelines for Metadata entry	

1.0 E-Parliament Digital Library

Hon'ble Speaker, Lok Sabha constituted a committee for making Parliament Library – a Digital Library and making micro-films of the documents for long term preservation. The committee met several times under the chairmanship of Shri S.S. Ahluwalia, Ex Member of Parliament, Rajya Sabha. The complete project of development of required software for the Parliament Digital Library and its hosting at the NIC national Data Centre was entrusted to NIC. The committee had an interaction with the various Ministries/Departments, in order to obtain the electronic data under their domain, which is required by the Hon'ble Members of Parliament for discharging their legislative duties. The committee took the following decisions.

- The Web Portal for the Parliament Digital library will be developed by the NIC.
- NIC would provide the necessary Hardware, Software and SAN storage for hosting the Parliament Digital Library web portal.
- In the first Phase, the electronic data available in the Lok Sabha Secretariat shall be published in the Digital Library web portal. The LSS shall create the metadata tags required for publishing the data in the searchable repository.
- The electronic data received from other sources such as Central State Library, national Archives, National Library of India and other Ministries Departments shall be published by the LSS after making their metadata.
- In the second phase, Lok Sabha Library shall do the scanning of other physical books and create their metadata for publishing in the Digital Library.
- NIC shall provide the technical consultancy for the project, development of Web portal and hosting of the Web portal at NDC
- The Lok Sabha shall coordinate will all the Ministries/Departments for receiving the electronic data, preparation of their metadata and publishing of the data in the Digital repository.

2.0 List of Type of Documents to be published in Parliament of Library.

1. Annual Reports
2. Parliamentary Committee reports
3. Commissions Reports
4. Demand for Grants
5. Gazette Notifications
6. Papers Laid on the Table of the House
7. Parliament Debates Part II
8. Parliament Debate Part I (Question & Answers)
9. Rare Books
10. Standing Committees reports
11. Trade Agreement
12. White Papers
13. PMA Collection
14. E-documents available in parliament Library
15. e-gazettes
16. Railway Budgets
17. General Budgets
18. Economic Surveys
19. Treaties
20. Foreign Trade agreements
21. Indian Policies
22. Finance Commission and other Commission's Reports
23. Five Year Plans

3.0 Digitization – Technical Standards & Best Practices followed

The proliferation of digital information has created the need for sound technologies, tools and standards for its archival and management. The techniques should be conducive to preservation and ease of access of the digital artefacts. Digital information is not enduring. It is fragile unlike information on traditional media such as paper. It is vulnerable to corruption, to alteration without detection and to obsolescence of storage and access technologies. At the same time, huge volumes of information in digital form can be stored more compactly and accessed with greater speed and ease than on paper. Changes in technology cause the media and the technical format of digital material to become unusable over relatively short timeframes. Some types of digital material like multimedia are closely linked to the hardware and software technologies that created them and cannot be accessed or managed outside these proprietary environments.

As more and more information is either ‘born digital’ or digitized from its analog form, a special focus is required to archive and manage it efficiently and effectively so that it is available over longer periods of time and is not lost to posterity. Keeping digital resources usable in the long term requires conscious effort and continual investment.

Digitization offers advantages like easy accessibility and retrieval, protection of data authenticity, data security besides the possibility to save and store in multiple formats. Once data is digitized, processing of files easier and faster compared to manual system. Furthermore, digitization enables access at multiple points at the same time, thus reducing delay in processing files.

3.1 Technical Standards - Scanning Mode

Scanning can be undertaken in the following ways:

3.1.1 Black & White Scan Mode: A *bitonal image* is represented by pixels consisting of 1 bit each, which can represent two tones (typically black and white), using the values 0 for black and 1 for white or vice versa. Speed is faster, but some information is lost in case of document is not in good quality.

3.1.2 Grey Scan Mode : A *grayscale image* is composed of pixels represented by multiple bits of information, typically ranging from 2 to 8 bits or more. At 8 bits, 256 (2^8) different tones can be assigned to each pixel. For obtaining best results in case of scanning of images is concerned, and also brightness is adjusted automatically.

3.1.3 Color Scan Mode: A *color image* is typically represented by a bit depth ranging from 8 to 24 or higher. With a 24-bit image, the bits are often divided into three groupings: 8 for red, 8 for green, and 8 for blue. Combinations of those bits are used to represent other colors. A 24-bit image offers 16.7 million (2^{24}) color values. It is best suited scan mode, however, file sizes becomes too heavy.

Therefore, it is suggested that greyscale mode is to be used for scanning of documents.

3.1.4 Best Practices for Digital Imaging for Textual Documents

Document Type	Resolution	File Format	Additional Note
Clean, high contrast documents	200 dpi bitonal	PDF/A*	
Documents with handwritten notes/markings, low contrast, half-tone illustrations, photographs or poor legibility	300 dpi grayscale	PDF/A	
Documents and items where color is important for accurate representation	300 dpi 24-bit colour	PDF/A*	

*As per Technical Standards for Interoperability Framework for e-Governance in India Guidelines (2012)

3.2 File Formats

The digitally scanned images are stored in a file as a bit-mapped page image. It can be formatted and tagged in dozens of different formats to facilitate easy storage and retrieval depending upon the scanner and its software. PDF-A searchable should be saved for the purpose of display and access.

3.3 Dublin Core Metadata Element Set (DCMES)

The Dublin Core is a set 15 data elements defined in the Dublin Core Metadata Element Sets (DCMES). One of the main purposes of defining DCMES is simple, broadly applicable metadata scheme that could be implemented by organizations and projects of all sizes. It has gained wide spread acceptance largely because of its simplicity, flexibility and applicability in any format. The List of metadata for each collection and instructions for filling up metadata are given in the Annexure.

3.4 File Naming Conventions

A file naming convention is a set of agreed-upon rules used to assign identifiers to digital objects in a collection. A good file naming convention ensures consistency and uniqueness and should be established before scanning begins. Please use the following guidelines to develop a file naming convention:

- Use 8 or fewer characters in the file name, whenever possible.
- Use a 3 character file extension (i.e pdf).
- Use only alpha-numeric characters, except for dashes, - , and underscores
- Do not use special characters, such as, . \ / : * ? " < > |, except for dashes or underscores. As These characters are often reserved for use by the operating system.
- All letters should be lower case.
- Do not use spaces in the file name.

3.5 Creation of Metadata and Uploading of Digital Files

LSS shall finalize the metadata parameters for each collection identified above for publishing in the Digital Library.

Annexure - 1

Sample List of Collections & Metadata tags**Trade Agreement**

1. Title
2. Date of Agreement
3. Ministry
4. Keywords
5. Type
6. PDF File Name

Books

1. Title
2. Year of Publication
3. Author/Editor
4. Publisher
5. Keywords
6. Language
7. Type
8. PDF File Name

Parliamentary Committees

1. Report No.
2. Title
3. Date of Presentation
4. Ministry
5. Keywords
6. Type
7. PDF File Name

Committees and Commissions

1. Title
2. Date of Presentation
3. Ministry
4. Keywords
5. Type
6. PDF File Name

Bills

1. Title
2. Name of Member, if Private Member's Bill
3. Date of Introduction
4. Date of Passing in Lok Sabha
5. Date of Passing in Rajya Sabha
6. Date of Ascent
7. Status
8. Keywords
9. Type
10. PDF File Name

Acts

1. Title
2. Date of Notification
3. Keywords
4. Type
5. PDF File Name

Ordinance

1. Title
2. Date of Notification
3. Keywords
4. Type
5. PDF File Name

White Papers

1. Title
2. Year
3. Ministry
4. Keywords
5. Type
6. PDF File Name

Guidelines for creating Meta-Data

- 1) All Dates should be in dd-mm-yyyy format

e.g. 02-09-2012

- 2) Multiple entries in the record should be separated by semicolon (;)
For example, in “keywords” field, multiple keywords should be entered as follow:

Budget;Annual Report;Year Book

- 3) First letter of each word in the title should be a capital letter. All Abbreviations should be entered in capital letter.

- 4) Author name should be entered as follows

e.g. Mishra SK

No space, no dot should be in between the SK

Multiple entries of author should be separated by semicolon (;)

e.g. Sinha RK;Bhatia KK;Mishra SK

- 5) Ministry Name : A standard list of all the Ministries has been developed by LSS, all the Ministry names should be entered as per the enclosed list.