# The Big Data Era: The role of Sri Lankan Parliamentary Library and Research Services

By

P. H. N. Premasiri, Librarian – Parliament of Sri Lanka

Email: nimal_p@parliament.lk; Premasiri.nimal@gmail.com

Tel:+94112777483, +94714089798


S. L. Siyath Ahemed , Assistant Librarian - Parliament of Sri Lanka

Email: siyath_a@parliament.lk

Tel: +94 11 2777231, Mob: +94 714902481

*Abstract*

*Libraries play an important role storing and managing information assets which are large in amount of data and those data in library need to be transformed into information or knowledge which then be used by the Parliament and researchers on legislature. Parliamentary Librarians might need to understand how to transform, analyze, and present data in order to facilitate knowledge creation. And they should know how to make big datasets more useful, visible and accessible to their client.*

*Big data technologies make it easier to work with large datasets, link different datasets, detect patterns in real time, predict outcomes, undertake dynamic risk scoring and test hypotheses.*

*This Paper briefly describes how and why big data is important to Parliamentary Library. And it explores the possibilities and its implications on librarians and enumerates some of the Big Data projects in Sri Lankan Parliament Library.*

*Sri Lankan Parliament library is also, in the era of moving towards emerging new technologies and offering reliable online resources and services to its users and to its institution.*

## 1. Introduction

Libraries play an important role in storing and managing information assets which are large in amount of data and those data in library need to be transformed into information or knowledge which then be used by Parliaments and researchers on legislature. Parliamentary Librarians might need to understand how to transform, analyze, and present data in order to facilitate knowledge creation. And they should know how to make big datasets more useful, visible and accessible to their client.

Sri Lankan Parliament Library's collection was originally for Members of Parliament and government institutions or other special researchers to find necessary information they need. However, this data becomes so large in size and the format is so complicated that might affect the efficient use. Most of the documents have not been used for data mining or big data technology. In the present context of the rapid development and the advancement of information technology, most of the Parliament Libraries all over the world are now involved in building up of digital repositories to present the best of their services to their clientele. Being conscious of this timely development in other parliament libraries, Parliament Library of Sri Lanka too envisaged in a project to set up its own digital library with data management system under the support and assistance of the Inter Parliamentary Union (IPU) under its parliamentary library development project.

The IPU sent Dr. Edmund Balnaves, the Consultant who supportively identified approximately 10 million pages of physical as well as more than 13000 borne electronic data documents in this library as unstructured data. Most of these records remain isolated from internet, requiring a study for solution as to how this data could be effectively exposed for better use with Big Data technologies.

At this juncture, Sri Lankan Parliament Library has initiated a data management system by using Dspace software on the recommendation by Dr. Edmund Balnaves in February 2017.

Accordingly Parliament Library of Sri Lanka started a project to establish a digital library covering existing born digital resources and the scanning of the valuable historical collection of documents in the Parliamentary Library.

A detailed implementation plan for a digital library resulted from an initial mission by the IPU consultant with a recommendation of implementation in two phases:

First, establishment of a digital library and import of existing library digital databases, and second the scanning of the historical collection.

Inter Parliamentary union (IPU) agreed to fund the first phase of this project. The objective of the first phase is to establish a digital library platform with data management system which:

1.  Enhances information available to members, including delivery of new search services such as scanned news clippings.

2.  Consolidates a number of existing library resources (including the catalogue) currently hosted on platforms with limited access for members.

The first phase of the digitization project has already completed and at presents the library is functioning on a very attractive digital library platform. This mission marked the completion of Phase 1. It delivered:

1.  Migration of important databases management by the library to DSpace - including the library catalogue and the implementation of a circulation system for the library.

2.  Development of a platform for electronic documents, including workflows for document submission and automated harvesting from other born-digital resources.

3.  Training of library staff in technical management of DSpace configuration and submission of documents to DSpace.

4.  Installation of DSpace in the target server - Undertaken by the IT department supervised by the DSpace expert.

5.  Defining metadata requirements for the repository and build document workflows for all document types.

6.  Developing and implementing conversion scripts to transfer metadata from WinISIS to DSpace preparatory to document submission of existing content.

7.  Establish backup procedures and Disaster Recovery procedures

The library now has a substantial; Metadata indexed comprising > 50,000 documents.

## 2.  Literature Review: What is Big Data?

Big Data is one of the most popular terms in this era. Big data is data that exceeds the processing capacity of conventional database systems. Everything is data. Everything we do, say, or write creates datasets. The data is too big, moves too fast, or doesn't fit the structures of your database architectures. To gain value from this data, we must choose an alternative way to process it.

There are no clear cut definitions of what constitutes big data. But the term "Big Data" describes that dataset with so large and very difficult to manage with traditional database tools. Big Data refers to huge volumes of Data that cannot be structured and processed using the traditional approach within the given time frame.  According to Wikipedia, 'Big data' refers to data sets whose size is beyond the ability of traditional software tools for capturing, managing, and processing the data.   In order to classify big data there is lots of mesh conceptions referring the term "Big Data". Use the term Big Data referring usually GB, TB, PB, EB or anything larger then in size. This term does not define pliantly. Even the small data can be Big Data depending on codex to be used.  If you try to attach a document that is in 100 MB in size to email but cannot be able to do so. As the email system would not provide facility to make attachment of the size. In this situation 100MB size of attachment with respect to emails can be refer to as Big Data.

Chunning Wang and others (2016) describes that Big Data as innovative techniques and technologies to capture, store, distribute, manage and analyze datasets that traditional data management methods are unable to handle.

Hoy (2015) noted "in recent decades has the ability to capture and analyze data really off; as more and more aspects of daily life are connected to computers and the Internet, this formerly ephemeral data is now being captured, stored, and analyzed, often with surprising results.

Hoy also points out Ed Dumbill's definition that "Big Data is Data that exceeds the processing capacity of conventional database system. The data is too big, moves too fast, or doesn't fit the structures of your database architectures. To gain values from this data, you must choose an alternative way to process it.

Although there are many definitions, Laney (2001) only has given a widely-accepted definition that the Big Data often characterized with using three dimensions or "three V's" of big data: volume, velocity, and variety.

- Volume refers to the sheer amount of data being created. It says how big of data could be classified as big data. Therefore, the size might vary based on the disciplines. Traditional software usually can handle megabyte and kilobyte sized data sets, while big data tools should be able to handle terabyte and petabyte sized data sets.
- Velocity refers to the speed with which data is being created.
- Variety refers both to the types of data being gathered and to the lack of uniform structure in the data. It makes big data sets harder to organize and analyze.

In the Knowledge Based Society, both enterprise application data and machine generated data, known as Big Data. Ward and Barker (2013) - Big data is a phrase describing the storage and analysis of large and or complex data sets using a series of modern technologies. The big data technology enables us to acquire deeper, more valuable insights from the data and make more timely decisions.

It has to be asked often that is library data really big data? If you consider just the metadata representing the collection of printed and electronic works held by libraries, it really cannot

be considered big data in its current meaning. Even when you consider the full-text of those works, the data management does not require "big data" techniques. Hessman, T. (2013) says that if consider the static collection in libraries, it might be harder for us to relate it to big data. In addition the database management system should be enough to store and to process library data, therefore, based on the definition of big data, there is no need for big data technology such as distributed systems to analyze the data in library.

In conclusion, Chunning Wang and others (2016) say library data could be treated as big data without any doubt due to its property of large volume; high velocity and obvious variety. In addition, library data are often less organized, lacking of standards and unique formats.

Big data technologies make it easier to work with large datasets, link different datasets, detect patterns in real time, predict outcomes, undertake dynamic risk scoring and test hypotheses.

This Paper briefly describes how and why big data is important to Parliamentary Library of Sri Lanka. And. it also explores the possibilities and its implications on providing services to parliamentarians and more over it elaborates some of the Big Data projects in Sri Lankan Parliament Library.

Sri Lankan Parliament library is also, in the era of moving towards emerging Information technologies and offering reliable online resources and services to its users and to its institution. It is observed that digitization of documentary sources in the Parliament Library of Sri Lanka for the objective of conserving and preserving, has been a long felt requirement.

At a time when most of the parliament libraries the world over are adopting such new emerging technologies as Big data, Sri Lankan Parliament Library and Research Services has embarked on a digital library project to adopt the new technological trends and to cater to the legislators day to day information requirements.

## 3. Background and Problem statement

According to Byron Heirdron (2013) libraries must become curators of digital data in order to adhere to their core mission: to protect and disseminate information. The Parliament of Sri

Lanka encounters a number of risks and problems in both preservation of and access to important current and historical resources.

The library has a respected and unique cultural record of Sri Lanka in its print collection. The older items are fragile and in need of conservation. Most of the materials in the collection are of archival value and they have been neglected and undervalued since the library is not in a position to provide physical access to them due to their decrepit nature. Many documents are very old (from 1830 onward) and in a fragile condition and not suitable for general access. Especially this collection includes parliamentary documents that would be of considerable value for Members of Parliament but which are largely only in print format.

Many publications catalogued and shelved in the main collection are confined to their original locations itself and there is no way of providing access to contents and Members of Parliament are unaware of the availability of many historical documents in the collection.

The library has a collection of 13,000+ documents of parliamentary record, as well as a substantial book and newspaper collection from 1924. They also have a number of visual materials such as maps and photographs. The publications are largely not on acid-free paper and are gradually showing greater deterioration. They will be increasingly inaccessible for use by Members and are at risk of becoming difficult to scan and preserve. It is pressing to proceed with digitization of at least that component of the collection older than 50 years. There is an opportunity, and timing is becoming pressing, to digitize these documents.

While the Parliament of Sri Lanka has made significant progress since 2005 in creating all documents in digital form, there is only partial coverage of documents were subject to this project most of the documents are scattered in every division.

## 4. Solutions

Parliament Libraries are uniquely suited to working with big data. Parliament libraries have a long tradition of being early technology adopters, and big data should be no exception. Thanks to ICTA, one of the Sri Lankan Government's Agency which works for promotion and advancement of Information Technology policy in Sri Lanka spurring their department to

open websites. Under this mandate, every department or institution must host a www.[institution].gov.lk URL, listing the available datasets in both human-readable and machine-readable formats. Unfortunately, these data sets sometime removed from site or the site has expired to retrieve data from those portals. Parliament Library is well-positioned to help Members of Parliament to understand on how and where to find these data sets and to preserve them for future use.

Another way Sri Lankan Parliament Library has involved with big data is by working within its institution to help with research data information. Parliament library helps for researchers for finding very specific data for their research and providing preservation option of their research, as well as the potential guide for sharing their data.

Recently, in February 2017, Parliament library of Sri Lanka launched a data management system powered by open source software Dspace , to provide access to the immense array of data available. The Dspace system is integral to the success of the Open Data Policy as it goes beyond organization and access to harvest data from Parliament existing portal and a system to increase availability of whole data created by parliament of Sri Lanka. At the time of writing, there are over 20000 datasets made available on database and this number will only continue to expand. In fact, users of parliament library can directly request a specific dataset to be released.

Almost all the parliamentary documents generated through the legislative process are accessible to, in the parliamentary website; www.parliament.lk . The Data in these documents are also accessible to, through the intranet of the parliament.  In addition, many documents related parliamentary process are downloaded from various governmental, statuary institutions and non-governmental bodies and the bulk of data in documents uploaded to library's data management system powered by Dspace software. There are also data extracted from collections of CDs which are presented to the Parliament by all ministries, departments and public enterprises.  The information relevant to the above data carries in parliament publications also digitally available in the library's data management system. Apart from these sources there is information relevant to the above subject are also published in the digital media. In support of research activities these data are readily made available in our library data management system.

The main objective of this timely important problem is to provide the Data Management System with facilities to solve the issues encountered in finding solutions for many issues on the use of big data.

## 5. Conclusion

While the capabilities of Big Data are just being realized and Parliament Libraries have traditionally served as stewards of information access, research instructors, and privacy advocates for Members of Parliament as well as for government and general public.

Within the recent tidal wave of available data resources, the parliament libraries's role remains consistent and even has the potential to expand. To accommodate for the rising data demand, Parliament libraries are adapting to change their service activities along with the fundamental skill set for managing Big Data.

In order to better understand the environment of Big Data Parliament Libraries should be responsible for acquiring hardware and software for storing and analyzing big data which would be of immense use of parliamentary libraries.

## 6. References

Doug Laney.(February 2001) "*3D Data Management: Controlling Data Volume, Velocity and Variety*.". META Group Research Note 6.

Bryan Heirdorn. (2013) "*The Emerging Role of Libraries in Data Curation and E-Science*," Journal of Library Administration 51 issue 7-8: 662-667.

Lauren Reinhalter & Rachel J. Wittmann. (2014) *The Library: Big Data's Boomtown*. The Serials Librarian: From the Printed Page to the Digital Age, 67:4, 363-372

Michael Teets and Matthew Goldner. (2013) *Libraries' Role in Curating and Exposing Big Data: Future Internet*, 5, 429-438

Chunning Wang, Shaochun Xu, Lichao Chen, Xuhui Chen. (*2016) Exposing library data with big data technology: A review.* Okayama, Japan, *IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, vol. 00, no. , pp. 1-6, 2016

Hessman, T. (2013) *Putting big data to work.* Ind week 2013, 262, 14-18