DESIDOC Journal of Library & Information Technology, Vol. 40, No. 5, Sept 2020, pp. 272-276, DOI : 10.14429/djlit.40.5.16330

Exploring Archives Space: An Open Source Solution for Digital Archiving

Mayukh Sarkar[#,*] and Sruti Biswas[$]

[#]*Aligarh Muslim University, Aligarh - 202 001, India*
[$]*Indian Institute of Technology (Indian School of Mines), Dhanbad - 826 004, India*
[*]*E-mail: msarkar1990@gmail.com*

ABSTRACT

The advent of digital and networking technologies has begun to embrace the genesis of the next-generation digital archive. The inclusion of cross-domain objects like manuscript documents, audio and video recordings, photographs, paintings, sculptures and other digitised cultural heritage materials increases the complexity of digital archiving in terms of preservation, collection, and discovery of these resources. Introducing a high definition information retrieval system to exhibit the library and museum's digital resources to a maximum number of users in an open-access environment can satisfy the S. R. Ranganathan's fourth law – save the time of reader as well as the staffs. Nevertheless, from the perspective of acquiring an advanced OPAC view (web-scale discovery interface) with index-based searching, metadata harvesting, and accessing the physical as well as digital holdings is always a better option for Archival Collections Management System (ACMS). This paper illustrates the fundamental notions and applications of ArchivesSpace, a useful open-source digital archiving toolkit of the contemporary world and analyses its relevance in digital language archiving.

Keywords: ArchivesSpace; Archival collections management system; Digital archive; Digital language archive; Language documentation; Resource discovery.

1. INTRODUCTION

The 'archives' are the primary sources of information that holds the key to preserve the socio-cultural knowledge of human civilisation. The practice of archiving official records, museum and heritage artefacts started a long time ago (3000 B.C.) and in the 17th century, it gets the essence of perfection in the form of modern archival science emanated from the French revolution. The phenomena of 'language documentation' superimposed with the notion of archiving process as sands of time flows. Though 'language' itself has multiple heterogeneous directions explained by many linguistic and scientific communities, in terms of documentation in an elementary and compact fashion denoted as "systematically recorded representations of both spoken and written forms of a language in their appropriate socio-cultural context.[5]" The advent of neoteric digital technology moulded the conventional approaches of achieving and language documentation techniques and leads to intensifying the documentation of endangered languages as well as other heritage documents, making language-primary data long-lasting and increasing its visibility to the users. Another great impact of this digital revolution it has reconstructing the traditional museum and library collection system by implementing a stable content management system (CMS), allowing forming a global network of archival community and sharing the resources in a common global platform. With this approach of cross-domain collaboration as well as building a cross-cultural knowledge society can satisfy the high thoughts of globalisation. There are several existing CMS's and Institutional Repository Systems (IRS) available to date, capable of maintaining and disseminating their resources. However, in terms of preserving the unique archival materials and documented languages (classified according to their disciplines) and accessing them across the archive communities, for that, a more robust CMS is required. Here lie the limitations of DSpace/EPrints like IR Systems or Drupal/Joomla like CMS's where provisions of importing descriptive, cultural metadata and other advance discovery features are limited, and that is the motivation behind developing another system for museums. This type of systems known by the Archival Collections Management System (ACMS) defined as the modular system that enables the staffs and users to make, alter, modify and publish the digital archival contents and maintaining them with the particular set of data formats, principles, workflows and ethics (related to archives and language documentation) within a network so other community members can access them globally. ArchivesSpace is an open-source, web-based ACMS toolkit licensed under Educational Community License, version 2.0. It is a cross-platform supported, written in java, and uses MySQL as a database can manage the archival information smoothly and represents enhanced search results. ArchivesSpace empowers the five major types of documentation functionalities through the proper representation of metadata (media and textual),

Received : 16 September 2020, Revised : 20 September 2020
Accepted : 25 September 2020, Online published : 04 November 2020

data migration, facilitating link data features with JSON-LD, protecting user rights, and finally, distribution of published outputs in different formats to a range of users. The research attempts to explore the insights of ArchivesSpace as a useful toolkit for digital archiving and language documentation.

2. LITERATURE REVIEW

Archives and language documentation is always a challenging task not only from the technical aspects but also embedding with ethical and socio-political issues. Seifart[1] et al. in this context discussed the development of regional language archive, and they addressed several problems faced by three South American countries (Peru, Argentina and Brazil) which lead to establishing networked initiatives for indigenous multilingual language documentation in Latin America. The School of Oriental and African Studies (SOAS), University of London implemented an archiving repository called Endangered Languages Archive (ELAR) to navigate the interactivity between linguists and digital archives. ELER has a robust framework (content analysis of archive queries, archive format guidelines, and services) for archiving documents, including training courses for Documentation Programme (ELDP). It is becoming essential to figure out the diverse issues related to documenting languages. After understanding its distinct nature and products, a unified effort of language documentation specialists and archivist as a team will answer many questions as indicated by Nathan[2].

The case study of the University of Nevada, Las Vegas libraries showed that since 2014 the library started using ArchivesSpace as their archival repository and according to Shein, Ou, Irwin and Lemus[3], sooner they initiated to develop local codes to extend the functionality of the application. The initiative helps to develop some new plugins (UNLV Spawn, MARCXML Exporter) for creating and customising the collection records, a python script (Multi Marc Exporter) for batch exporting MARCXML records, and a MARC record editor (Connexion macros) for reformatting fields, inserting, deleting values in MARC record. Additionally, they have also adopted three plugins (UNLV Custom Reports Plugin, LCNAF Plugin, and Overlay Plugin) for exporting reports, importing authority name-subjects, cleaning up the duplicate .csv and .ead files, as well as UNLV EAD Export Plugin and a stylesheet to auto-generating the customised PDFs. Through surveying 103 member institutions of the ArchivesSpace community and analysing the quantitative facts, Toov and Wick[4] tried to portray the achievements and future implication of this system under the headings of ten major categories. The significant findings of the study specified that post-implementation experience of ArchivesSpace was very satisfactory to most of the respondents compared to the previous system. In terms of ArchivesSpace functionality, the Accessions helped to standardise the existing process; the system allows to create individual repositories for different units of an institution, moderate respondents (30%) occasionally uses Digital Object functionality, 72% do not utilise Subject/Agents, location functionality.

3. OBJECTIVES OF THE STUDY

The research attempts to fulfil the following objectives:

- To introduce ArchivesSpace as robust ACMS toolkit for digital archiving.
- To describe the development, system architecture, and salient features of archives space.
- To identify the advantages of archives space in digital language archiving.

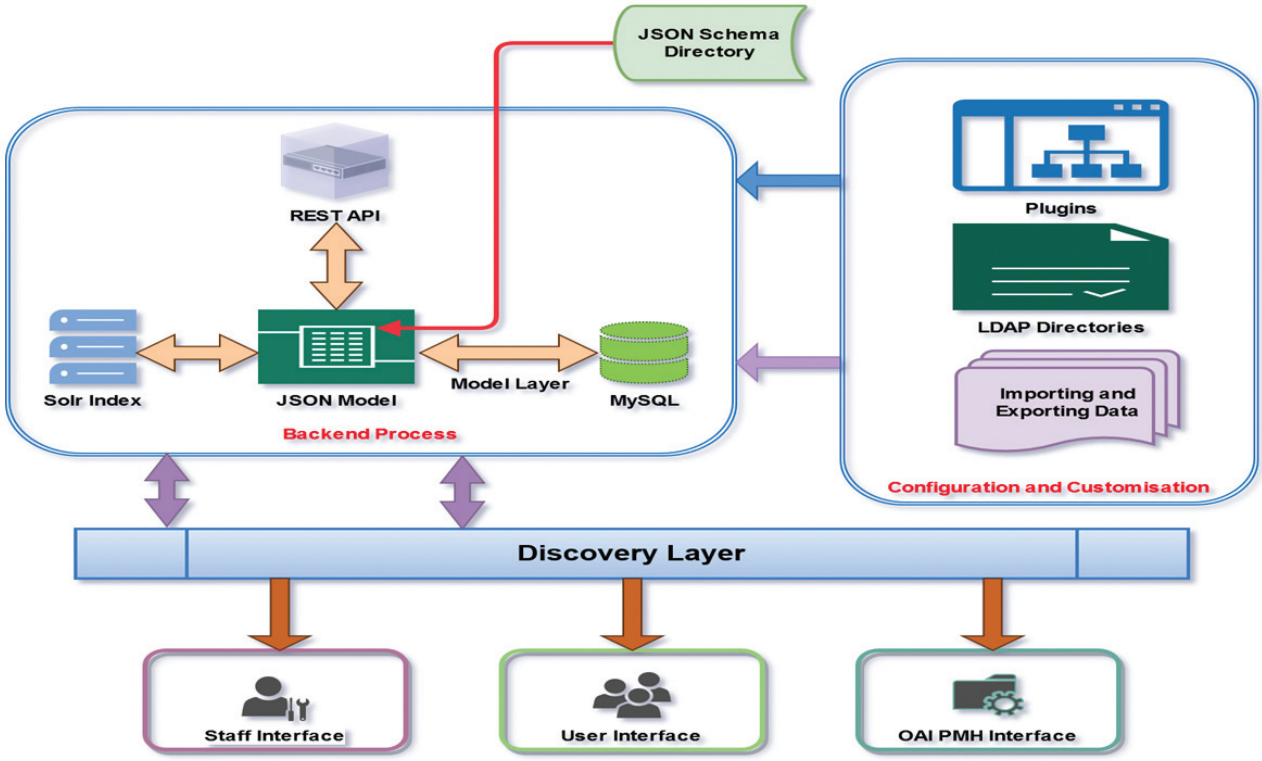

Figure 1. system architecture of ArchivesSpace.

4. METHODOLOGY

This qualitative study concerning the ArchivesSpace ACMS offers insight into the toolkit driven by descriptive research methodology. The paper investigates its developing history and draws the system architecture by understanding its workflow, followed by a range of literature reviews. The study further described the fundamental features of ArchiesSpace by exploring its modules and finally concludes correlating and examining the advantages of it with the notion of digital language archiving.

5. DEVELOPMENT AND SYSTEM ARCHITECTURE

Archives space is essentially the successor to 'Archivists Toolkit' (AT) developed by the Andrew Mellon Foundation and 'Archon' of the University of Illinois Urbana-Champaign. Both systems look forward to the vision of creating a sustainable ecosystem in 2009, and the responsible bodies (New York University, University of Illinois Urbana-Champaign, University of California San Diego, and Andrew W. Mellon Foundation) have agreed to merge two applications. As a result, the first beta version of ArchivesSpace released in September 2013 under the organisational home of LYRASIS. It could refer in the context that LYRASIS has also started working with DuraSpace (primarily known for solutions such as DSpace, Fedora and VIVO) since 2019 to broaden the impact and extend its reach to global communities. Besides that, the latest version of Archives space 2.7.1 (Frebruary 2020) powered by enhanced features and infrastructure improvements with 129 schemas and three new database integrations.

The system architecture of ArchivesSpace (Fig. 1) grounded in JSONModel, concerned with managing different archival record type through defining the properties of class instances and validation rules for each type of record. JSON schema directory has listed the schema definitions based on which JSONModel captures the validation rules for ArchivesSpace. The RESTful API consumes and produces JSONModel instances using the program rest.db. On the other hand, the model layer connects the relational database (MySQL) to JSONModel instances. The full-text searching made possible through its Solr index. Besides, the main. rb program is responsible for triggering all the controllers and models including JSON to load all record schemas from the file system, connecting to the database and handling the system startup in this backend process. So, it is needless to say that the ArchivesSpace blackened process support CRUD operations (Create, Read, Update, and Delete). The second powerful part of ArchivesSpace is its area of configuration and customisation, allows integration to numerous plugins (aspace_ yale_accessions, extended_advanced_search, material_types, payments_module), importing and exporting data (CSV, EAD, XSL) and authenticating against LDAP (Lightweight Directory Access Protocol) directories. The architecture finalised with the accessible interface of ArchivesSpace, represented through the discovery layer that a user or staff accessed for their purpose. Apart from the general use of public user interface (PUI) and staff interface, there is a separate OAI-PMH interface permitting other systems to harvest its records using the standard requests. Though there are 25 plugins available at present, there is an open provision to the communities for developing a new plugin or modifying the existing one as per their requirements.

6. FUNDAMENTAL FEATURES

Unlike the web-scale discovery and IRS systems, ArchivesSpace enriched with the discovery features of an ample search space with Relevance ranking, Faceted search option, Refinement of actual queries, Flexible searching and sorting result by descriptive date format, Integration of plugins and other archival systems, Harvesting of metadata and report generation, Local and remote access. However, the unique features that make it sturdy enough for managing the archival collections in the digital environment categorised under the following headings:

6.1 Web Based System

ArchivesSpace developed as a web-based system that provides hassle-free installation and uses the opportunity to the users. Most of the open-source discovery or IRS systems consist of a complicated method of installation (command-based) to activate the service in a domain. Compared to those systems, installation of ArchivesSpace is mainly web-based and run with a single command after downloading the required version (.zip file) available at GitHub[6] followed by its technical documentation.

6.2 Specially Designed Modules

ArchivesSpace facilitates the ACMS function through its four main modules (Archival Object, Accessions, Resource, and Digital object), and six supporting modules (Agent, Repository, Staff User, Location, Subject, and Event). Apart from these modules, there are few sub-record definitions (Date, Extent, External Document, Rights and Collection Management) that connect the hierarchy of records like Accessions, Resource, or Digital object in a one-to-one relationship. These modules are defined below with their functionalities:

6.2.1 Main Modules

The main modules deal with the core functions of the ArchivesSpace System so that staffs and users get involved with the system and start interacting with it by creating and editing records.

6.2.1.1 Archival Object

The archival object module defines principles, rules for the compilation of ArchivesSpace items (accessions, resources, digital objects, and their components) aggregated into a single archival object record further transcribed and enlarged by the particular object type warrant. Additionally, it has also described the rules for linking and validating the objects. The significant properties of archival object record denoted by archival object type, identifier, title, publish, restrictions apply, and linked repository.

6.2.1.2 Accession

This module helps to record accession information about the collection submitted to the archivists. It has two required

fields – Unique accession identifier (consisting of four parts) and Accession date. The provisions for linking other types of records with accession records are also available according to one-to-many relation (extent, date, rights, location, deaccession external document, collection management sub-records), and many-to-many relation (resume, name, subject records).

6.2.1.3 Resource

It provides the finding aid (resource record) described the collection which could display in the PUI. There is a long list of fields available in resource record in parent level, component level with sub-records and linking options. Users are empowered to create, edit and delete the resource record.

6.2.1.4 Digital Objects

This module stores the information (digital content files with their web location and metadata) about a digital object or a collection of these objects (both simple and complex architecture). Along with the digital objects, surrogates, or born-digital material, this module also accommodates other digital library tasks like describing the audio, visual image, and ETD collections.

6.2.2 Supporting Modules

The supporting modules create and connect the supporting records with the main module records and extend the functionality by developing a many-to-one relationship.

6.2.2.1 Agent

The agents refer to the persons, families, or corporate bodies that perform the role of a creator, source, right owner, etc. of the resource. The agent record describes an agent through Agent types, System control data, Linked context records (accession record, resource record, rights record), and Linked sub-records (name forms, name contact, external documents) that makes the resource more productive and helps to identify the agent entities in the discovery system. It is compliant with ISAAR (CPF) standard, PREMIS Agent alignment and allows export-import of MADS and EAC encoded agent records.

6.2.2.2 Repository

The repository module identifies, links information regarding the repository-level records of the resources, and separates one repository records, data values to another repository stored inside the same database.

6.2.2.3 Staff User

The functionality of this module is to record the information of staff/user information for administrators to control their access and grants permissions on the staff interface.

6.2.2.4 Location

Location module designed to describes the shelving location of archival materials. Here location module is restricted to track only the physical materials, the location of web materials (URIs) defined in Digital object of main modules.

6.2.2.5 Subject

The subject module specially designed for assigning topical terms (subject heading) to the accessions, resources or digital objects at any level of description. Generally, the subject record has five parts such as 'Required' fields (primary subject heading), 'Optional' fields (supplemental information), 'Linked Records', 'Linked sub-records' and 'System control data' (auto-generated). It can also export from DC, MODS, VRA, EAD, and MARCXML records.

6.2.2.6 Event

Events represent a particular course of action (archival and collection management workflow) linked with the one or more agents and archival objects within a specific date-time period.

6.3 Other System Integrations

There are provisions for other system integrations like Digital Preservation and Archive Systems (Archivematica, Preservica, Archive-It), Content Publication Systems (Drupal, CollectionSpace, Aviary), Reference and Reading Room Systems (Circa, ExLibris Alma), Import/Export Integrations (EADChecker, ArchivesSpace Preprocessor), and more.

6.4 Enhanced Search Result

ArchivesSpace processed a federated search although the repositories, collections, digital materials, accessions, subject, names and classifications of the application and presented result in a faceted manner such as title, subject, creator etc. The search terms could further filter by setting its limit (all records, collections, digital materials, etc.), keyword, title, creator, subject, notes, identifier and year range. The user has complete access to the indexed data with Solr search and index engine.

7. ARCHIVES SPACE IN DIGITAL LANGUAGE ARCHIVING

Looking down the memory lane, the digital language archiving started evolving as a discipline for past 10 to 15 years taking feeds from the concept of language documentation, digitisation, and digital preservation. The development continues with the help of digital tools (e.g. Lexique Pro, SayMore, Mukurtu Mobile, Audacity) and standards for archiving language data including different formats (audio, video, manuscript and many more). The fundamental objective is to revitalise, re-learn, maintain, and access these languages as well as the cultural identity for the people as they evolve every day and the original group of people who communicate and have knowledge of specific dialects has departed[10]. Apart from that, some people deal with the various forms of information regarding languages, such as oral history, music, and folk cultures, pervasively related to the notion of digital language archive. Now coming back to the ArchivesSpace, it has some already enabled features and future provisions that could exclusively contribute to the digital language archiving:

- ArchivesSpace appears to be a new jargon in the field of linguistic training where the institutions commit to depositing the intellectual products, training programs,

lectures, workshop materials of the linguists, field researchers that processed for the long term preservation of under-documented languages.

- The multi-level record system module provides sophisticated and in-depth information (descriptive metadata) about language resources.
- It creates a virtual bridge between the heritage language community and the native speakers in the same space where they gather to fulfil different aims such as educational, cultural, intellectual curiosity, or explore their own identity.
- The discovery features extend the limits of modern developments and open up new possibilities to a new dimension by integrating the language archiving networks, databases and standards such as DELAMAN, ELAR, Pangloss Collection, DOBES Portal, Living Tongues Initiative, and International Standards for Language Engineering Metadata Initiative (IMDI) etc.
- As the archived language data disseminate to a broader community, it may help to realise the adequate and unique documentation of languages.
- Finally, concerning the legal and ethical issues, archival institutions have to contend with aspects such as privacy, restrictions and access to the archival materials. It has always been a challenge to cope up with the international standards and ethics with the institutional policies. Institutions can maintain the privacy policy of these materials through the ArchivesSpace by only providing permission to access the metadata of the resources.

8. CONCLUSIONS

According to the census report[7] of 2011, there are more than 19500 languages spoken in India. Till now different government institutions (Central Institute of Indian Languages, Indira Gandhi National Centre for the Arts), NGO's (Bhasa Research and Publication Centre, Gujarat), linguists, activists, archives, museums, and tribal libraries, have significant works on language documentation with particular focus on endangered languages. Nevertheless, it is unfortunate, UNESCO[8] has classified 197 languages are in extremis situation in India. In another survey study by Devy[9], it estimated that currently, the number of endangered languages is approximately 600 and close to 250 languages have already been extinct for the last 60 years. This scenario requires faster progress in the process of documentation of these languages and other related cultural resources and, after that, ArchivesSpace can make it accessible to the broader community. It is true that at present, very few institutions in India (e.g., National Centre for Biological Sciences, Bangalore) compared to the other western countries have implemented it for managing their archival resources. However, there will be a promising future waiting, when more institutions extract the facilities of ArchivesSpace to manage not only digital language objects but also other archival materials and increase the visibility and access to these resources to larger populations.

REFERENCES

1. Seifart, F.; Drude, S.; Franchetto, B.; Gasché, J.; Golluscio, L. & Manrique, E. Language documentation and archives in South America. *Lang. Doc. Conserv.*, 2008, **2**(1), 130-140.
2. Nathan, D. Digital archives: Essential elements in the workflow for endangered languages documentation and revitalisation. *Lang. Doc. Descr.*, 2008, **5**, 103-119.
3. Shein, C.; Ou, C.; Irwin, K. & Lemus, C. Open-source opens doors: A case study on extending ArchivesSpace code at UNLV libraries. *J. Contemp. Archival Stud.*, 2017, **4**(1), 1-16.
4. Toov, R. & Wick, A. Making it work – understanding and expanding the utility of ArchivesSpace. *J. Archival O.*, 2018, **14**(1/2), 35-54.
5. Furbee, N. Louanna. Language documentation: Theory and practice. In Language documentation: Practice and values, edited by Lenore A. Grenoble & N. Louanna Furbee. Amsterdam, 2010, 3-24.
6. GitHub. ArchivesSpace technical documentation. https://github.com/archivesspace/tech-docs (Accessed on 10 May 2020).
7. Registrar General, India. Census of India 2011. https://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf (Accessed on 4 June 2020).
8. Moseley, Christopher. Atlas of the world's languages in danger, 3rd ed. UNESCO Publishing, Paris, 2010. http://www.unesco.org/culture/en/endangeredlanguages/atlas (Accessed on 4 June 2020).
9. Devy, Ganesh Narayandas. Peoples linguistic survey of India (PLSI Series). Orient Blackswan, India, 2014-2019.
10. Holton, Gary; Berez, Andrea, & Camber, Wendy. Preserving and enhancing native language resources in tribal libraries, archives, and museums. *In* 2015 International Conference of Indigenous Archives, Libraries, and Museums.

CONTRIBUTORS

Mr Mayukh Sarkar is currently working as a Research Assistant at Aligarh Muslim University, India. He has completed the Master of Science degree in Library and Information Science (MS-LIS) from Documentation Research and Training Centre, Indian Statistical Institute, Bangalore, His research interest includes knowledge management, knowledge organisation, human information behaviour, resource discovery systems and services, digital libraries and archives.
Contribution in present study: He has designed the research theme, developed the system architecture, and described salient features of ArchivesSpace.

Ms Sruti Biswas currently working at the Central Library, Indian Institute of Technology (Indian School of Mines), Dhanbad, India. She completed her Master's in Library and Information Science from Rabindra Bharati University, Kolkata. Her research interest covers a wide range in the areas of Information economics, digital libraries, information literacy, and information management.
Contribution in present study: She reviewed the literature and tested the advantages of ArchivesSpace correlating with digital language archiving.