Libraries as Data Repositories for Smart City projects

S Bharanidharan *Vijay Krishnan* *G Krishnamoorthy*

Abstract

Libraries play a key role in the development of their communities. Information available in a library is valuable to the community for its development. Archives and records about a place or geographical community is a repository of the history of developments in that location. These can span various aspects of governance and administration from land use to water supply, from transportation to electricity, and others. From villages to cities, various records are collected and historical data is available. Indian customs and traditions could impact such records. Computerization of records has made them easily accessible.

Smart city implementations can involve data collection and analytics for the benefit of their communities. There are privacy concerns around data from IoT and other technology platforms. Libraries have contributed to smart city developments. In places where there is conflict around control or access to such data, libraries can offer data archival, governance and analytical services so that smart city project data can benefit the intended communities.

Keywords: Data Repository, Library, Library, Smart City, Urban Development

Libraries have historically been places where people from the community gather, learn, and conduct discussions and activities that interest and benefit them. Records spanning various aspects of life in the community were maintained by administrative bodies. Records over time got transferred to archives from which the history of various aspects of communities is written, and many libraries maintain archives. Computerization of records and archives has made accessible this rich history of various facets of communities, quickly and easily.

Among many definitions, Smart city is described as an urban area that uses different types of electronic Internet of things (IoT) sensors to collect data and then use these data to manage assets and resources efficiently, per Wikipedia (2019). They, generally, aim to improve the quality of life via better delivery of services to mostly urban communities. Data helps proactively identify the requirement for various services, with analyzing and forecasting using past data or getting live streaming data from sensors being adopted. In areas where IoT and sensor data collection is implemented, data privacy, access and governance issues arise. Since smart city data pertains to various departments of civic administration and is of interest to various stakeholders, the role of libraries in managing data repositories for smart city projects can be considered.

1. Library and Community Development

Libraries are open to all and create a common bond to all living in a community as per the American

Library Association (2019). This is true for public libraries where visitors need not be members in order to avail library services. Institution and academic libraries are restricted mostly to the organizations that they serve, but even they are open to access by the public in their communities in a limited manner. Even if the general public cannot access all services at such libraries, they are able to participate in public events and programs that are conducted. The role of the library as a community gathering place emerges repeatedly, and it is evident that libraries stand alone in many communities as a gathering place.

The main reason for visiting the library was to look for information on a subject as per surveys conducted in Toronto and Vancouver, and presented by Leckie and Hopkins (2012). Library collections and downloads can include books, periodicals, professional journals, travel materials, audio books, DVDs, videos, music and business resources. Research suggests that the search for relevant information and its subsequent use in productive activity may be an integral characteristic of the construction of contemporary public culture. The library is, thus, a key site of cultural consumption and production, as well as a facilitator of civil society.

2. Records and Archives of Communities

As per the State Archives (2004), in British ruled India, for revenue administration, the country was divided into four divisions with a Diwan Peishkar in charge of it taking on the responsibilities of a district collector and magistrate. Each division was divided into Taluks presided over by Tahsildars, and further into villages where Pravartikar or Adhikari were the officials responsible for the village, which is the smallest unit of administration in the country. They maintained various records about the village and that practice is in place to date.

Various records are maintained in, about and for the villages. Census records pertain to population, births, deaths, gender and demographics. Land records pertain to land ownership, land use, conversion of land use for other activities and related economic information. Police records pertain to crime, passport issue, background checks and other information related to law enforcement. Health records available with dispensaries and government hospitals at the village level can show incidences of diseases, mortality, and other related medical information. With regard to India, it is important to note that Indian customs and traditions can lead to data of one village being recorded in another. As per a study by Okabe and Surijit (2012), in most cases mothers go to their native villages or for institutional delivery at the nearest town, and births are registered outside the village. It concluded that village-level statistical databases today require that village-level birth registers cover those, whose place of usual residence is the village under consideration. Records over history are available in various archives, and in many libraries.

An archive is an accumulation of historical records or the physical place they are located as per Wikipedia (2019). Historians generally understand archives to be records that have been naturally and necessarily generated as a product of regular legal, commercial, administrative, or social activities. Archives are considered the factories and laboratories of historians. Along with private studies and public libraries, archives provide material with which to build history. Tendency to regard archives as neutral and unproblematic store of historical facts

without scrutinizing the decisions about selection, arrangement, preservation and retention taken by those responsible for the care of their contents over successive generations is a legacy of the historiographical developments of the nineteenth century per Walsham (2016) who also notes that contemporaries were beginning to apply the word archives to repositories that performed the function of preserving its written traces rather than simply safeguarding documentation of legal and business transactions.

Computerization of records and archives is making it possible to not only access the required information, but also to study historical changes while ascertaining possible root causes of the impact of policies, programs or other related developments. To give an example, from historical archives at district level and high resolution remote sensing datasets, it was possible to study large scale land transformations as demonstrated by Tian, Banger, Bo and Dadhwal (2014). Records management has developed into an area of study and practice, contributing to the accuracy and update of records relevant to civic authorities. Open data initiatives as exemplified by Data.Gov.In (2012), have gone a step further and published datasets, documents, services, tools and applications for public use to increase transparency in the functioning of the government and to open avenues for innovative use of such data.

Libraries can be involved in data collection for smart city across many types of data categories, while some data categories may not concern them. It is also possible for various libraries to concern themselves with different data categories that they work with, based on their specialization.

Table 1

Data categories for smart city implementations per Government of India Ministry of Housing and Urban Affairs (2019) with suitability of libraries to handle that data category

Sector	Library Suitability for Data Collection
Art and Culture	Yes
Commerce	No
Economy	Yes
Education	Yes
Environment and Forest	Yes
Finance	No
Food Yes	
Governance and Administration	No
Health and Family welfare	Yes
Home Affairs and Enforcement	No
Housing	Yes
Industries	No
Information and Broadcasting	Yes
Information and Communications	No
Infrastructure	No
Labour and Employment	Yes
Power and Energy	No
Science and Technology	Yes
Social Development	Yes
Statistics	Yes
Transport	No
Travel and Tourism	No
Urban	No
Water and Sanitation	Yes
Water Resources	Yes

Specialized sectors where government and market/ industry activity is high require institutional bodies to handle the data categories. In Table 1, the suitability has been arrived at based on the ongoing operation of a suitable institution of higher education in that particular sector, whereby its library can handle that data category. For example, Environment and Forest data can be handled by institutes where Forestry and/or Environment management/sciences are offered.

3. Smart City and Data

The term Smart City remains unclear to its specifics and therefore, open to many interpretations. In some definitions, smart city concepts integrate information and communication technology (ICT), and various devices connected to IoT networks in order to optimize the efficiency of civic operations and services that interface with citizens, as per Wikipedia (2019). Civic administrators interact with the community and city infrastructure/services to monitor events in the city and to calibrate the response of civic services to those events. ICT is used to enhance the quality, performance and interactivity of urban services, to reduce costs and resource consumption among others while smart city applications help manage urban flows and allow for real-time responses.

Real time responses are usually in response to real time data. Such data is available through various means. In historical times, town criers provided data and the community responded. Today, emergency response numbers are available for the public to report incidents and get help from civic services. Police are equipped with mobile communication equipment that can relay news as it happens. Reporters can broadcast breaking news events from satellite broadcast vans. All these though involve human involvement in generation of the data and informing those who can act on it. ICT interventions in many smart city initiatives aim to automate this via use of sensors and IoT devices that can communicate and store the data on centralized systems, from where mission control dashboards can relay real time events to interested personnel.

Data that is thus collected by IoT devices and sensors, as part of smart city initiatives, begets us the question as to who should be taking care of it. Who governs that data, who gets access to that data, what is collected and what is not, how can that data be used - these and more questions arise from data governance and privacy standpoints. Identity privacy, query privacy, location privacy, footprint privacy and owner privacy are dimensions of a citizen privacy model proposed by Martinez-Balleste, Perez-Martinez and Solanas(2013). Data over-collection referring to smart phone applications collecting more data than its original function while within the permission scope, is rapidly becoming one of the most serious potential security hazards in smart city applications as observed by Li, Dai, Ming and Qiu (2016). It was found based on extensive experimental results that putting all user data on the cloud can greatly improve data security.

4. Libraries and Smart City Data

Efforts to reinvent themselves as hubs, makerspaces and community centres offering spaces for social gathering, targeted learning and production, including entrepreneurial activity and innovation, are strategies for maintaining and extending the traditional remit of the public library as indicated by

Leorke, Wyatt and McQuire (2018). It is also informed that public libraries figure strongly in smart city strategies. It was also observed that public libraries certainly can play an important role in the future development of Flemish and Dutch cities by Vallet (2013). It concludes that when the identified generic and specific strategic roles of the Flemish and Dutch public libraries are embedded in city innovation projects, they definitely become part of a broader network focusing on the overall future urban development. Thus, libraries are both a recipient as well as contributor to smart city efforts.

With the role of public libraries being an integral part of smart city initiatives, either as new developments added via smart city projects or as part of the existing city administrative team that is working on the smart city initiative, they can be considered for various roles in the smart city initiatives. With their investments in ICT for staff as well as infrastructure including digital library and cloud computing, leading public libraries are candidates for handling data related to the smart city initiatives with attendant management of data governance and data privacy aspects.

4.1 Network

In the section Records and Archives of Communities, we see how data regarding our villages and communities was maintained, how librarians are playing a role in the computerization of records and archives, and also the need for collecting data from geographically dispersed records and archives so that all information regarding an individual can be aggregated. This aggregation of historical records and archives on various metrics and initiatives of villages, towns, cities and communities can help smart city initiatives learn from historical outcomes, study changes over time, and can help plan services to be in alignment with the requirements and modeled on prior successful initiatives in that region.

With the extensive work done by librarians in archive and record management, libraries can start by computerizing the records and archives, if not done already, and by networking those in order to support aggregated and integrated view of that geographically dispersed historical data. Libraries of affiliating universities, usually based in major cities, can be the nodal centers for specialized equipment and staff, while branch libraries at various institutions can serve as data collection and data entry centers, with exceptions for specific skill based centers already in operation. Such library networks in India have been described by Viswanathan (1991). Infrastructure for such a library network can be paired with existing investments in optic fibre cable capability being made across the country, Figure 1 shows the one for RailTel.

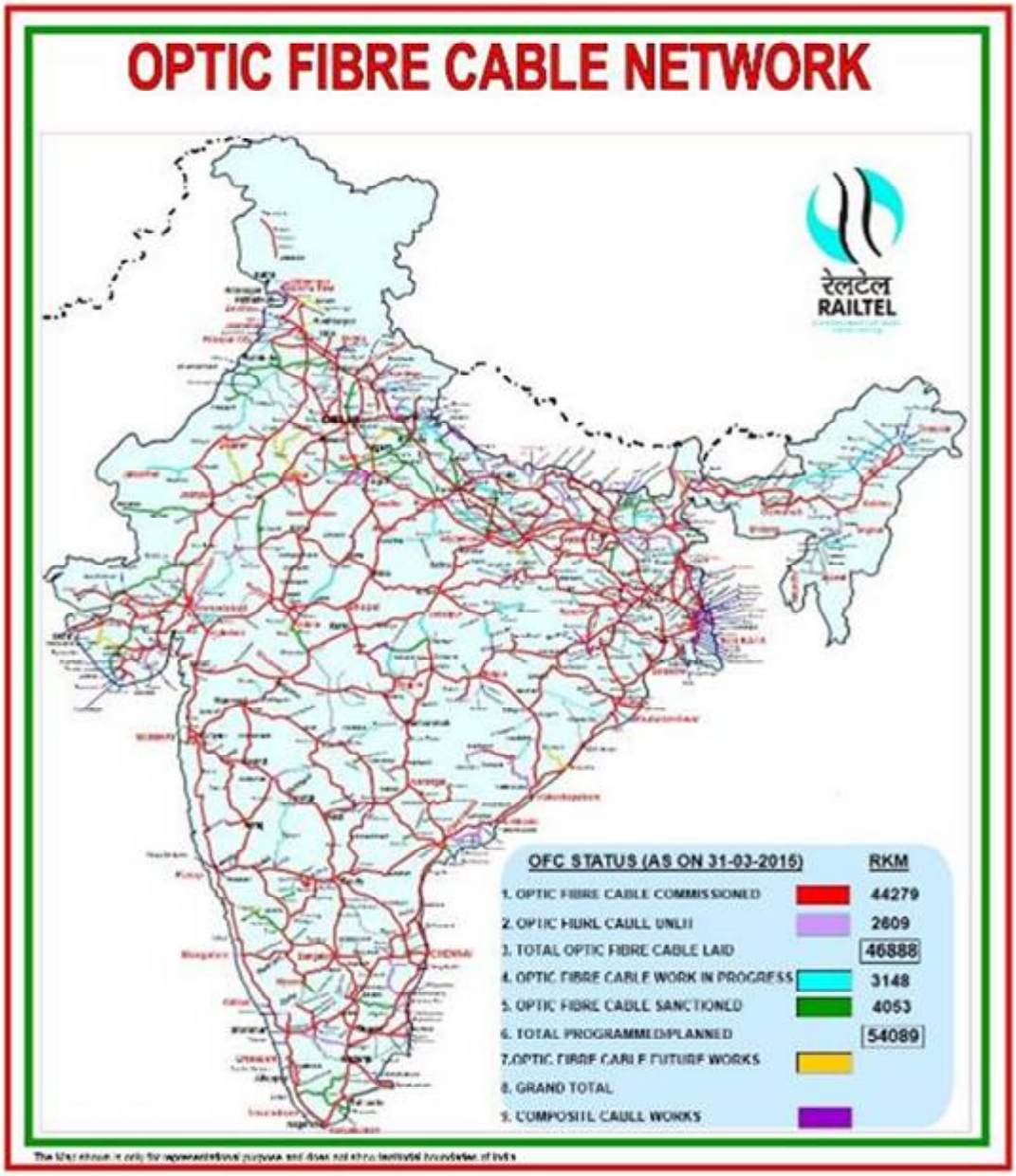

Figure 1: RailTel Optic Fibre Cable Network, Adapted from "About Us - RailTel Infrastructure & Capabilities" Copyright 2016-2017, RailTel Corporation of India Ltd.

This pairing with existing infrastructure can reduce costs and allow excess capacity to be used for other projects and initiatives. Overall, with libraries usually being located at or near places with rail connectivity,

this represents a macro view of the kind of network that would be required for archives and records of various libraries to be aggregated and integrated.

It is also required to have city wide fiber networks for Internet of Things (IoT) devices, sensors and real-time monitoring equipment including CCTV, Fire Alarms, etc. While it is not expected that libraries shall set them up, they do have to connect to that network in order to help aggregate, organize and analyze relevant data for the purpose of records and archives maintenance apart from creating suitable repositories and reports. Robinson (2018) postulates that while smart city infrastructure is enabling efficiency, cost reduction and improved living for city occupants, it needs edge networks to create these efficiencies. Cloud-connected Edge networks with support for multiple clouds can allow for various capabilities at different points of time and it would be required for libraries to have a connection to the cloud providers involved in order to access the data being live streamed or its summary being uploaded from edge computing devices.

4.2 Storage

One has to consider the storage solution for archive and records data. We must examine its size, nature and access requirements to do so. Historical data is used to track change over time, and in performing analysis that can contribute to the planning stages of smart city projects. Access to such data is usually in a planned manner and hence, need not use high speed or high availability infrastructure. Offline or cool storage with suitable access mechanisms can be implemented. The nature of data being organized also governs the choice of storage. With historical archive and record data being in various stages of organization from scanned images to digitized sculptures, specialized applications could be needed to access them. Managed object cloud storage can not only offer a low cost option, but also help ensure a backup location separate from the physical library centers. Per Kovacs (2017), Amazon S3 and Azure Blob Storage are available from $0.002 to $0.004 per GB per month, with capacity for storing large objects around 5TB. This gives enough scope for rich data capture including high resolution videos, if required. AWS (2019) informs that Amazon Glacier offers highly affordable long-term storage that can replace tape for archive and regulatory compliance.

For IoT devices and sensors, real-time monitoring solutions and other devices on the Edge network, storage has to support high speed writing and access for quick analysis and real time decision making. Artificial Intelligence and Machine Learning solutions are needed for real-time interventions and actions. Ideally, a variety of cloud providers are to be supported for various types of AI and ML solutions. They need access to the streamed data or to summary data from Edge computing devices that can analyze and summarize the data before transmission. In either case, there is need for a high velocity storage for intra-city smart city network devices generating high volumes of high variety formats like unstructured and loosely structured data, and with the need to support AI and ML use cases. This needs careful consideration per Dimitrovich (2016) with regard to velocity, variety and volume of data among other factors. In case of specialized solutions like policing systems with face recognition and other compute intensive requirements, Amazon FSx for Lustre can be considered per AWS (2019).

4.3 Data Ingestion and Analytics

With IoT use cases, Phillips (2018) informs the need for a flexible data management infrastructure, for enabling higher level functionality in the application layer, with its services including data normalization, policy enforcement, asset modeling, user roles, audit trails, and multi-tenancy. Various enterprise solutions on top of basic cloud services can be adopted.

1. AWS (2019) provides multiple ways to ingest real-time data generated from websites, mobile apps, and internet-connected devices, with a simple way to capture and load streaming data or IoT device data involving use of Amazon Kinesis Data Firehose, Amazon Kinesis Video Streams, and AWS IoT Core.

2. Microsoft (2019) presents Azure Stream Analytics as an easy-to-use, real-time analytics service that is designed for mission-critical workloads, easily extensible with custom code and built-in machine learning capabilities. It offers rapid scalability with elastic capacity to build robust streaming data pipelines and analyze millions of events at sub-second latencies. Microsoft (2019) also offers Event Hubs as a fully managed, real-time data ingestion service that is simple, trusted and scalable. It allows users to stream millions of events per second from any source to build dynamic data pipelines and immediately respond.

3. Freedman (2016) showed it is also possible to setup a data analysis pipeline, where real-time streams are collected from edge devices, gateways, or other clouds, and then processed by Apache Spark Streaming applications, which in turn generate derived streams for further processing, data aggregates, or trigger other real-time events.

4. Google Cloud (2019) informs that Stream analytics from Google Cloud can make data organized, useful, and accessible from the instant it is generated. Built on autoscaling infrastructure of its core components Cloud Pub/Sub, Cloud Dataflow, and BigQuery, this streaming solution reduces complexity by provisioning the exact resources needed to ingest, process, and analyze fluctuating volumes of real-time data.

5. Distributed In-memory storage and analytics can bridge high velocity and high volume data per Williams et. al. (2014). Solutions range from analytics using Microsoft Power BI,

4.4 Deployment

With the massive number of nodes for a national library network, cloud and datacenter deployments to provision and setup resources could be a complex and time consuming process apart from requiring high cost cloud computing solution architects and experts. With various processes for procurement of services from IT consultants, it can be beneficial to have it included in the services provided by a cloud vendor.

Multi-cloud vendor supporting unified cloud platforms like Drootoo (2019) offer templates for use cases from computing to storage and virtual machines, enabling rapid deployment of performance optimized and best practices adherent computing infrastructure that has already been setup and configured by experts. Customization of templates is possible and helps departments and vendors to

make changes and share that computing environment, for deployment by others. This can be adopted by centers which do not have cloud computing experts available, so that a standardized smart city cloud computing infrastructure can be adopted across many centers in a rapid manner.

4.5 Data Dissemination

Smart city data is for dissemination to various user groups from civic bodies in charge of city infrastructure and services to city residents who want to know the status of various amenities. Since smart city data collection potentially involves personally identifiable information, privacy is of much concern, as discussed, while distributing that data. In order to follow privacy guidelines in access and distribution of such data, data security principles including data encryption at rest is applied. Applications, on web and on mobile, become the delivery vehicles for smart city information so that users can benefit in various ways

- Environment data – climate, weather conditions, pollution status, water availability, etc. – can be used for planning people movement from individual travel plans, to emergency migration of communities during natural disasters.
- Sensor data – fire alarms, cctv and other monitoring systems, access control systems, biometric access control, etc. – can be used to alert users of hazardous activities and events. Policing and security can be improved manifold even with personnel resource constraints by application of AI and ML to auto-analyze such data.
- Fiscal data – finance, commerce, industry, etc. – can be used to guide users on investment and savings opportunities, apart from employment prospects in a particular region.
- Education data – education, economy, commerce, industries, statistics – can be used to plan higher education and learning
- Governance data – economy, commerce, industries, governance and administration, statistics – can be used to plan investments, budgets and subsidy as per the status.

Cities can use the data to monitor various dimensions per UN HABITAT (2019) -

1. Productivity – Products and Services per capita, Age demographics based employment metrics
2. Quality of Life – Life expectancy, mortality rate, literacy rate, Green area per capita
3. Environmental Sustainability – Pollution and Carbon metrics, Solid and Water waste treatment, Renewable energy consumption
4. Infrastructure and Development – Metrics on Shelter, Water supply, Medical facilities, Internet and Mobile penetration, public transport, traffic accidents, road and highways development
5. Equity and Social Inclusion – Poverty rate, Youth unemployment, Slum households
6. Governance and Legislation – Voter turnout, Revenue collection, Days to start a business.

5. Conclusion - Libraries as Data Repositories for Smart City projects

Poon (2019) recorded the observations of a communication director for an association of leading libraries across North America that there is no strong model for handling private and public data, whereby

the library as a local institution that is most trusted can be given the responsibility, and that is a whole new paradigm for how data can be handled in a local government. In open data initiatives in various cities, libraries are participating in various capacities from hosting the data, to making it available for public access on a web portal. City boards of trade are looking at libraries as an acceptable option to oversee data governance for smart city projects. They believed that development-focused agencies do not have the mandate or resources to develop technology policies for the larger city, nor have resources to provide independent enforcement or oversight over emerging technology practices.

Ahmad, JianMing and Rafi (2019) performed a study showing a strong correlation between the required competencies and skills of librarians for the implementation of Big Data analytics in academic libraries. With libraries and librarians investing in AI, Big Data and Data Analytics capabilities, libraries can add value to such data by processing it and making the results available for public access in data repositories. This can contribute to data driven planning and development for smart city initiatives.

References

1. Ahmad K., JianMing Z. and Rafi M. (2019). An analysis of academic librarians competencies and skills for implementation of Big Data analytics in libraries: A correlational study. Data Technologies and Applications. Vol. 53 No. 2, pp. 201-216.
2. American Library Association. (2019). Impact on Community Development. Available at http://www.ala.org/tools/research/librariesmatter/impact-community-development (Accessed on 31/07/2019)
3. American Library Association. (2019). Connecting People and Ideas. Available at http://www.ala.org/tools/research/librariesmatter/category/connecting-people-and-ideas (Accessed on 31/07/2019)
4. AWS. (2019). Data Lakes and Analytics on AWS. Available at https://aws.amazon.com/big-data/datalakes-and-analytics/ (Accessed on 31/07/2019)
5. AWS. (2019). Cloud Storage with AWS. Available at https://aws.amazon.com/products/storage/ (Accessed on 31/07/2019)
6. Dimitrovich S. (2016). Picking the Right Data Store for Your Workload. AWS Startups Blog. Available at https://aws.amazon.com/blogs/startups/picking-the-right-data-store-for-your-workload/ (Accessed on 31/07/2019)
7. Drootoo. (2019). SMART CITY. Available at http://drootoo.com/case-studies/smart-city/ (Accessed on 31/07/2019)
8. Freedman M. (2016). Spark Streaming and IoT. Available at https://databricks.com/session/spark-streaming-and-iot (Accessed on 31/07/2019)
9. Government of India. (2012). About Open Government Data (OGD) Platform India. Available at https://data.gov.in/about-us (Accessed on 31/07/2019)
10. Government of India Ministry of Housing and Urban Affairs. (2019). Sectors. Smart city. Available at https://smartcities.data.gov.in/ (Accessed on 31/07/2019)

11. Harries S. (2012). Records Management and Knowledge Mobilisation - A Handbook for Regulation, Innovation and Transformation. A volume in Chandos Information Professional Series. (pp 145-168). Chandos Publishing.

12. Kovacs G. (2017). Comparing AWS and Azure Storage Pricing And Features. Available at https://cloud.netapp.com/blog/aws-vs-azure-cloud-storage-comparison (Accessed on 31/07/2019)

13. Leckie GJ, and Hopkins, J. (2012), The Public Place of Central Libraries: Findings from Toronto and Vancouver. American Library Association. February 28, 2012.Available at http://www.ala.org/tools/research/librariesmatter/public-place-central-libraries-findings-toronto-and-vancouver (Accessed on 31/07/2019)

14. Leorke D., Wyatt D., and McQuire S. (2018). More than just a library: public libraries in the 'smart city'. City Cult. Soc. 15, 37–44 (2018)

15. Li Y., Dai W., Ming Z., and Qiu M. (2016). Privacy Protection for Preventing Data Over-Collection in Smart City. IEEE Transactions on Computers. Vol. 65, no. 5, pp. 1339- 1350, 1 May 2016.

16. Martinez-Balleste A., Perez-Martinez P. A., and Solanas A. (2013). The pursuit of citizens' privacy: a privacy-aware smart city is possible. IEEE Communications Magazine. Vol. 51, No. 6, pp. 136-141, June 2013.

17. Microsoft. (2019). Azure Stream Analytics. Available at https://azure.microsoft.com/en-us/services/stream-analytics/ (Accessed on 31/07/2019)

18. Microsoft. (2019). Event Hubs. Available at https://azure.microsoft.com/en-in/services/event- hubs/ (Accessed on 31/07/2019)

19. Okabe J., and Surjit V. (2012), Village-Level Birth Records: A Case Study. Review of Agrarian Studies. Vol. 2, no. 1, Available at http://ras.org.in/village_level_birth_records (Accessed on 31/07/2019)

20. Phillips M. (2018). Azure IoT vs. AWS IoT: Read Before Building Your IIoT Platform. Available at https://dzone.com/articles/azure-iot-vs-aws-iot-read-before-building-your-iiot-platform (Accessed on 31/07/2019)

21. Poon L. (2019). Should Libraries Be Keepers of Their Cities' Public Data?. Available at https://www.citylab.com/life/2019/02/libraries-public-information-city-data-digital-archive/581905/ (Accessed on 31/07/2019)

22. RailTel. (2017). RailTel Infrastructure & Capabilities. Avalable at http://railtelindia.com/blog/index.php?option=com_content&view=article&id=79:railtel-infrastructure-capabilities&catid=94&Itemid=529 (Accessed on 31/07/2019)

23. Robinson S. (2018). Smart city infrastructure needs edge networks. Available at https://www.cisco.com/c/en/us/solutions/internet-of-things/smart-city-infrastructure-edge-networks.html (Accessed on 31/07/2019)

24. State Archives. (2004). About Us. Available at http://keralastatearchives.org/main1.htm (Accessed on 31/07/2019)

25. Tian H., Banger K., Bo T., and Dadhwal V.K. (2014). History of land use in India during 1880- 2010: Large-scale land transformations reconstructed from satellite data and historical archives. Global and Planetary Change, Vol. 121 October 2014, pp. 78-88

26. UN HABITAT. (2019). THE DIMENSIONS. City Prosperity Initiative. Available at http://cpi.unhabitat.org/ (Accessed on 31/07/2019)

27. Vallet N. (2013). Becoming partners in urban development. Library Management. Vol. 34 No. 8/9, pp. 650-663.

28. Viswanathan T, Mittal R and Lakshmi V V. (1991). Library networks in India. Annals of Library Science and Documentation. 38(2) (1991) 39-52.

29. Walsham A. (2016). The Social History of the Archive: Record-Keeping in Early Modern Europe. Past & Present. Volume 230, Issue suppl_11, November 2016, Pages 9–48.

30. Wikipedia. (2019). Archive. Available at https://en.wikipedia.org/wiki/Archive (Accessed on 31/07/2019)

31. Wikipedia. (2019). Smart city. Available at https://en.wikipedia.org/wiki/Smart_city (Accessed on 31/07/2019)

32. Williams J. W. [et. al.]. (2014). Bridging high velocity and high volume industrial big data through distributed in-memory storage & analytics. Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, 2014, pp. 932-941.

About Authors

Dr. S. Bharanidharan - Computer Scientist, Anna University, Chennai
Email: s_bharani@annauniv.edu

Mr. Vijay Krishnan, Research Scholar, Anna University, Chennai

Dr. G. Krishnamoorthy, University Librarian, University Library. Anna University, Chennai