

The development of a public library data collection and analysis platform and its uses: a case study of the National Library Big Data Platform in Korea

Jungyeoun Lee¹ and Nahyun Kwon²

¹Department of Library and Information Science,
Joongbu University, SOUTH KOREA

²Department of Library and Information Science,
Myongji University, SOUTH KOREA

e-mail: jyonlee@joongbu.ac.kr; nkwon@mju.ac.kr (corresponding author)

ORCID ID: J.Lee: 0000-0003-2543-959X

N.Kwon: 0000-0003-0384-2345

ABSTRACT

Public libraries reflect the contemporary way of life. In the era of the digital data age, actions carried out through libraries can be traced and analyzed to monitor the demand for library services. This study presents the development and utilization of Korea's National Library Big Data Platform (NLBDP), which was constructed to collect operational data from public libraries in the country and to build national open data platform for the participating libraries and for public use. This study analyzed 17 application cases that utilized NLBDP in order to understand the way the platform has been used. The findings showed that the types of data analyzed so far have been somewhat limited. It suggests needs for further enhancing the data quality and for diversifying data sources, such as individual local library operational data, surveys, demographics, news, and social media. The results also revealed that NLBDP has been utilized for seven distinct purposes: library operations and strategic planning; social trend analysis; collection management; redesigning library space; recruiting new patrons; library marketing; and non-library applications. Building an infrastructure for networked data analysis not only enables evidence-based decision-making in public libraries, but also extracts timely topics that can be developed as useful services for their community. The findings of this study highlighted the necessity to formulate national library data policies and gather data capable of measuring the social impact of libraries.

Keywords: National library; Big data; Library policy; Library data trend analysis; Public libraries.

INTRODUCTION

Data serves as the imprint of human behavior. In this era centered around data, organizations are fostering their growth by conducting logical analyses of data and making informed decisions. Public libraries are no exception to this trend. Data collected from public libraries can be used to track citizens' everyday life experiences and expectations toward libraries.

Traditional library records, such as collections, users, and circulation records, have been analyzed for library management on a local library level. With the advancement of big data technologies, the integration and analysis of data gathered from thousands of libraries on a national scale has become achievable. Joshi (2015) argued that libraries can extract values for library operations and improve services through collecting, storing, processing, and analyzing big data. Garoufallou and Gaitanou (2021) also claimed that public libraries can create value-added services for the community using big data. Zhan and Widen (2018) emphasized the importance of big data systems in delivering public library services in Finland. Data-driven library management can equip library managers and policy-makers with tools for scientific decision-making. It also allows public libraries to enhance collection management and user services to meet community needs.

Nonetheless, it is difficult to collect and analyze large amounts of big data with existing data management and analysis systems. It requires a considerable amount of costs and capable librarians to operate the system. Li et al. (2017) argued that human resources, literature resources, technical support, service innovation, and infrastructure construction are needed in implementing a library big data utilization framework successfully. Thus, it would be especially challenging for small and medium-sized public libraries (Bertot, Butler and Travis 2014; Lawton and Burn 2015; Pawar 2016).

Due to the challenges inherent in implementing a big data system at a local library level, there have been efforts directed towards establishing a national level library big data system. In South Korea, the National Library of Korea (NLK) took the initiative to build national big data platform in 2014 and has accumulated years of experience. Both the collected data and services developed within the framework of the big data platform have been utilized as open data by both public libraries and the commercial sector in the country. Thus, this study aimed to examine how public libraries utilize this national initiative to engage in data-driven library operations and develop library services for their community. The findings of this study would help identify the areas where the platform can be improved as open data platform for public libraries and the general public.

LITERATURE REVIEW

Collecting public library data at the national level is not an entirely new concept. Annually since 1988, the US Institute of Museum and Library Services (2023) has been collecting library statistics from 9,000 public libraries nationwide using the Public Libraries Survey (PLS). The data is collected through each state library, and any personally identifiable information is subsequently removed before being made to the general public for research purpose. The data include information about library visits, circulation, collection size, public service hours, staffing, electronic resources, operating revenues and expenditures, and the number of service outlets. The PLS also provides definitions of data elements. However, the PLS data primarily focus on input and output data, with a dearth of outcome and impact data.

More recently, the decline in library use has addressed the need to measure the social impact and values of library services (Creaser 2018). It addressed the importance of library outcome data, and there have been efforts to collect such data. The Project Outcome, an outcome evaluation project of the Public Library Association (PLA), provides useful complementary datasets to the PLS. Public Library Association (2018) also developed data analytics toolkit

through Project Outcome in 2015 to advocate the library's impact to the society. Project Outcome has also made it possible to collect nationwide quantitative and qualitative data concerning key aspects of public libraries' performance. As of January 2022, a total of 2,057 libraries are engaged in this initiative. Additionally, state libraries across the United States are investing efforts to the collection and use of data. For example, the State of North Carolina annually releases the statistics on the budget, collections, staff, facilities, and programs of all public libraries collected from county library systems. According to Chow and Tian (2021), a decade's worth of public library data spanning from 2005 to 2014, encompassing variables such as service population, staff, budget, circulations, and programs, were extracted to scrutinize the impact of public libraries on the quality of life of state residents. In addition, census data from the US Department of Labor and Statistics, household annual income, education level indicators, and data on employment and poverty were also extracted.

Singapore, a city-state, serves as an exemplary case of data-based library policies implemented at the national level. The National Library Board (NLB) of Singapore systematically collects and analyzes data from 26 public libraries, integrating the insights into library management practices (National Library Board 2019). The NLB attempted to innovate libraries by analyzing lifestyle trends and responding to the worldwide decline in library visits by furnishing spaces and contents (Dresel et al. 2020). It also developed a data analysis platform collaboratively with the Singapore Land Authority, a geo-spatial analysis service institution, considered new establishments and expansions according to population growth, and predicted the future through need analysis data, rules, and algorithm analysis. The NLB built a recommendation service based on predictions and circulation records (Ong 2014) and developed the Reach Index in 2015 as an annual performance indicator (Lee 2018). It measured service performance goals every year to implement library establishment and remodeling policies. In short, Singapore's NLB can be regarded as a representative data-driven practice of the collection and analysis of national public library data.

In Korea, the National Library Big Data Platform (NLBDP) was developed in an effort to implement a big data system of public libraries by the Presidential Committee on Library and Information Policy of Korea in 2014. It was launched under the project name "Library Big Data Analysis and Utilization System Establishment" to lay the foundation for Korea's library data collection, analysis, and utilization technologies. The NLK and public libraries are the main pillars, which have agreed to provide and jointly use information on holdings and circulations. As of April 2023, 86 percent of public libraries in the country are participating in this project by contributing their holdings, users, and circulations data. The platform consists a total of 170,961,600 bibliographic data of collection holdings, 34,158,156 user membership data, and 2,042,062,737 circulation data, which were collected from 1,490 participating public libraries since 2014 (National Library of Korea 2023).

The structure of the NLBDP, as shown in Figure 1, encompasses participating library data, bibliographic data from the NLK, bookstore sales data, and general public domain data. The collected data are stored in the storage infrastructure (e.g., Hadoop, RDBMS) after data cleansing in a data pre-processing system (e.g., morpheme analyzer, duplicate processor, and anomaly data detection algorithm). The analysis system processes these data through statistical analysis, text mining, and visualization.

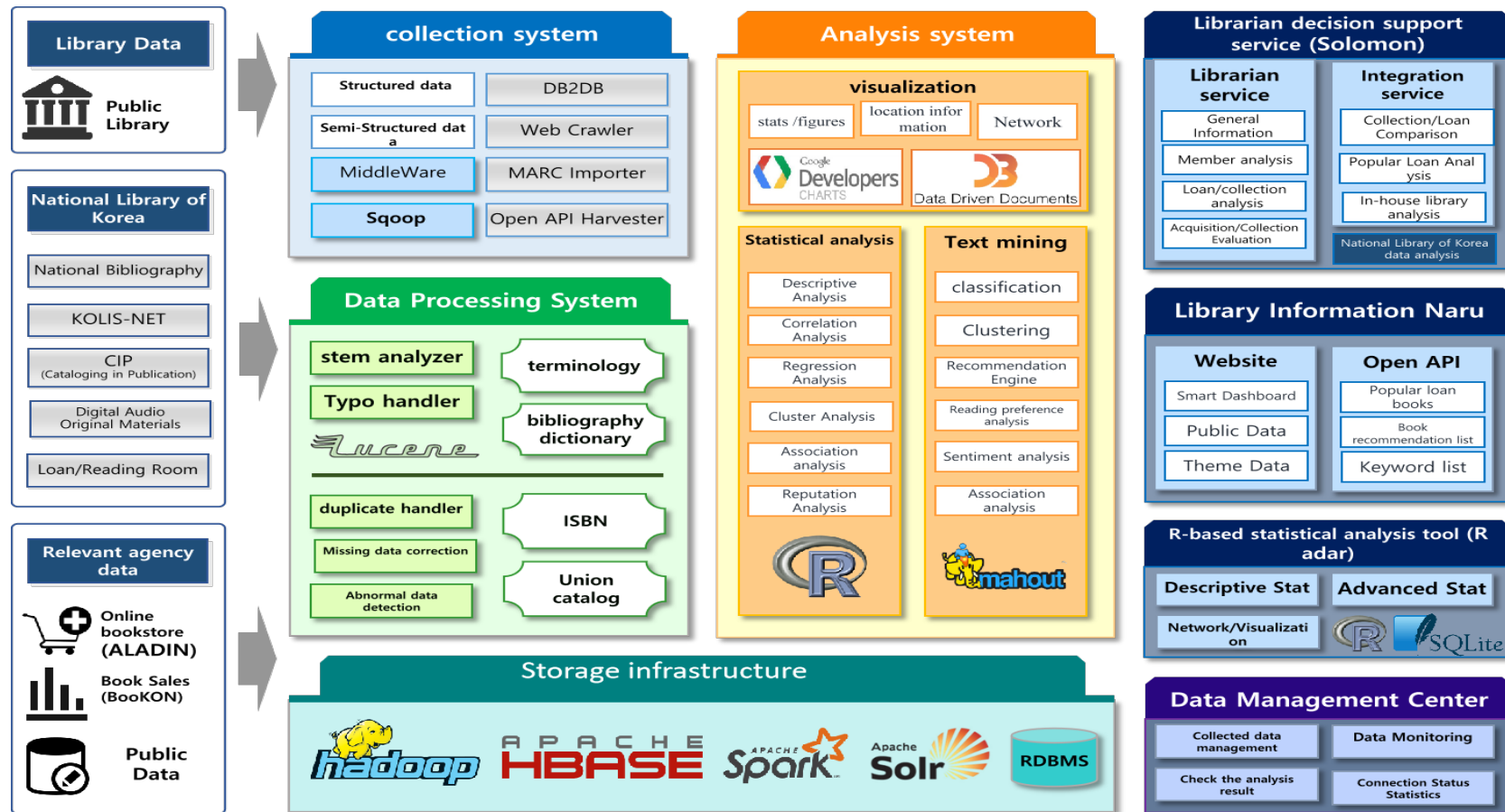


Figure 1: Diagram of the National Library Big Data Platform Configuration (National Library of Korea 2023)

Collection data, user data, and circulation data are collected from individual member public libraries, as shown in Table 1. Personal information in user data and their circulation records are protected through conversion into an encrypted user identification number. The final service platform comprises a decision support service (namely, Solomon), dashboard service (namely, Library Information Platform), an open data service (OpenAPI) and R-based statistical analysis tool (namely, Radar) services. A brief overview of the characteristics of these constituent systems is presented.

Table 1: Types of Collected Data

Category	Collection Items
Collection data	Bibliographic data of library holdings in the collection
User data	Encrypted user number, gender, age, residence zip code, circulation record, subscription library
Circulation data	Circulation records by collection, circulation records by user, etc.

First, “Solomon”, the decision support service, collects and analyzes various data sets, such as local library’s collection data, user data, circulation data, bookstores’ sales data, and public data such as local geo-statistical data, to assist in decision making for library administrations and service planning. Second, the “Library Information Platform” system supports the shared utilization of library data by interested researchers, developers, and libraries. It provides information on the status quo of public libraries nationwide, such as the most circulated titles by gender, age, and region, and provides various types of analytical data, such as national bibliographic data, circulation data, and circulation rankings. In addition, the “OpenAPI service” enables data from participating libraries to be linked and utilized in real time. Lastly, “Radar” is an R-based statistical analysis tool. It provides core statistical functions such as descriptive statistics, regression analysis, crossover analysis, ANOVA, and the t-test.

Previous studies have examined the NLBDP in terms of its values, technical issues in database construction and design, and performances of its application programs (Cho 2019; Kwon, Lee and Lee 2018; On and Park 2020). Yet, few research have investigated real-world applications of the NLBDP for library operations and service development by participating member libraries and the general public, which was one of the major motivations of developing the NLBDP.

OBJECTIVE AND METHOD

This study purported to analyze the types of data and the purposes of utilizing the NLBDP, the big data platform of public libraries. The NLK has been holding annual national competitions for best practice of NLBDP utilization to encourage more participations from public libraries and to promote the platform utilization. A casebook was published in both print and online every year to inform the details of the platform utilizations of each best practice. This study employed content analysis as the research design in analyzing the 17 best practices published in the casebooks between 2016 and 2020 (National Library of Korea 2016-2020). Specifically, the following two research questions were developed for this study:

- (a) What types of data do the best practices utilize from the NLBDP? How are the external data utilized in linking with the NLBDP?; and
- (b) What are the participating institutions' purposes of utilizing NLBDP?

Two coders analyzed the contents of the 17 best practices published in the casebook in terms of the type of data utilized, external data use, purposes of external data use, and purposes of NLBDP utilizations. To achieve satisfactory inter-coder reliabilities, the recoding process was repeated until it reaches the complete agreement between the two coders.

RESULTS

The results are reported in the sequence of two research questions.

Types of Data Analyzed

The types of data utilized by the 17 best practices could be categorized into

- (a) NLBDP data (i.e., data from the NLBDP's services),
- (b) local library data (i.e., data created by and stored in individual local libraries), and
- (c) external data (i.e., data obtained from external sources).

The 17 cases (C1 – C17) are listed in Table 2, along with their names, year of establishment, and the types of data utilized.

The most frequently used data type was Solomon of NLBDP, which was 15 of the 17 cases (88.2%), followed by Library Information Platform of NLBDP (47.1%). While nine of the 17 cases utilized NLBDP data only (52.9%), the other eight cases utilized NLBDP along with other data sources, such as local library data, surveys, demographic information, and social media (47.1%). Of the eight cases that utilized local data sources, six utilized NLBDP along with their own library data. Four cases used their library operational data (23.5%) and two cases used user survey data administered by the local library (11.2%). Because patrons' personal information had been removed from the NLBDP database, these libraries analyzed their operational data along with the NLBDP data to bring more meaningful analysis. Four cases utilized these external sources along with NLBDP: one library used national statistical database (5.9%) and three libraries used social media data (17.6%).

The findings inform that the type of data utilized by 17 best practices tends to be straightforward, as over half of the cases exclusively utilized NLBDP data. However, this pattern appears to be shifting. The use of NLBDP by public libraries has been on the rise in more recent years, as evidenced in Table 2. Out of the seven public libraries chosen in 2018 and beyond, six of them incorporated data sources other than NLBDP.

The analysis further revealed challenges in data analysis when linking NLBDP with external data. The process of data cleaning and verification was found to be essential during the data linkage. It is imperative that the definitions of data in both the NLBDP and the library management system align, and the data linkage must undergo meticulous scrutiny. For example, Guro Library (C4) reported that it took five months to figure out the reasons behind the top-ranked books in the NLBDP receiving low rankings in their library. It turned out that this problem occurred when data in NLBDP did not process duplicate copies of the same book properly. This case demonstrates the challenges associated with integrating nation-

The Development of a Public Library Data Collection and Analysis Platform and Its Uses

wide library data into a unified system, suggesting the importance of data verification when establishing links to external data sources.

Table 2: Library Big Data System Use Cases
(*Commercial companies)

Year	Case	Library Name	Year	NLBDP Data			Local Library Data		External Data	
				Solomon	Library information platform	Radar	Library operation data	User survey data	Statistical data	Social media
2016	C1	Bucheon City Library, Bucheon	1985	⊙	-	-	-	-	-	-
	C2	Dapsimni Library, Dongdaemun-gu, Seoul	2014	⊙	⊙	-	-	-	-	-
	C3	Seoknam Library, Incheon	2008	⊙	-	-	-	⊙	-	-
2017	C4	Guro Library, Seoul	1984	⊙	-	-	⊙	-	-	-
	C5	Jechon City Library, Jechon	1996	⊙	-	-	-	-	-	-
	C6	Dapsimni Library, Dongdaemun-gu, Seoul	2014	⊙	⊙	-	-	-	-	-
	C7	Gayang Library, Daejeon	2001	⊙	-	-	-	-	-	-
	C8	Dalsung County Library, Dalsung	2014	⊙	-	-	-	-	-	-
2018	C9	Kangdong Library, Seoul	1984	⊙	-	⊙	⊙	-	-	⊙
	C10	Naver*	1999	-	⊙	-	-	-	-	-
	C11	Brain Cola*	2015	-	⊙	-	-	-	-	-
2019	C12	Gangnam Gu Library, Seoul	1997	⊙	⊙	-	-	-	-	⊙
	C13	Kumam Library, Incheon	2017	⊙	⊙	⊙	⊙	-	-	-
	C14	Seoul Children's Library, Seoul	1979	⊙	-	-	⊙	-	-	-
2020	C15	Sujung Library, SeongNam	2000	⊙	⊙	-	-	⊙	⊙	-
	C16	Yonghak Library, Daegu	2010	⊙	-	-	-	-	-	⊙
	C17	Mapo Lifelong Learning Center, Seoul	1980	⊙	⊙	-	-	-	-	-
Total (17)			Case	15	8	2	4	2	1	3
			%	88.2%	47.1%	11.7%	23.5%	11.7%	5.9%	17.6%

Purpose of Data Analysis

When categorizing the 17 cases based on the purpose of data analysis, seven distinct categories emerged: (a) library operations and strategic planning, (b) social trend analysis, (c) collection management, (d) redesigning library space, (e) recruiting new patrons, (f) library

marketing, and (g) non-library applications via OpenAPI. The first six categories pertain to applications within public libraries, whereas the final one pertains to usage within the private sector. The following sub-section presents each of the seven categories.

(a) Library Operations and Strategic Planning

The first type of data analysis purpose was for library operation and strategic planning. In these cases, as shown in Table 3, statistical data of individual libraries, demographic data of the local community, and surveys assessing citizens’ needs were analyzed alongside NLBDP data.

Table 3: Applications in Library Operations and Strategic Planning

Utilized Data	Application Examples (Case no.)
<ul style="list-style-type: none"> • User data • Collection data • Circulation data • Local library operational data • Statistical data (e.g., demographics) • User survey data 	<ul style="list-style-type: none"> • Annual Assessments and Planning (C1, C3, C12) • Decisions on new library locations (C12) • Decisions on library kiosk locations (C12)

The data employed for the development of library service planning include analyses demographic characteristics of library users and non-users, and usage statistics segmented by geographic regions within service areas. Quarterly and annual circulation statistics were analyzed by gender, region, and age of the patrons. These analyses helped library administrators decide a new library location in a high traffic area and to locate sites to open new library kiosk.

(b) Social Trend Analysis

The second purpose of NLBDP analysis relates to identifying community needs and social trends. As shown in Table 4, data were extracted from circulation statistics and turnover rates of library collections, and the most highly circulated titles nationwide using the Network Library Service, i.e. an NLBDP service that enables a local library to compare its usage data with those of other libraries and bookstores.

Table 4: Applications in Social Trend Analysis

Utilized Data	Application Examples (Case no.)
<ul style="list-style-type: none"> • Collection data • User data • Circulation data • All popular books • Social media data • Online bookstore sales 	<ul style="list-style-type: none"> • Quarterly list of recommended books (C1,C11,C12) • 50 book titles for patrons in different stages in life cycle (C4) • Book curation service (C4,C5,C6,C7,C9,C11, C13) • Reading program development (C2,C3,C6) • Selection of popular topics that attract social attention (C6,C7,C8,C14,C16) • Selection of popular writers, cultural programs, user groups (C2,C5,C6,C8,C9,C16)

As an example of social trend analysis type, firstly, users were grouped based on library circulation statistics to formulate a library service that meets citizens’ interest. Secondly, to develop topics of interest for the community, the latest social trends were identified by

analyzing the words appearing in social data. It informed choices on how to curate related collections and programs. This initiative used media attentions, the latest publication trends, and social media data from platforms such as NAVER Analytics and TEXTOM. For example, Yonghak Public Library (C16) identified “Corona” (i.e., COVID-19) as the latest topic of social interest via TEXTOM, which prompted librarians to curate collections related to psychology and immunity, and plan cultural programs, such as book talks and exhibitions.

(c) Collection Management

The third purpose of NLBDP utilization relates to collection management, which involved the analysis of data on circulation and collection turnover rates (Table 5). The most common practice of collection development was to identify “priority books” or “specialized topics” and to allocate budget for the related topical areas. Another library (such as C1, C6, C13, C17) established an acquisition plan based on circulation data from other libraries of similar size and topics reflecting most popular books and social trends in the country. After extracting a list of books that have not been circulated despite having significant values, librarians moved them to more prominent locations within the library to increase visibility and encourage circulation.

Table 5: Applications in Collection Management

Utilized Data	Application Examples (Case no.)
<ul style="list-style-type: none"> • Circulated books • Nationally popular books • Collection turnover rates • Books never circulated 	<ul style="list-style-type: none"> • Establishing a scientific collection development plan (C3, C17) • Reallocation of acquisition budget (C1,C7,C13) • Identifying titles of high turnover rates for purchase (C13,C16) • Enhancing collection turnover rates (C1,C4,C6,C14)

(d) Redesigning Library Space

The fourth type of purpose relates to redesigning library space. It uses the data from changing the library space layout (C9) to improving library homepage interface (C6), as shown in Table 6. Based on the library circulation statistics, reservation statistics, and turnover rates of collection, curating corners section for priority books on special topics can be created. This service will keep the collection up-to-date and promote active use. In addition, titles with little usage were moved to repository, and potentially interesting titles with little usage were developed as themes for book curations by moving to a location that attracts patrons’ attention (C8, C9, C16, C17). Another library (C1) developed customized book recommendation services for individual patrons on the library website based on OpenAPI linkage.

Table 6: Applications for Redesigning Library Space

Utilized Data	Application Examples (Case no.)
<ul style="list-style-type: none"> • Circulation data • Reservation statistics • Collection turnover rate • Seat reservation data • Popular books • OpenAPI linkage 	<ul style="list-style-type: none"> • Displaying titles with high usage, develop a special subject corner, and displaying monthly themed books (C9) • Moving little used books to repository (C8,C9) • Developing little used books as recommended book curation (C16, C17) • Search system linkage development (C1) • Homepage guide to popular books by type/age (C6)

(e) Recruiting New Patrons

The fifth purpose relates to recruitment of new patrons. In the cases presented herein, the region, date, and time of use were extracted and analyzed. As shown in Table 7, libraries developed programs that analyzed local demographics to identify topics of interest to users with low usage rates (C8). They identified the characteristics of non-users by their age, gender, postal codes, and so forth, based on demographic statistics of local residents and library membership data. The results helped the libraries develop new cultural programs that targeted potential users of different age groups (C6, C7, C8). Using a wider range of data could expand these efforts to attract new patrons. For example, planners of outreach services could identify related local organizations, such as local welfare centers, childcare centers, and senior centers. This evidence-based approach of new patron recruitment appeared to be particularly popular amid the global trend of declining public library patrons.

Table 7: Applications for Recruiting New Patrons

Utilized Data	Application Examples (Case no.)
<ul style="list-style-type: none"> • Circulation data • User data • Collection data • Collection turnover • Local demographics 	<ul style="list-style-type: none"> • Identifying low-use groups (C8) • Newly opened night program for men (C6,C7) • Special program for children with Dad (C8,C12) • Development of an outreach program delivering books to childcare classes (C15)

(f) Library Marketing

The sixth purpose of NLBDP analysis is for library promotion and marketing. Local library usage statistics were frequently utilized for this purpose. These initiatives provided the community and local residents with library status data, such as library membership status data, age distributions of library members, the annual number of visitors, collection characteristics, circulation rates by subject, as well as trendy topics of social interest. Intuitive graphics were used to design promotional materials. For example, R statistical program was used to visualize the results of a text mining of circulation collections as a word cloud to inform keyword analysis, or visual materials such as infographics were produced and posted on library newsletters and websites to inform local residents. Table 8 details the application examples in the case libraries.

Table 8: Applications in Library Marketing

Utilized Data	Application Examples (Case no.)
<ul style="list-style-type: none"> • Circulation data • User data • Collection data • “Radar” (the NLBDP statistical software service) 	<ul style="list-style-type: none"> • Library newsletter (C5,C7,C9) • Improving the library website (C10,C12) • Infographics (C2,C6,C9)

(g) Non-library Applications

The last type of NLBDP use relates to private sectors’ use for profitable service development. Two private companies, Naver and Brain Cola, used the OpenAPI big data system to develop corporate products and services for the general public. Naver, the most dominant Internet portal site in Korea, developed a service that provide real-time, up-to-date information about public libraries in Korea, including library service hours, the number of in-house library users, the number of vacant seats in reading rooms, collection size, and library location (C10).

Whereas, Brain Cola developed a mobile application that provides book recommendation services by linking with bookstore data. Using the OpenAPI of the NLK and the OpenAPI of "Library Information Platform," an algorithm was developed by analyzing bibliographic data of library holdings and most highly circulated books by period, gender, and age (C11). These cases demonstrate examples of making national-scale library data available to the public through an OpenAPI to facilitate product development by the private sector.

DISCUSSION

With the advent of the data-based era, the need for evidence-based library practices has increased to accurately grasp community needs using data and to facilitate decision-makings in library administration. While service innovations would be difficult without utilizing big data, according to Igarashi, Koizumi and Widdersheim (2020), collection development that meets community needs is difficult to establish because privacy issue prevent libraries from utilizing patrons' circulation data. Thus, big data systems have been constructed at the national level in many countries, including the United States and Singapore. These countries have employed data-driven library planning using circulation data, reading room reservation data, and external data, such as national census data and various social data. There has been similar efforts in South Korea to develop the NLBDP, a national big data platform of public libraries since 2014. This platform is particularly significant in that it intended to serve as an open data platform for all public libraries in the country as well as the public in general. NLBDP allows anyone to utilize the national public library data and its value-added services to practice evidence-based library operations and service planning without much analytical knowledge and skills. Unlike other systems, NLBDP is designed as a real-time online system. Individual public libraries can conduct community analysis and service improvement by utilizing nationwide dataset of circulations, collections, and user data in real time, which is difficult to build and operate in individual library level (Lawton and Burn 2015). This study attempted to analyze 17 best practices that utilized NLBDP platform in order to identify the way the data platform has been utilized so far and the places for further improvement. Specifically, the types of data utilized for analysis and the purpose of utilizing NLBDP platform were analyzed.

According to the findings, eight out of the 17 best practices utilized the data provided by NLBDP only. The other nine cases utilized NLBDP in linking with external data sources, such as data in the local library's database system, survey data administered by the local library, and various statistical data in the national statistics and social media data. For meaningful analysis, more libraries seemed to be linking with external data sources (C9,C12,C15,C16).

Seven purposes of NLBDP utilizations were also identified: (1) library operations and strategic planning, (2) social trend analysis, (3) collection management, (4) redesigning library space, (5) recruiting new patrons, (6) library marketing, and (7) non-library applications via OpenAPI. Table 9 tabulated the purpose of NLBDP utilization with the best practices. Social trend analysis was found as the most popular purpose, followed by collection management and library marketing. Relatively less used purposes were non-library applications and library operations and strategic planning. This suggests a need for developing NLBDP as a more attractive data platform by incorporating more diverse and useful data to private developers.

Table 9: Best Practices by Purposes of NLBDP Utilization

Purpose of NLBDP Utilizations	Case no. of Best Practices
(1) Library operations & strategic planning	C1, C3, C12
(2) Social trend analysis	C1, C2, C3, C4, C5, C6, C7, C8, C9, C11, C12, C13, C14, C16
(3) Collection management	C1, C3, C4, C6, C7, C13, C14, C16, C17
(4) Redesigning library space	C1, C6, C8, C9, C16, C17
(5) Recruiting new patrons	C6, C7, C8, C12, C15
(6) Library marketing	C2, C5, C6, C7, C9, C10, C12
(7) Non-library applications	C10, C11

While the NLBDP has indeed helped local public libraries initiate data-driven service practices across South Korea, the study also identified several areas for further improvement. These findings provide three implications from the study. First, it was found that the types of data utilized by 17 best practices were largely limited to collection, user, and circulation data. Although operating budgets, types of staffing, and facilities are important inputs for library operation and policy-making, these data are not yet included. Kwon, Lee and Lee (2018) claimed the importance of incorporating input data into the NLBDP when libraries establish strategic planning. They analyzed the types of NLBDP data and suggested the data desirable for further addition to build more useful data platform. Some of those data include operating budget, staffing, facilities, and the national library statistical system. Kwon, Lee and Lee further suggested that unstructured data, such as book reviews, library social media data, and search keyword logs, should be incorporated into the platform to enable more useful multi-faceted data analysis.

Secondly, the NLBDP should give more efforts to collect outcome data such as user satisfaction and library performance assessment. This is because local libraries are facing greater challenges to demonstrate their social impact to their local government. The data-driven era values strategic plans that are based on objective and transparent decision making process. As with the US Project Outcome’s performance index and Singapore’s Reach Index, the ultimate purpose of data collection and analysis is to measure the social impact of libraries. Kwon, Lee and Lee (2018) also addressed a need to incorporate outcome data, such as user satisfaction, library impact, and service quality. Therefore, it is essential for the NLBDP to supplement data that can measures social impact and service performance.

Thirdly, while data linkage with the national geographic information portal is useful to analyze local communities as reported in the case of the NLB of Singapore, few of the cases actively utilized geographic information although they are easily available at the street level from the national open data statistics. Two reasons may explain the lack of utilization of external statistical data: training and data incompatibility.

An explanation that relates to training issue show that although the NLBDP was designed to be utilized without any prior knowledge in big data analytic skills, there appear to be significant psychological and practical barriers to carrying out data analysis. Anyone can easily analyze the NLBDP and customize the results, yet it is still important to provide online/offline support and training for public librarians as demonstrated in the Project Outcome of the Public Library Association in the United States.

The other explanation attributes to incompatibility of data between different data sources. One of the cases reported errors in the data that were transferred from individual libraries to the NLBDP database. The key to big data analysis is the accuracy of the source data. When comparing the data from the local ILS with the data from the NLBDP, data analysts should invest sufficient time in data cleaning and verification to identify problems and ensure the data consistency. Similar problems in data linkage were reported by On and Park (2020) who examined data consistency between NLBDP data and the data from the integrated library system of 23 public libraries in the city of Daejeon. The comparisons of the two systems yielded many discrepant results, which occurred in the process of uploading local library collection data to the big data platform. Supporting On and Park's (2020) report, the findings alert data analysts utilizing national-federated datasets to understand the reliability of data through cross-validation and to derive a semantic analysis. This is of vital importance for a national data construction and sharing system, such as the NLBDP. The PLS of the United States can be exemplary in that it provides documentations with data definition along with the questionnaire, thereby enhancing accurate data utilization and verification. Continuing efforts to secure the reliability of data are needed.

CONCLUSIONS

In conclusion, public libraries in the data era are equipped with powerful tools for library operations and service planning. The national efforts to develop open public library data platform are spread across many countries. It helps libraries focus on creative, convergence-oriented thinking and the analytical ability to detect changes in the community needs and to plan more timely library services. Lessons learned from the experiences of the NLBDP of South Korea could help other countries to develop more effective federated-public library data platforms.

ACKNOWLEDGEMENT

This paper was supported by Joongbu University Research & Development Fund, in 2022.

AUTHOR DECLARATION

There are no conflict of interest involving either of the authors of this work.

REFERENCES

- Bertot, J. C., Butler, B.S. and Travis, D.M. 2014. Local big data: the role of libraries in building community data infrastructures. *Proceedings of the 15th Annual International Conference on Digital Government Research*: 17-23.
- Cho, J. 2019. Analysis of reading domain of men and women elderly using nook lending data. *Journal of the Korean Library and Information Society*, Vol. 50, no. 1: 23-41. Available at: <https://doi.org/10.16981/kliss.50.201903.23>.

- Chow, A. and Tian, Q. 2021. Public libraries positively impact quality of life: a big data study. *Public Library Quarterly*, Vol. 40, no. 1: 1-32. Available at: <http://doi.org/10.1080/01616846.2019.1632105>.
- Creaser, C. 2018. Assessing the impact of libraries: the role of ISO 16439. *Information and Learning Science*, Vol.119, no. 1/2: 87-93. Available at: <https://doi.org/10.1108/ILS-05-2017-0037>.
- Dresel, R., Henkel, M., Scheibe, K., Zimmer F. and Stock, W.G.. 2020. A nationwide library system and its place in knowledge society and smart nation: the case of Singapore. *LIBRI*, Vol. 70, no. 1: 81–94. Available at: <https://doi.org/10.1515/libri-2019-0019>.
- Garoufallou, E. and Gaitanou, P. 2021. Big data: opportunities and challenges in libraries, a systematic literature review. *College & Research Libraries*, Vol. 82, no. 3: 410. Available at: <https://doi.org/10.5860/crl.82.3.410>.
- Igarashi, T., Koizumi, M. and Widdersheim, M. 2020. Capturing citizens' information needs through analysis of public library circulation data. *LIBRI*, Vol. 70, no. 2: 127-141. Available at: <https://doi.org/10.1515/libri-2018-0137>.
- Institute of Museum and Library Services. 2023. *Public Library Survey*. Available at: <https://www.ims.gov/research-evaluation/data-collection/public-libraries-survey>.
- Joshi, P. 2015. Analyzing big data tools and deployment platforms. *International Journal of Multidisciplinary Approach & Studies*, Vol. 2, no. 2: 45–56.
- Kwon, N., Lee, J.Y. and Lee, J.Y. 2018. *A study on next-generation big data service operation plan*. Daejeon: Korea Institute of Science and Technology Information.
- Lawton, A. and Burns, J. 2015. A review of competencies needed for health librarians: a comparison of Irish and international practice. *Health Information and Libraries Journal*, Vol. 32, no. 2: 84-94. Available at: <https://doi.org/10.1111/hir.12093>.
- Lee, A. 2018. Measuring population reach of Singapore's libraries: the use of a Reach Index. *Journal of Library Administration*, Vol. 58, no. 2: 193-203. Available at: <https://doi.org/10.1080/01930826.2017.1392222>.
- Li, J., Lu, M., Dou, G. and Wang, S. 2017. Big data application framework and its feasibility analysis in library. *Information Discovery and Delivery*, Vol. 45, no. 4: 161-168. Available at: <https://doi.org/10.1108/IDD-03-2017-0024>.
- Pawar, A.M. 2016. Big data mining: challenges, technologies, tools and applications. *Database Systems Journal*, Vol.7, no. 2: 28-33.
- National Library of Korea. 2023. *Library Information Platform Homepage*. Available at: <https://data4library.kr/utilizationCase>
- National Library of Korea. 2016-2020. *Library big data use case book*. Seoul: National Library of Korea.
- National Library of Board. 2019. *The remaking of Singapore's public libraries*. Singapore: National Library Board.
- On, J.M. and Park, S.H. 2020. Big data analysis for public libraries utilizing big data platform: a case study of Daejeon Hanbat Library. *Journal of the Korean Society for Information Management*, Vol. 37, no. 3: 25–50. Available at: <https://doi.org/10.3743/KOSIM.2020.37.3.025>.
- Ong, G. 2014. *A smarter library: using data analytics to improve resource management and services at NLB*. Singapore: Civil Service College.
- Public Library Association. 2018. *Project Outcome*. Available at: <https://www.ala.org/pla/data/performance measurement>.
- Zhan, M. and Widén, G. 2018. Public libraries: roles in big data. *The Electronic Library*, Vol. 36, no. 1: 133-145. Available at: <https://doi.org/10.1108/EL-06-2016-0134>.