Research On Contemporary Issues in

MEDIA RESOURCES AND INFORMATION AND COMMUNICATION TECHNOLOGY USE

A Festschrift

in Honour of

PROFESSOR IYABO MOTOLAGBE MABAWONKU

Edited by

**WOLE MICHAEL OLATOKUN
AMOS OYESOJI AREMU
AIREN ADETIMIRIN**

Research On Contemporary Issues in

MEDIA RESOURCES AND INFORMATION AND COMMUNICATION TECHNOLOGY USE

A Festschrift

in Honour of

PROFESSOR IYABO MOTOLAGBE MABAWONKU

∽ঌ঺঻ৎ঺ৎ঴ঌ∾

Edited by
WOLE MICHAEL OLATOKUN
AMOS OYESOJI AREMU
AIREN ADETIMIRIN

CONTENTS

Big Data and the Emerging Role of Libraries

Oluwaseun I. Obasola

Abstract

Technological advancement has led to a massive increase in the amount of data being generated and accumulated daily. As a result, libraries no longer manage only print collections. They are also involved in managing materials that are born digital from an ever- growing range of sources using new technologies. This has led to the emergence of big data and new services in libraries. As the library's tradition changes, so also the research landscape changes. How is the library responding to this contemporary challenge for research? This chapter presents an overview of big data projects involving libraries. Some publications on big data projects were retrieved from databases such as LISTA and Google. Relevant websites and blog post were also consulted. The discussion focused on the emergence of big data and the library's emerging role in the research landscape. The review reveals the libraries' engagement with big data and challenges encountered, and recommends ways of addressing these challenges.

Introduction

The term 'Big data' is not a new term. Its existence dates back to 19th Century (1941) during the information explosion era when information leapt in bundles. Although the concept was not tagged with a name, there was an attempt to quantify the growth rate of the volume of data generated. In 1999, the term was officially tagged in a publication titled 'Visually Exploring Gigabyte Datasets in Real Time' by the Association for Computing Machinery. The following year (2000), Francis X. Diebold's presentation at the Eighth World Congress of the Econometric Society in the United Kingdom described the concept as an "... explosion in the quantity of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology." (B. Marr, 2015; G. Press, 2016).

According to some authors (L. Doug, 2001; M. Hilbert, 2015), 'Big data' is a concept with five attributes: volume, velocity, variety, veracity, and value. The first attribute – volume –refers to the size of the data being generated every day. As noted by the International Business Machines blog (R. Jacobson, 2013), "Every day, we create 2.5 quintillion bytes of data......" implying that a significant amount (90%) of data in the world today was created in the last two years alone. The second attribute – velocity – refers to the rate at which the data is being generated every second, and variety is about the type of data being generated and the lack of uniformity in the data structure. Veracity refers to the disorderliness or trustworthiness of the data, the value is the profit or benefit that can be derived from big data.

In Hoy's (2014) perspective, while data in simple term is the idea that computers can collect an enormous amount of information about billions of different things and find useful patterns in that information. With big data technology, information can be stored and used to make intelligent decisions for the good of the society. This is made possible with the help of distributed systems, stored in different locations, connected by computer networks. Some examples of big data generated every day include Twitter and Facebook posts, WhatsApp chats, call logs, student records, customer transaction details in supermarkets or banks, video clips on Youtube, pictures on Flickr, libraries' collection, factory production activities surveillance and navigation systems (Mutula, 2016)

In the library setting, big data refers to its ever-growing collections. Each information product in the library is data to a potential researcher depending on the context and the nature of the investigation. For example, a researcher using relevant articles in online academic databases for a systematic review or text mining changes the conceptualisation of published articles as research output to research datasets for scientific enquiry. Likewise, research outputs such as pictures, video clips and audio recordings from a particular investigation can also be re-used by another researcher in a different context as research dataset; hence, the description of big data in the library as information products such as print materials and electronic resources, like e-journals, e-books, library catalogues, images/pictures, magazines, qualitative/quantitative datasets, repositories and digitised materials. Others include those materials that are born digital like usage statistics, user details, chat on a library's social media platforms (blogs, Facebook page, twitter), video files, audio files and transaction logs with suppliers and publishers.

Zhan and Widén (2017) in their paper titled 'Understanding big data in librarianship' described the term (big data) as data sets with large size and fast-growing speed in different file formats which can complicate data handling techniques but can also boost the creation of technological solutions. Dumbill's(2013) description of big data as a complex dataset that cannot be processed using the conventional database systems indicates that more advanced systems are needed to derive substantial benefit from big data. Some of the tools used for big data analytics include Hadoop and Mahout. The application of these tools to harness the potential of big data has resulted in a shift in scientific practice leading to the advent of e-research, implying that big data underpins e-research (Emmot, 2006).

Gray's definition of e-research as a concept that exploits technologies for computation, data curation, analysis, visualisation and collaboration highlights the complexity involved in using big data for research. It also reveals the magnitude of issues surrounding the management, storage and preservation of big data. This calls for collaboration by research libraries, researchers, universities, funders, software and hardware technologists, national organisations and policymakers to harness the opportunities it offers (Soehner, Steve and Ward, 2010). This implies that the library has to

take on a new role in the research arena, changing the library's role into Library 4.0, a smart library that can analyse user needs and present findings intelligently. As such, big data is considered an essential concept for the development of future libraries (Noh, 2015).

As the scholarly culture moves towards online research, allowing investigations to cross disciplinary boundaries, there is a need to explore the role of libraries in big data analytics. What are the emerging roles of librarians and libraries in big data analytics? This paper presents some big data initiatives that have emerged amongst stakeholders to highlight the library's response to the change in the research landscape (Charles Henry, 2012). The study presents an overview of libraries' involvement in big data initiatives across the globe. It provides answers to the following research questions:

- What are the existing big data projects involving libraries?
- What are the emerging and current roles of librarians/libraries?
- What are the challenges of the emergence of big data in e-research?

Methods

The study procedure followed the review plan recommended by (T. Dyba, T. Dingsoyr, and G. K. Hanssen, 2007). The approach involved the following:

- Definition of search terms and strategies
- Identifying data source (databases) for literature search
- Literature searching using search terms and strategies
- Selection of relevant studies.

Search Strategy

To cover a wide range of publications in Library and Information Science(LIS), data sources such as Google and LISTA were used for literature searching. Relevant journals such as big data journal and blog posts on big data were also consulted. These data sources were selected to cover a significant proportion of relevant publications; as they were identified as relevant academic databases. Google search engine was included to ensure that all relevant documents were retrieved since most data sources are indexed by Google.

To ensure the search terms retrieve relevant articles for the study, the researcher and the research assistants tested the search terms developed using some pilot searches and the results obtained were vetted by other information professionals at the British Library in London and University of Ibadan Library System in Ibadan, Nigeria. Based on the research questions and initial searches during the planning stage, the following search strings were found suitable for the study:

'big data and librar* '

'data and librar*'

'data science and librar*'
'e-science and librar*'
'data project and librar*'
'big data projects and librar*'
'data initiatives and librar*'
'big data initiatives and librar*'

To construct the search string, all the search terms were combined by using 'AND' Boolean operator. The conjoining method was used to ensure the retrieval of relevant publications only. The search was conducted on the title and the abstract, to avoid missing out relevant publications.

Inclusion and Exclusion of Studies

Some of the studies retrieved were still irrelevant to the focus of the study despite the appearance of the keywords in the title or the abstract. Afterwards, a manual sorting of the publications retrieved was conducted to retain only publications which are relevant to the study. Only articles written in English language were selected for the review. Publications that met any of the following criteria were included in the review:

- Reports or Powerpoint presentation on big data projects involving libraries or librarians
- Studies that focus on big data projects involving libraries
- Studies that describe big data projects and highlighting the role of librarians/libraries.

Publications that do not meet the above criteria were excluded. Those excluded were:

- Studies that only mention the search term, but does not focus on big data project
- involving libraries; and
- Studies that focused on big data in general (projects that do not involve the library).

The protocol for identifying primary studies for this study was initiated in November 2016 and concluded at the end of November 2017. The search was limited to publication after 2000 up to 2017. The year 2000 was selected as a delimiter because preliminary search revealed no significant publications before this date.

These steps resulted in twelve citations which make up the final set of primary studies used for the investigation. For details on the search details and procedure for identifying the primary studies, see Tables 1-2 and Figure 1.

Table 1: Details of the search strategy

Database	Search String	Preliminary Articles Retrieved	Result after Reviewing the Title and the Abstract	Final Articles Collected
Google	big data and librar* data and librar* data science and librar* e-science and librar* data project and librar* big data projects and librar* data initiatives and librar*' big data initiatives and librar*	400	45	11
LISTA	big data and librar* data and librar* data science and librar* e-science and librar* data project and librar* big data projects and librar* data initiatives and librar*' big data initiatives and librar*	321	7	1

Figure 1: Procedure for screening articles recommended by T. Dyba, T. Dingsoyr, & G.K. Hanssen. (2007)

Literature search from Google and LISTA

400 Citations from Google

321 from LISTA

52 articles after screening the titles and removing duplicates

Screening of abstracts and full-text of 52 articles for eligibility

- 45 articles screened from Google and
- 7 articles screened from LISTA

Forty articles excluded

12 articles included

(8 articles, 3 websites and a blog)
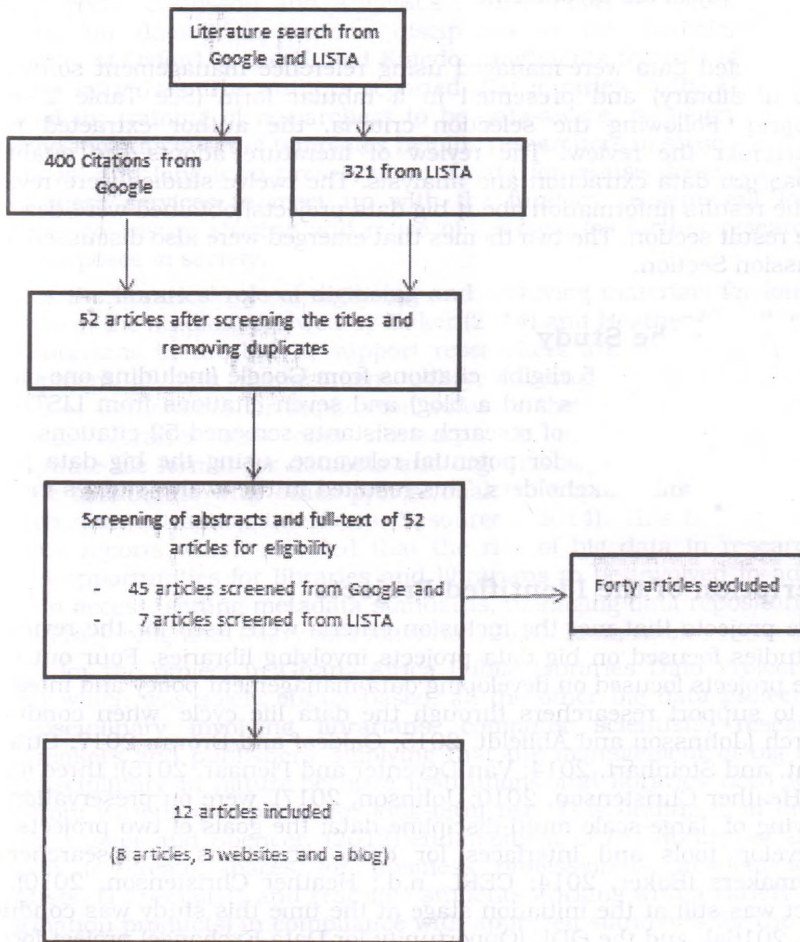
Data Extraction and Analysis

Electronic templates were used in extracting data from the 12 studies retrieved. The data extraction template was pre-tested with the first five studies included in the review and adjusted as considered necessary. Data extracted using the templates are highlighted as follows:

25

- Project title, author, date of publication;
- Details of the project including aim, activities and stakeholders involved;
- Location, setting;
- Project result/outcome.

The extracted data were managed using reference management software (a Mendeley library) and presented in a tabular form (See Table 2 in the Appendix). Following the selection criteria, the author extracted twelve citations for the review. The review of literature adopts the qualitative approach to data extraction and analysis. The twelve studies were reviewed and the results (information about big data projects) obtained were described in the result section. The two themes that emerged were also discussed in the Discussion Section.

Findings of the Study

The author retrieved 45 eligible citations from Google (including one citation each from three websites and a blog) and seven citations from LISTA. The author with the support of research assistants screened 52 citations. These abstracts were screened for potential relevance, using the big data project aim, activities and stakeholders. This resulted in the twelve studies included in the study.

Description of the Identified Projects

Twelve projects that met the inclusion criteria were used for the review. All the studies focused on big data projects involving libraries. Four out of the twelve projects focused on developing data management policy and integrated tools to support researchers through the data life cycle when conducting research (Johnsson and Åhlfeldt, 2015; Oakleaf and Brown, 2017; Strasser, Wright, and Steinhart, 2014; Van Deventer and Pienaar, 2015); three (CERL, n.d.; Heather Christenson, 2010; Johnson, 2017) were on preservation and archiving of large-scale multi-discipline data; the goals of two projects were to develop tools and interfaces for data extractions by researchers or policymakers (Baker, 2014; CERL, n.d.; Heather Christenson, 2010), one project was still at the initiation stage at the time this study was conducted (EIFL, 2016a) and the ODE (Opportunity for Data Exchange) project focused on an investigation of research data services in Europe. The projects were implemented in seven countries, namely the United States, the United Kingdom, Germany, Ghana, Singapore, South Africa and Australia (See Table 2 for more information about the projects). Forty potential papers on big data projects retrieved were excluded at full-text screening because the papers only mentioned the search term but did not involve any library as part of the project stakeholders.

Discussion

Libraries and Librarians Engagement with Big Data

The studies reviewed show that libraries are important stakeholders in big data projects. Johnsson and Åhlfeldt's (2015) report on DAMARO project focusing on data from various disciplines at the Bodleian Libraries, University of Oxford in the United Kingdom highlights the role of libraries in big data initiatives. The authors reported that libraries in the UK had been involved in: training of researchers to be data-aware, development of data curation tool kits (a set of templates helping researchers to structure the data workflow), and long-term preservation of data for re-use. Researchers need all of these services to meet up with the funders' requirement for data retention to ensure sharing and reuse of datasets for further research and other purposes in society.

Besides the libraries' role of digitising and archiving materials for long-term use, two of the studies reviewed by Baker (2014) and Heather (2010) showed that librarians in the bid to support researchers are now involved in the development of data management policy, investigator's toolkit and data models or frameworks for application software that promote easy access to datasets. Libraries also offer advisory role on metadata profile, an appropriate file format for datasets and organisation of datasets. Librarians likewise collaborate with other professionals to improve access to datasets through data extraction interfaces (Resources, 2014). This finding confirms previous reports which revealed that the rise of big data in research has created opportunities for libraries and librarians to be involved in advocacy for open access, setting metadata standards, managing data repositories and data literacy programs (Mary and Mavrinac, 2010; Rani, n.d.).

Except for the project in Ghana-Africa Public Libraries Data Project (EIFL, 2016a) which was solely library-based, all the other big data projects were multidisciplinary involving librarians, computer scientists, researchers, policymakers and publishers. Implying that the management of big data is multidisciplinary in nature because institutions and libraries need to store data from various disciplines for long-term use. This finding is in line with Jahnke, Asher and Spencer, (2012) endorsement of the library as an ideal location to create spaces for business professionals, academics, data scientists, IT specialists and graduate students working with a variety of data (information products) in compliance with approved standards for the benefit of the society.

The literature search retrieved only two documents on big data projects involving libraries in Africa (South Africa and Ghana). While the first project in South Africa focused mainly on the development of data management policy and repositories to support the use of data, the latter which involves data collection on public libraries in Africa was still at the initiation stage at the time the literature search was conducted. This is unlike Johnsson and Åhlfeldt report in the UK which revealed that most libraries in Europe (9 out of 13 cases in League of European Research libraries) are involved in data

management. Other authors also got similar reports in the US, Australia, South Africa and Singapore (Choudhury and Hanisch, n.d.; Heather Christenson, 2010; Johnson, 2017; Resources, 2014). Indicating that libraries in the developed countries are actively involved in the management of data in academic institutions and at the national level, it also revealed that the emergence of big data promotes the library's expertise in collection development and preservation of research datasets and other information products from research. According to the studies reviewed, libraries in the developed countries could take part in big data projects because of the funding and support received from funders and national research bodies (government) in each country. Libraries in the developing countries require similar support to survive the change in the research terrain.

Challenges of Big Data in Libraries

The studies reviewed have revealed some challenges of big data. Fuller (2015) and Zhan and Widen (2017) identified the development of meta-data templates as an issue with big data. The authors reported that due to the variety attribute of big data, developing a formal meta-data standard for big data is difficult because the formal dataset modelling processes are not applicable to big data. This makes the organisation, accessibility and managing of big data for use/re-use complex. This difficulty with managing big data was also acknowledged at the University of Pittsburgh (2007) with other challenges such as data capture, storage, timeliness, sharing, incompleteness, scalability and security issues.

Another challenge identified by this review is the need to adequately monitor big data analytics to avoid misleading correlation among data sets. Big data can misrepresent valid results or trends from data drawn from many sources. An example is the Flu Trends project by Google which exaggerated flu outbreak in over 25 countries in 2008. This review also acknowledges that the application of big data for e-research requires new skill sets and tools. The demand for data librarians is growing every day, and there are few professionals in the industry to fill the gap. Mutula in 2016 also reported this skill gap. Consequently, librarians need to acquire knowledge about big data analytics tools and develop data literacy skill to support users. Otherwise, the challenge may limit libraries involvement in big data analytics (Gordon-Murnane, L. 2012). To address this challenge, Library Schools in the United States and organisations in the United Kingdom have developed courses aimed at equipping information scientists with the requisite skill sets in data science, system analysis, data mining and programming, Artificial Intelligence, and more (Penn. nd; SAS Academy).

Besides the challenge with the skill gap and metadata standards, the studies reviewed also shows issues with data privacy or ownership of data, and the infrastructure to support big data in libraries. Because big data systems can collect an enormous amount of data about people, activities or an object, there is the possibility for damaging disclosure which may not be intentional on the part of the organisation holding the data. Many hackers work

tirelessly to break into data banks of prominent organisations; and once they succeed, there will be data breaches and a loss to such organisations and their customers (Hoy, 2014).

Conclusion and Recommendation

This review has shown that libraries and librarians have an essential role to play in managing big data and the delivery of other data services. As the demand for data services grows, librarians need to gain data curation and literacy skill sets to meet up with the need for data librarians. Library schools need to include training on data management in its curriculum to fill the skill gap in the information industry. Library schools may also have to collaborate with data centres in research institutes or academic institutions to evolve training modules on data management to equip future librarians with the required skills and tools for data services.

Librarians likewise need to collaborate with other professionals to develop data standards and metadata profile to support long-term preservation of data for knowledge creation in the emerging information terrain.

To ensure the use of big data is appropriately monitored to maximise its gains, librarians and libraries need to get involved in offering data advisory services to promote healthy use of data, and educating data users on the implication of data abuse or breaches. Governments and institutions need to support libraries and other knowledge organisations to evolve data standards and policies that will improve the use of big data in businesses and researches and promote healthy data sharing to maximise the gains of big data in different sectors of the society.

References

Marr, B. (2015). Why only one of the 5 Vs of big data really matters. *IBM Big Data and Analytic Hub.* Available at: http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters. Accessed August 2016.

Dumbill, E. (2013). Making Sense of Big Data. *Big Data, 1*(1), 1-2. doi:10.1089=big.2012.1503. Available at: https://www.liebertpub.com/doi/10.1089/big.2012.1503. Accessed December 2016.

Press, G. (2016). A very Short History of Big data. Big Data Trend. Available at: https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#63dbd1565a18 . Accessed March 2017.

Doug, L. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *META Group Research, 949.* Available at : https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Accessed January 2017.

Mavrinac, M. A. (2010). Transforming Libraries: Big data.*News letter of the University of Toronto Mississauga,9*(1).Available at: https://tspace.library.utoronto.ca/bitstream/1807/24942/1/BigData.pdf . Accessed July 2017.

Hilbert, M. (2015). Big Data for Development: A Review of Promises and Challenges. . *Development Policy Review., 34,* 135-174. Available at : https://onlinelibrary.wiley.com/doi/full/10.1111/dpr.12142 . Accessed May 2017.

Hoy, M.B. (2014). Big Data: An Introduction for Librarians. *Medical Reference Services Quarterly, 33*(3), 320-326. Available at : https://www.ncbi.nlm.nih.gov/pubmed/25023020 . Accessed July 2017.

Dyba, T. Dingsoyr, T. and Hanssen, G.K. (2007). Applying systematic reviews to diverse study types: An experience report.P*aper presented at the Proceedings of the First International Symposium on Empirical Software Engineering and Measurement.* Available at :https://ieeexplore.ieee.org/document/4343750/ . Accessed April 2017.

Baker, J. (2014). The British Library Big Data Experiment. Retrieved from http://blogs.bl.uk/digital-scholarship/2014/06/the-british-library-big-data-experiment.html.

CERL, U. of C. and the. (n.d.). The Digital South Asia Library. Retrieved May 9, 2018, from http://dsal.uchicago.edu/about.html.

Choudhury, S. and Hanisch, R. (n.d.). *The Data Conservancy□: Building a Sustainable System for Interdisciplinary Scientific Data Curation and Preservation.* Retrieved from https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/34018/47_Choudhury_DataConservancy.pdf.

EIFL. (2016a). Africa Public Libraries Data Project. Retrieved May 9, 2018, from http://www.eifl.net/events/africa-public-libraries-data-project.

EIFL. (2016b). Africa Public Libraries Data Project. Retrieved April 26, 2018, from http://www.eifl.net/events/africa-public-libraries-data-project.

Gordon-Murnane, L. (2012). Big data: A big opportunity for Librarians.

Online, *36*(5), 30–34.

GSLIS Center for Informatics Research in Science and Scholarship (CIRSS) University of Illinois. (n.d.). Retrieved from http://cirss.lis.illinois.edu/

Heather Christenson. (2010). Hathi Trust: A Research Library at Web Scale. *Lrts, 55*(2), 93–102.

Hoy, M. B. (2014). Big Data□: An Introduction for Librarians. Matthew B. Hoy and Tara Brigham , Column Editors Big Data□: An Introduction for Librarians, (August), 37–41. https://doi.org/10.1080/02763869.2014.925709

Johnson, V. (2017). Leveraging Technical Library Expertise for Big Data Management. *Journal of the Australian Library and Information Association, 66*(3), 271–286. https://doi.org/10.1080/24750158.2017.1356982

Johnsson, M. and Åhlfeldt, J. (2015). Research libraries and research data management within the humanities and social sciences. Project report. Retrieved from http://lup.lub.lu.se/record/5050462.

Mary, B. and Mavrinac, A. (2010). Newsletter University of Toronto Library, (1), 9–11.

Mayernik, M. S. Choudhury, G. S., DiLauro, T., Metsger, E., Pralle, B., Rippin, M. and Duerr, R. (2012). The data conservancy instance: Infrastructure and organizational services for research data curation. *D-Lib Magazine, 18*(9–10). https://doi.org/10.1045/september12/mayernik

Mcdonald, R. H. (2015). Academic Libraries and Big Data. Retrieved from: https://www.slideshare.net/rhmcdonald/academic-libraries-and-big-data-trends-in-collection-publication-preservation-and-access.

Mutula, S. (2016). Matula S Big Data_17102018_155757.pdf. *Afr. J. Lib. Arch. & Inf. Sc, 26*(2), 093-96.

Oakleaf, M. and Brown, M. (2017). Institutional Learning Analytics: How Can Academic Libraries Connect? - YouTube. *Coalition for Networked Information.* Retrieved from https://www.youtube.com/watch?v=hT42hzoNDXI.

Penn. State College . Information Science and Analytics, Modeling and Informatics. Retrieved from: https://ist.psu.edu/node/1633.

University of Pittsburgh (2007). Big Data Analytics Specialisation. Retrieved from:http://www.ischool.pitt.edu/ist/degrees/specializations/big-data.php.

Jacobson, R. (2013). IBM Blog. Retrieved from https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/

Rani, B. R. (n.d.). Big Data and Academic Libraries. Retrieved from http://km.ptar.uitm.edu.my/documents/10180/1099975/Big+data+and+Academic+Libraries.pdf/74bbc20c-cfc0-4cae-ba4b-ead791b8f406?version=1.0.

Reinhalter, L. and Wittmann, R. J. (2014). The Library: Big Data's Boomtown: Edited by Rick J. Block. *Serials Librarian, 67*(4), 363–372. https://doi.org/10.1080/0361526X.2014.915605

Resources, C. O. F. (2014). *Business Information* (Vol. 2). Retrieved from http://www.asiandatascience.com/wp-content/uploads/2014/09/BI_Asean_Sept_2014_FINAL.pdf.

SAS Academy for Data Science , UK. Certified Data Scientist with Self-Paced elearning. Retrieved from: https://www.sas.com/en_gb/training/academy-data-science.html.

Strasser, C., Cook, R., Michener, W. and Budden, A. (2012). Primer on Data Management☐: What you always wanted to know. *DataONE*, 1–11. https://doi.org/10.5060/D2251G48.

Strasser, C., Wright, S. and Steinhart, G. (2014). DataONE for Librarians, 1–6. Retrieved from https://www.dataone.org/for-librarians.

Van Deventer, M. and Pienaar, H. (2015). Research Data Management in a Developing Country: A Personal Journey. *International Journal of Digital Curation, 10*(2), 33–47. https://doi.org/10.2218/ijdc.v10i2.380.

York, J. (2009). This library never forgets: Preservation, cooperation, and the making of HathiTrust Digital Library. *Archiving 2009: Final Program and Proceedings, 2009*(1), 5–10.

Zhan, M. and Widén, G. (2017). Understanding big data in librarianship. *Journal of Librarianship and Information Science*, 096100061774245. https://doi.org/10.1177/0961000617742451.

Zilinski, L. D., Barton, A., Zhan, T., Pouchard, L. and Pascuzzi, P. (2016). RDAP Review: Research Data Integration in the Purdue Libraries. *Bulletin of the Association for Information Science and Technology, 42*(2), 33–37. Retrieved from https://www.asist.org/publications/bulletin/decemberjanuary-2016/rdap-review-research-data-integration-in-the-purdue-libraries/

Oluwaseun Obasola is the System Librarian at E. Latunde Odeku Medical Library, College of Medicine, University of Ibadan, Oyo State, Nigeria. She holds a Bachelor of Science degree in Computer Science with Economics from Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria; a Master's degree and PhD in Library, Archival and information Studies from the Library School, University of Ibadan, Oyo State, Nigeria.

Appendix

No	Project Title and Initiation Date	Project Objective(s)	Stakeholders	Location	Outcome	Authors
	Table 2: Overview of Big Data Projects Involving Libraries					
1.	DAMARO 2011-2013	Development of data management policy and integrated tools to support researchers through the data life cycle	Bodleian Library, University of Oxford and JISC(Joint Information Systems Committee)	United Kingdom (UK)	Data Policy Integrated tool for Data management	(Johnsson and Åhlfeldt, 2015)
2.	Purdue Library Data Project 2000	To develop a set of templates (a data management plan) that will help researchers to structure the data workflow in research	Purdue University Library and International Museum and Library Services (IMLS)	United States (US)	Data Curation Profiles Toolkit	(Oakleaf and Brown, 2017; Reinhalter and Wittman, 2014; Zilinski et al, 2016)
3.	DataONE Project 2009	The goal of the project is to preserve and provide access to multi-scale, multi-discipline, and multi-national data. Stakeholders The project provides scientific data archiving for ecological and environmental data produced by scientists.	National Science Foundation, NSF Computer and Information Science and Engineering Directorate (CISE), NSF INTEROP Programs, Leon Levy Foundation, National Aeronautics and Space Administration (NASA), Microsoft Research, Gordon and Betty Moore	US	Data Primer Librarian Outreach Toolkit *Investigator Toolkit*	('GSLIS Center for Informatics Research in Science and Scholarship (CIRSS) the University of Illinois,' n.d.; Strasser, Cook, Michener, and Budden, 2012; Strasser, Wright, and Steinhart, 2014)

No	Project Title and Initiation	Project Objective(s)	Stakeholders	Location	Outcome	Authors
			Foundation and United States Geological Survey (USGS). Other stakeholders 11 working groups and user groups involving libraries/libra rians and academic institutions and research centres.			
4.	HathiTrust Digital Library. 2008	The project focuses on large-scale digitization of the 'big data' in 25 libraries to ensure long, term preservation of their content.	University of Michigan, Indiana University, libraries of the Committee on Institutional Cooperation, University of California and the University of Virginia	US	A Digital Library preserving and archiving the holdings of 25 libraries	(Heather Christenso, 2010; Mcdonald, 2015; York, 2009)
5.	Data Conservan cy Project	The goal of the project is to broaden data infrastructure development and gain better understanding of the challenges in collecting, preserving and curating different types of research data.	Sheridan Libraries at Johns Hopkins University Sloan Digital Sky Survey data and US National Virtual Observatory	US	A data-centric architecture, discipline-agnostic data model and a data feature extraction framework that facilitates data integration and cross-disciplinary queries	(Choudhury and Hanisch, n.d.)
No	Project Title and Initiation	Project Objective(s)	Stakeholders	Location	Outcome	Authors

34

	Date					
6.	ODE - OPPORTUNITIES FOR DATA EXCHANGE 2010	To investigate emerging best practices in sharing, re-using, preserving and citing data, the drivers for these changes and barriers impeding progress	European Organization for Nuclear Research (CERN), Alliance for Permanent Access (APA), CSC - IT Centre for Science, Alfred Wegener Institute for Polar and Marine Research (AWI), Science and Technology Facilities Council (STFC), The British Library, German National Library, Association of European Research Libraries (LIBER), and International Association of Scientific, Technical and Medical Publishers (STM)	Germany	Three reports, 1. Ten Tales of Drivers and Barriers in Data Sharing 2. Integration of Data and Publications 3. A final report on the project	ODE website http://www.dnb.de/EN/Wir/Projekte/Archiv/ode.html
7.	The Digital South Asia Library	To provide access to important and rare resources on South Asia to scholars, public officials, business leaders, and other users	University of Chicago, the Centre for Research Libraries, academic institutions, libraries and organisations in the US, Europe, South Asia	US	Digital Library providing access to rare materials on South Asia	(CERL, n.d.)

No	Project Title and Initiation Date	Project Objective(s)	Stakeholders	Location	Outcome	Authors
			and Australia			
8.	Africa Public Libraries Data Project 2016	The aim of the project is to gather data about the presence, use, services and resources of public and community libraries in Africa. It seeks to develop research methods for sustainable and ongoing data collection about public and community libraries in Africa.	EIFL(Electronic Information for Libraries), African Library and Information Associations and Institutions (AfLIA) and the Technology and Social Change Group, University of Washington Information School (TASCHA),	Ghana	Project is still at the initiation stage.	EIFL Website (EIFL, 2016b)
9.	British Library Big Data Experiment 2014	The focus is to open up digital collections at the British Library to researchers in Arts and Humanities.	British Library, UCL Computer Science, and UCL Centre for Digital Humanities	UK	Platforms or interfaces that link researcher and allow interrogation of BL's public domain digital collections using the Microsoft Azure cloud infrastructure	BL's blog (2014) (Baker, 2014)
No	Project Title and Initiation Date	Project Objective(s)	Stakeholders	Location	Outcome	Authors
10.	Shell Australia Technical Library	To develop a Geophysical Database Management	Shell Library staff and Technical Data	Australia	Robust Database for storing and managing	(Johnson, 2017)

	2016	System	Management Team involving geophysical data analysts, geological data analysts, IT specialists, project managers and workflows/applications specialists		Geophysical data	
11	Singapore Library Data Mining Project 2013	To optimize the library's collection to improve access to information resources using big data technology (Hadoop)	National Library of Singapore and 25 branch libraries	Singapore	A title recommendation service that offers more relevant data to its users via its websites; Mining of reading patterns based on the libraries' past records; Enhancement of the search and discovery experience of users and improvement in the use of digitised resources.	(Resources, 2014)
12	SARIS Project 2004	Improve access to electronic research literature	Ford Foundation, University of Pretoria Library, Council for Scientific and Industrial Research (CSIR), Data Intensive Research Initiative of South Africa	South Africa	Data management Data management policy Data repository	(Van Deventer and Pienaar, 2015)

			and the Network of Data and Information Curation Community (NDICC)				2016 (TCL Wegwijs Artikel 2016)