Data Librarianship: Open Data Awareness, Perceptions and Services in Medical Libraries in Nigeria

Ebele N. Anyaoku
[1]Medical Library, Nnamdi Azikiwe University, Awka, Nigeria.
[2]University Library, PAMO University of Medical Sciences, Port Harcourt, Nigeria (Sabbatical).

Abstract

Background*: Open data is a relatively new initiative and its management and access is a very new service for many libraries, requiring development of new skills through training.*

Purpose/Setting/Participants: *The research assessed open data awareness, perceptions and services in medical libraries in Nigeria. Fifty – one medical librarians from 36 medical education institutions and hospitals responded to the survey.*

Methods: *Questionnaire was the instrument for data collection. The questionnaire was created in SurveyMonkey and distributed through the email list and WhatsApp group platform of the Medical Library Association of Nigeria (MLA-N).*

Results/outcomes*: Results showed that a high percentage of the medical librarians are aware of the open data initiative, they are also aware and have a basic to extensive understanding of open data values and benefits. Over half are aware of the various topics in open data management. The medical librarians showed positive attitude to data sharing and librarians' involvement in data management services. However, the awareness and positive attitudes did not translate to data management. Data management services are yet to be implemented in the libraries.*

Conclusions/recommendations*: The paper recommends that data resource collection development should be integrated into the library collection development policies and processes. It concludes that the future and prospects for data management in medical libraries in Nigeria may depend on a well-developed data librarianship training programme.*

Keywords: Open data, Open movement, Open data management, Data services, Research Data Management, Medical Libraries.

Introduction

The drive for unrestricted access and freedom to use intellectual and knowledge resources has led to the various open movements such as open science, open research, open educational resources and many others. These initiatives advocate free access to information resources such as research reports, and academic literature to the public at no cost. Open data is a relatively new initiative and its management and access is a very new service for libraries which have served the information needs of the society for centuries.

Data are symbols, numbers, alphabets, and figures, stored and presented in a formalized mode, derived from observations or measurements suitable for interpretation, or processing by humans or automatic means. Data are the raw material for actionable information and knowledge creation. Datasets like other intellectual outputs can be copyright protected. Licensing may be proprietary with usage and sharing restrictions placed on end users by the copyright owner thereby restricting access to vital resources that may be required for societal sustainability and development.

The increasing permeation of the open movement and the accompanying technological advances that support information access and sharing propelled the emergence of open data movement. Open data are data that reside in the public domain or have open license with no-cost access and licensing restriction; they are available for free to the public with no usage and sharing restrictions. Furthermore, *Open Data Charter* defined open data as digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere. Attributes of open data include availability, accessibility, reusable, unrestricted access, free sharing and redistribution by others. Constituent of open data include open research data, open health data, open government data and other scientific and social indices data.

Proponents of open data note that they can empower governments, citizens, civil societies and private sector organizations to work toward better outcomes for public services in areas such as health, education, public safety, environmental protection, human rights, and natural disasters (Open Data Charter). Linde, Wessels, Sveinsdottir, and Noorman (2014) stated that unrestricted and digitally facilitated access to data would enable faster progress in science through minimizing duplication of effort and offering scientists a wider range of data to use for re-analysis, comparison, integration and testing. It would contribute to improve the way science and scientific data can be used in relation to social goals.

Open Health Data

Healthcare is data intensive. Public health data are often aggregated information that can help in making decisions regarding health issues (D'Agostino, 2018). Health data are amassed from numerous sources such as electronic health records (EHRs), research studies, generic databases, medical imaging, patient portals, genomic sequencing, pharmaceutical research, public records as well as government data produced or commissioned by government (NEJM Catalyst, 2018). These, as Martin et al. (2017) noted, provide opportunities for ongoing health research, using a mine of research data that are potentially useful for assessing and informing the health of nations.

Many benefits are associated with open data in healthcare. Proponents note that the availability of open data can empower citizens and support clinicians, care providers, and researchers make better decisions, and stimulate new innovations (eHealth Ireland). Verhulst (2014) highlighted the following value propositions for using more open data in healthcare:

- Accountability: The use of data to hold healthcare organizations and providers accountable for treatment outcomes.
- Choice: Providing open data to help patients make informed choices from among the healthcare options available to them.
- Outcomes: Improving treatment outcomes by using open data to make the results of different treatments, healthcare.
- Economic growth and innovation: Using open data to fuel new healthcare companies and initiatives, and to spur innovation, new discoveries.

Data Management Services in Medical Libraries

Traditionally, the main duties of medical librarians are to collect, process and disseminate relevant, accurate and up-to-date information in print and electronic formats for healthcare and research. However, technologies are expanding and creating new responsibilities as well as modes of services for medical librarians. Such emerging responsibility is that of data management services. Medical libraries as health information disseminating institutions have the potential to manage and disseminate public health data. This management of data goes beyond research data emanating from the parent institutions; it also includes other datasets relevant to healthcare.

Galetto (2016) defined data management as an administrative process that includes acquiring, validating, storing, protecting, and processing required data to ensure the accessibility, reliability, and timeliness of the data for its users. The work of data management covers factors such as how to

- Create, access, and update data across a diverse data tier
- Store data across multiple clouds and on premises
- Provide high availability and disaster recovery
- Use data in a growing variety of apps, analytics, and algorithms
- Ensure data privacy and security
- Archive and destroy data in accordance with retention schedules and compliance requirements (Oracle, nd.)

Data librarianship is an emerging field. Many librarians have extended their services through involvement in the collection, curation and creating access to datasets. Burnett (2013) noted that librarians are becoming more involved in the development of principles and best practices for managing digital data for long term use. They have highly relevant information standards and organisational skills, including expertise in setting up file structures, knowledge of work flows and collection management, describing data in accordance with established metadata schemes and controlled vocabulary, collection curation / preservation and service provision in form of help desks, training, availability of subject specialists etc.

Other Librarians' roles as listed by *Foster* in open data management include: advocating and raising awareness, providing infrastructures to share data, and training and supporting researchers to open their research workflows.

Librarians are highly involved in information literacy training, open data management can be incorporated into the programme to inculcating data literacy which can lead to increased willingness to data sharing and use. Cliggett (2013) pointed out a number of barriers qualitative researchers cite when considering archiving and data sharing. Among these are concern for privacy, data ownership and challenges for interpretation, protection of study participants, sensitive data and confidentiality. In their role as data literacy instructors, librarians are proponents of data privacy. They are the professionals who should point out to researchers all the ways that qualitative data – apparently "anonymised" – can become less anonymous once matched with other data.

Literature suggests that librarians are increasingly embracing the new role of data services. Many libraries are fully utilizing technology to create awareness and offer the service to their users. This is attested by Yoon and Schultz (2017) who examined the research data services in academic libraries in the United States through a content analysis of 185 library websites. The results showed that there is a wide variation among library data management services and programs according to their web presence.

Medical librarians in Nigeria have the potential to be advocates of open data for the public good. The major objective of the study therefore was to assess the level of open data awareness, perceptions and services offered in medical libraries in Nigeria. This will show their readiness and contributions to data management. The following research questions guided the study.

Research Questions

1. What is the level of awareness of medical librarians of open data concepts and resources?
2. What are the data services offered in medical libraries?
3. What is the attitude of the medical librarians to research data sharing?
4. What is their attitude to data management services?
5. What are their open data training needs?

Methods

The research surveyed medical librarians in Nigeria. Questionnaire was used for data collection. The questions were developed based on literature review. However, question one (Awareness and Level of Understanding of Some Open Data Concepts) was adopted from the GODAN Action Open Data Management e-learning course questionnaire (https://www.collaborateandlearn.org). The options were modified.

The questionnaire was created in SurveyMonkey tool (http://www.surveymonkey.com/) and distributed through the email list and WhatsApp group platform of the Medical Library Association of Nigeria (MLA-N). Fifty – one medical librarians from thirty - six medical education institutions and hospitals responded to the survey. Greater response was obtained from the WhatsApp social media collector (44) than Email invitation (7). Results were analyzed using simple percentages. The results are presented in tables.
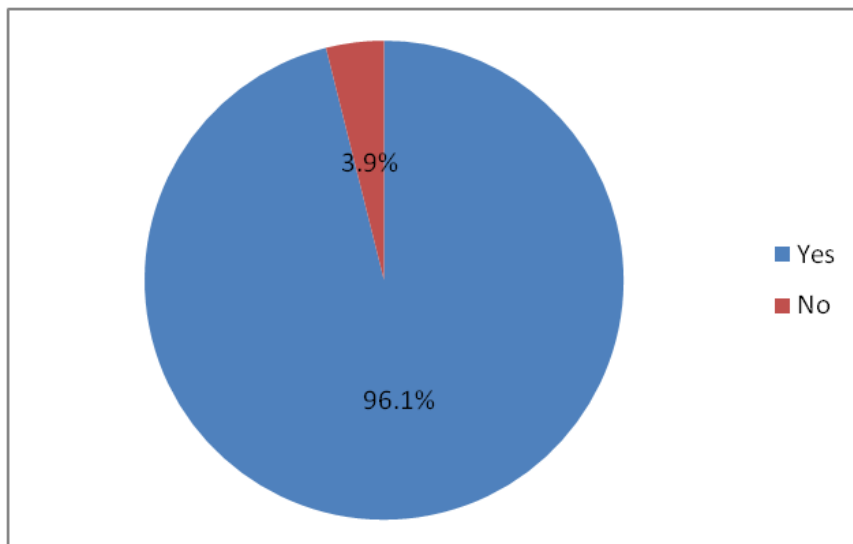
Results

Respondents' demography

Analysis shows that 54% (27) of the respondents are female, 46% (23) are male. In terms of educational qualifications, 54% (27) have a Master's degree; 26% (13) have a Bachelor's degree, and 20% (10) have a PhD. For work experience, 30% (15) had worked 6 - 10 years; 26% (13) had worked for 11 - 15 years; 22% (11) had worked 21 years or more, 14% (7) had worked 1-5 years, and 7.8% (4) had worked 16 – 20 years. One respondent skipped the demography question.

Awareness of Open Data Initiatives

Figure 1: Medical Librarians Awareness of open data initiative

Respondents were asked to indicate their awareness of open data initiative. Results as shown in figure 1 indicate a high awareness of open data initiative. Almost all 49 (96.1%) claim awareness while 2 (3.9%) indicated no awareness.

Awareness and Understanding of Open Data Concepts

Respondents awareness of open data concepts and resources are shown in Tables 1 and 2.

Table 1: Awareness and Level of Understanding of Some Open Data Concepts

	I understand the topic extensively	I have a basic understanding of the topic	I am aware of the topic but have no understanding	I am not aware of the topic
Value and benefits of open data	17 (33.3%)	22 (43.1%)	9 (17.7%)	3 (5.9%)
Open data Standards and Vocabularies	5 (9.8%)	21 (41.2%)	21 (41.2%)	4 (7.8%)
Open data licensing	6 (11.8%)	20 (39.2%)	18 (35.3%)	7 (13.7%)
Open data curation	6 (11.8%)	15 (29.4%)	22 (43.14%)	8 (15.7%)
Data citation	12 (23.5%)	23 (45.1%)	13 (25.5%)	3 (5.9%)

N=51

As shown in Table 1, most librarians (39, 76.5%) are aware and understand the value and benefits of open data. Most (35, 68.6%) have some understanding of data citation. Equal numbers, 26 (51%) have some understanding of open data standards/vocabularies, and also open data licensing, but less than half, (21, 41.2%) have some understanding of open data curation.

Table 2: Awareness and Access of Open Data Resources

	Aware and accessed	Aware but have not accessed	Not Aware at all	Missing
re3data	8 (16%)	17 (34%)	25 (50%)	1
DataCite	14 (28%)	22 (44%)	14 (28%)	1
dataMED	18 (36%)	19 (38%)	13 (26%)	1
Dryad Data Repository	6 (12%)	18 (36%)	26 (52%)	1
Global Health Observatory Data Repository - World Health Organization	17 (34%)	22 (44%)	11 (22%)	1

Some medical related open data resources and repositories were listed and respondents were requested to indicate awareness and access of the websites. As shown in Table 2, the best known were the Global Health Data Observatory (WHO) and DataMED. For example, 39,

(78%) of the respondents claim awareness of the Global Health Data Observatory, and out of this group 17(34%) claim to have accessed the resource. Two resources, re3data and the Dryad Data Repository were little known, with DataCite in the middle position (22, 44% of the respondents were aware, but had not accessed the resource).

Data Services.

To find out data services offered in the 36 medical libraries represented in the survey, responses from individuals from the same institutions were merged. Result shows that for the options provided, the only positive indication was from 17 respondents from 14 individual libraries who claimed that the data service they offer is the availability of an institutional repository that supports data services. By contrast, 39 respondents from 20 libraries indicated they do not offer data services in their libraries. None of the respondents indicated depositing datasets or researchers' data in the repositories. None indicated offering data services such as advocating or training researchers to deposit their data. This implies that these libraries are yet to commence data management services.

Despite this, respondents indicated certain data services they are capable of offering to support open data use in their institutions (Table 3).

Table 3: Services Medical Librarians Can Perform To Support Open Data Initiatives

Data management services	Frequency	%
Locating datasets for patrons	25	50
Identifying and preparing datasets for deposit into the repository	9	18
Creating metadata for datasets.	8	16
Participating with researchers on a research and data management project	18	36

Attitude of Medical Librarians to Data Management Services

The attitude of medical librarians to research data sharing and services (Tables 4 and 5) are largely supportive of data sharing. A large number, 42 (85.7%) agreed to make their data openly available. Similarly, only 13 (26.5%) agreed that research data are personal and confidential property and as such should not be shared. Again, only 14 (28.6%) agreed that research data can be subject to manipulation by owners, so reuse should not be encouraged. This indicates that majority rejected the negative statements. Contrarily, slightly over half 26 (53.1%) agreed that shared data can be misused or misinterpreted by others.

Table 4: Librarians Attitude to Data Sharing

Data Sharing	Strongly Agree / Agree	Neither	Strongly Disagree / Disagree	Missing
I am willing to make my research data openly available	42 (85.7%)	6 (12.2%)	1 (2.00%)	2
Research data are personal and confidential property it should not be shared.	13 (26.5%)	2 (4.1%)	34 (69.4%)	2
Research data can be subject to manipulation by owners, so reused should not be encouraged	14 (28.6%)	9 (18.4%)	26 (53.1%)	2
Shared data can be misused or misinterpreted by others	26 (53.1%)	6 (12.2%)	17 (34.7%)	2

Table 5: Attitude to Data Management Services in the Library

Data Services	Strongly Agree / Agree	Neither	Strongly Disagree/ Disagree	Missing
I am willing to locate datasets for users as part of my duties	48 (98%)	1(2%)	0	2
Institution researchers should be mandated to store their research data in library repositories	44 (89.8%)	5 (10.2%)	0	2
Data management is not part of the librarians' duties	2 (4.1%)	2 (4.1%)	45 (91.8%)	2
Data management is too technical; many librarians will not understand the processes	10 (20.4%)	1 (2. %)	38 (77.6%)	2
Our users do not need data so it will be a wasted effort in its collection management	2 (4.1%)	0.00	47 (95.9%)	2

Respondents' opinions were sought on various issues pertaining to medical librarians' involvement in data management. Almost all respondents (98%) are willing to locate datasets for users as part of their duties (Table 5). The majority (89.8%) agreed that institution researchers should be mandated to store their research data in repositories managed by the library. They were optimistic about the purposes of data management (96% disagreed that that data management was wasteful) and their competence in handling data, most disagreeing that data management was too technical, or beyond their understanding.

Table 6: Training Needs on Open Data

	Frequency	%
Open data management	35	71.4%
Using open data in research	35	71.4%
Locating and searching open data repositories	30	61.2%

However, Table 6 shows that large percentage (>60%) of the respondents acknowledge they need training on open data management in libraries.

Discussion

The study sought to find out the open data awareness, perceptions and services performed in medical libraries in Nigeria. Result shows that large majority (96.1%) of the respondents claim to be aware of open data and have some understanding of the value and benefits. About half of the respondents indicated awareness and understanding of various components of open data management such as open data standards and vocabularies, open data licensing and open data citation.

For awareness of open data resources and repositories, many claim awareness of the resources listed. However, experience in accessing the resources was generally much lower. Tenopir et al. (2014) in a similar high positive finding, noted that library directors and librarians who are aware of research data management issues more likely responded to the survey, so results possibly showed an inflated picture of research library involvement in research data services. The same may be true for the results of this research, considering that majority of the respondents (74%) have a postgraduate level of education and are therefore more likely to have encountered the concepts.

However, the result is significant because it reveals that a good number of medical librarians in Nigeria are observant of global developments in the field of open data management.

Despite the awareness of open data management, the medical librarians are yet to commence data services in their various libraries. However, some respondents indicated availability of institutional repositories which is a basic technical infrastructure for data management.

Findings showed that the respondents have a positive attitude to data sharing and librarian involvement in data services. Large majority (85.7%) agree that they are willing to make their data openly available. A majority disagreed with the statement that research data are personal and confidential property it should not be shared. They also disagreed that research data can be subject to manipulation by owners, so reuse should not be encouraged.

The willingness to make their data openly available is commendable if they can substantiate it by initiating the processing and depositing of their research data into their institutional repositories. However, it has been noted that there is an inverse relationship between favourable dispositions to data sharing and taking positive action to share the data (Nelson, 2009, Berghmans et al., 2017). Berghmans et al. (2017) observed that researchers acknowledge the

benefits of open data, but data sharing practices are still limited. Reasons include: fear of work being scooped or misused, not enough training in data sharing, unclear rewards, research data management and privacy issues, proprietary aspects, confidentiality and ethics. Librarians need to be aware that researchers may avoid some research data management activities, just as users have avoided depositing research reports and publications into repositories. Librarians may need to assist users with aspects such as confidentiality and ethics.

Respondents' opinions were sought on various issues pertaining to medical librarians' involvement in data management. Almost all respondents (98%) are willing to locate datasets for users as part of their duties, most agreed that institution researchers should be mandated to store their research data in library repositories, and most disagreed with the assertion that data management is not part of the librarians' duties. Most respondents seemed to agree that data management was an activity needed by their users. The awareness and positive attitudes attest to the medical librarians' acceptance of the new responsibility of data services, and the role of data librarian. It may be correct to assume that these health information professionals, if provided with enabling policies and resources can join their counterparts in other part of the world in developing data services.

Many of the medical librarians are willing to be trained on open data management services. This is probably what they need to empower them to launch data services in their libraries. The training could be formal or informal, and Barbrow et al. (2017) present a range of options for North American librarians that could be adapted for Nigeria.

Recommendations

Based on the findings of the study, the following recommendations are made:

1. Medical librarians should innovatively advocate for data services in their libraries. They should create awareness and promote use of data resources including through information literacy programmes.

2. They should integrate data resources into their collection development policy. Open datasets should be managed like open access collections. Bibliographic records for open data resources should be included in the library catalogues. Links should be created to open data sources on their websites.

3. They should endeavour to attend training on data management. They can participate in online trainings/webinars organized by different stakeholders. Importantly, library schools in Nigeria should incorporate open data management in their curriculum. This is urgently needed to inculcate open data knowledge and skills.

Conclusion

Following the increasing awareness of the importance of open data for research and decision making in healthcare, there is need for data management services. The involvement of medical librarians in the process is imperative. This research shows medical librarians in Nigeria are yet to take up this responsibility fully. Librarians need to develop information systems that will support their roles in collecting, preserving, and creating access to healthcare data in all ramifications. It is important to note that the future and prospects for open data management in libraries in Nigeria depend on a well-developed data librarianship training programme.

References

Barbrow, S., Brush, D. and Goldman, J. (2017). Research data management and services: Resources for novice data librarians. *College & Research Libraries News*, *78*(5), p.274.

Berghmans S, Cousijn, H., Deakin, G. Meijer, I. Mulligan, A. Andrew P. et. al. (2017). *Open data: The researcher perspective.* https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf

Burnett P. (2013). *What is the role of a librarian in Research Data Management*? http://blog.inasp.info/research-data-management-role-librarians/

D'Agostino, M, Samuel, NO, Sarol, MJ, de Cosio, FG, Marti, M, Luo, T. et al. (2018) Open data and public health. *Rev Panam Salud Publica,* 42:. https://doi.org/10.26633/RPSP.2018.66

eHealth Ireland (nd.). *Open Data for Health*. https://www.ehealthireland.ie/Strategic-Programmes/Open-Data-for-Health/

Cliggett, Lisa, (2013). Qualitative Data Archiving in the Digital Age: Strategies for Data Preservation and Sharing. *Anthropology Faculty Publications. 1.* https://uknowledge.uky.edu/anthro_facpub/1

Foster (nd.). *Libraries: roles and opportunities on Open Science.* https://www.fosteropenscience.eu/content/libraries-roles-and-opportunities-open-science

Galetto, M. (2016). *What is data management?* https://www.ngdata.com/what-is-data-management.

Oracle (nd.) *What is data management?* https://www.oracle.com/database/what-is-data-management/

Linde, P. Wessels BA, Sveinsdottir, T, Noorman, M. (2014). How can libraries and other academic institutions engage in making data open? In Polydoratou P. and Dobreva, M. (Eds.), Let's put data to use: Digital scholarship for the next generation doi:10.3233/978-1-61499-409-1-3

Martin EG1, Law J, Ran W, Helbig N, Birkhead GS.(2017). Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on 3 Open Data Platforms. *Journal of Public Health Management and Practice*, 23(4):e5-e13. doi: 10.1097/PHH.0000000000000388.

Nelson, Bryn (2009) Data sharing: Empty archives. *Nature* 461, 160-163 https://www.nature.com/news/2009/090909/pdf/461160a.pdf

NEJM Catalyst (2018). *Healthcare big data and the promise of value-based care.* https://catalyst.nejm.org/big-data-healthcare/

Open Data Charter. *Principles* https://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf

Tenopir, C. Sandusky RJ. Allard, S., Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research,* 36(2), 84-90.

Verhulst, S., Noveck, BS. Caplan, R., Brown, K. and Paz, C. (2014).*The open data era in health and social care: a blueprint for the National Health Service* http://www.thegovlab.org/static/files/publications/nhs-full-report.pdf

Yoon, A. and Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College & Research Libraries*, 78(7), 920-933. https://doi.org/10.5860/crl.78.7.920