Data sharing practices of health researchers in Africa: a scoping review protocol

Oluchukwu Loveth Obiora¹,² • Benita Olivier¹,² • Dorothy Agnes Shead¹,²,³ • Aletha Withers⁴

¹Department of Physiotherapy, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, ²The Wits-JBI Centre for Evidenced-based Practice: A JBI Affiliated Group, Johannesburg, South Africa, ³Department of Anatomy, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, and ⁴Department of Paediatric Surgery, Chris Hani Baragwanath Academic Hospital, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

ABSTRACT

Objective: The aim of the review is to map the existing evidence regarding the data-sharing practices of health researchers in African countries. This review will also identify perceptions; barriers; facilitators; ethical-, legal-, and author-reported recommendations; institutional- and funding-related aspects that are being considered by African health researchers on data sharing in Africa and, as a result, identify areas for development and improvement in health care on the continent.

Introduction: The sharing of health-related data has been widely discussed in the literature. However, sharing health-related data has yet to become a common practice among health researchers in Africa, which bears a large burden of the global health diseases. The sharing of health research data could lead to greater development and improvement in health care in Africa.

Inclusion criteria: This review will incorporate studies that report on data sharing among health researchers in Africa. All primary, secondary, and gray literature that report on the practice of data sharing among health researchers in Africa will be included. Studies on data sharing on topics other than health-related data will be excluded. No language restrictions will be applied.

Methods: The JBI scoping review methodological framework will be adopted. An initial search of databases such as MEDLINE (PubMed), Scopus, LILAC, and Web of Science will be conducted. All search results will be screened and relevant data extracted by two independent reviewers. Data extracted will be exported into JBI SUMARI. The findings will be presented in the final scoping review report and illustrated in a PRISMA flow diagram.

Keywords: Africa; data sharing; health; information dissemination; open science


Introduction

Globally, there is growing advocacy for efficient sharing of individual-level data among health researchers, and campaigns to encourage and facilitate access to health research data.¹-⁴ Many research financiers, such as the National Institute of Health (NIH), the Wellcome Trust, the Bill and Melinda Gates Foundation, as well as several high-impact journals, require researchers to share their raw data as a criterion of funding to optimize their value and usefulness.¹,²,⁵,⁶ According to the Research Council in the United Kingdom, “publicly funded research data are a public good, produced in the public interest, which should be made available with as few restrictions as possible in a timely and responsible manner.”⁷(p.3)

Data sharing supports open science and involves making data readily accessible, comprehensible, reproducible, replicable, and verifiable.⁷,⁸ Data sharing refers to making health research data accessible to other investigators. This involves storing health research data in trusted repositories with regulated access, and making it accessible to other investigators on request.⁸ It promotes public trust, honesty, and completeness in science, and is rapidly turning into a new gauge for scientists’ accountability and
sincerity. Various types of data can be shared and these include the raw data and the cleaned data set, metadata, summary-level data such as summary-level results posted on registries, lay summaries, and publications. The benefits of sharing health research data cannot be over-emphasized. It can be very beneficial in addressing global health emergencies, enhances the opportunities for scientific collaboration and alliances, promotes the advancement of knowledge application to health products and policies, and enriches research and scientific skills. Sharing of health research data can also be instrumental in improving the timely discovery of health threats and disabilities, improving investigations into options, and tracking real-time responses. Furthermore, it can inform interventions and policy decisions, improve assessment capacity and performance measures, and increase the volume for public engagement, transparency, and responsibility. The merging of data from minor studies into huge datasets generates considerable statistical capability, presents feasible opportunities for comparative effectiveness research, and also shows respect and recognition of the research participants’ altruism.

According to Pisani et al., “if health researchers made their data available to colleagues, there would be less duplication of research and fuller use of study results. Data could be combined across time and countries to answer new questions, improving health policy. Data sharing would save time, effort, and money—it would probably also save lives.”

Unfortunately, despite these compelling benefits of sharing health research data, when compared with non-health research data, many research studies have reported low frequencies of raw data sharing in health research. There is no doubt that there are legitimate reasons why data is withheld, and these reasons can include issues of privacy and legal constraints. A 2014 systematic review identified significant impediments to the global sharing of health data, including technical, motivational, monetary, political, legal, and ethical impediments. There have been concerns about the capacity of the primary researchers to ensure that re-identification of data will not occur, especially when reverse engineering and the merging of related datasets may expand the possibility of recognizing the participants. Also, gathering data for research and implementing the procedures needed for the effective management of data demand considerable effort, competence, and resources. Shortage of resources required for efficient data sharing has been recognized as an impediment to research data dissemination in developed countries and as a serious impediment in low-resource settings. Much of the health data are derived directly from patients, clients, and populations, such as from clinical records, and demographic and survey data. The practice of data sharing has been widely discussed in literature. However, it is filled with contextual problems that make researchers, specifically those in developing countries, reluctant or cautious to share data. It is yet to become a common practice among health researchers in Africa, which bears a large burden of the global health diseases and has a rapidly growing population. Traditionally, it was not a usual practice for a published study to include a link to the raw data used for the study. This limitation, assisted by technological, social, and cultural factors among others, has slowed down the advancement of scientific knowledge. However, in this era of growth in information and great technological advances, several data repositories permit researchers to exchange their data utilizing a citable digital object identifier (DOI). There have also been some new policies and platforms to advance openness in research. The growing size of health research being carried out in Africa has the capacity to produce datasets of considerable value to researchers who are trying to confront the high prevalence rate of disease in Africa. Lessons learned from the data sharing in genomic research indicate that challenges or barriers to individual-level data sharing in developing countries might be unrelated in importance and differ significantly in moral principles from those emanating in the developed countries. Health decision-making is becoming progressively sophisticated and the utilization of generated data is vitally important both at individual, local, national, regional, and international levels. Therefore, there is a pressing need to explore the data-sharing practices of health researchers in Africa.

In Africa, some factors, which may be systemic, technological and/or cultural, define how researchers approach the challenges, barriers, and unique circumstances that surround the collection and sharing of health data in both urban and rural communities. These factors are rooted in political and historical structural issues, which the scientific community alone may not easily solve. Therefore, despite the
global support for open data, the feasibility of sharing research data for African scientists is still complex.\textsuperscript{19} It is true that health researchers are encouraged to make their data free, accessible, interoperable, and re-usable,\textsuperscript{22} however, the manner in which they achieve this relies heavily on their independent approaches and institutional assistance.\textsuperscript{20} A scoping review approach is the most relevant choice for a review topic that has such a wide-ranging scope,\textsuperscript{23,24} and this review seeks to map the existing evidence regarding the data-sharing practices of health researchers in African countries. Importantly, it will provide a summary of the evidence, identify gaps to guide future research aimed at improving the sharing of health research data in Africa, and identify areas for development and improvement in health care on the continent. To the authors’ knowledge, there are no systematic studies assessing the data-sharing practices among health researchers in Africa. A preliminary search of JBI Evidence Synthesis, the Cochrane Database of Systematic Reviews, PubMed, Figshare, Open Science Framework, and PROSPERO was completed in September 2020 and found no relevant complete or in-progress reviews on this topic.

**Review questions**

- What are the data-sharing practices (including geographical locations, types of data, data-sharing platforms used) of health researchers in Africa?
- What are the perceptions, barriers, facilitators, and author-reported recommendations of African health researchers on data sharing in Africa?
- What ethical, legal, institutional, and funding-related aspects are being considered by health researchers involved in the sharing of health research data in Africa?

**Inclusion criteria**

**Participants**

This review will incorporate studies that report health research data sharing among health researchers in Africa, irrespective of their specialty or years of experience. The health researchers may include individuals or groups who are conducting health research through an academic or tertiary institution or a research entity. They may include physiotherapists, nurses, doctors, and other researchers, irrespective of their age, gender, and other sociodemographic factors, who generate health-related data that can be used for research. Clinicians will be included if they share research data, which means they will be referred to as researchers in this context.

**Concept**

The concept of interest is the sharing of health-related data in Africa. In this scoping review, health research data refers to all health data generated for research by individual health researchers, health and medical facilities, institutes, or organizations. The various types of data that can be shared include raw data and the cleaned data set, metadata, summary-level data, such as summary-level results posted on registries, lay summaries, and publications.\textsuperscript{11}

This scoping review will examine how health research data is shared with regard to geographical locations; types of data; data-sharing platforms used; ethical, legal, institutional, and funding-related aspects that influence the sharing of health data; individual-level data sharing; clinical- or patient-oriented data; and perceptions, barriers, and facilitators to data-sharing practices among health researchers. All primary, review, and gray literature that report on the data-sharing practices among health researchers in Africa will be included.

**Context**

The focus of this scoping review will be on the sharing of health research data by health researchers residing in Africa, whose focus is the collection of data from study participants who live in Africa. This is because in Africa, some context-related factors (eg. systemic, technological, and cultural factors) define how researchers approach the challenges, barriers, and unique circumstances that surround the collection and sharing of health data in both urban and rural communities.\textsuperscript{10,19} No restrictions will be implemented in terms of language on the articles to be included in the study. Furthermore, data-sharing is partly technology-driven\textsuperscript{10,16} and there has been a proliferation of research studies in this area in the last 10 years (see Figure 1). This scoping review will be limited to articles published from January 1, 2011 to June 30, 2021. This timeframe will reflect current developments in the data management and sharing space.

**Types of sources**

This scoping review will consider all types of quantitative, qualitative, mixed-methods studies and reviews that report on the sharing of health research data among health researchers in Africa. If a study...
Reports on both health and non-health-related data, the reviewers will determine if the two can be separated. Unpublished articles, opinion pieces, policy documents, and dissertations that include information on health research data sharing among health researchers in Africa will also be evaluated for inclusion in this review.

Studies on data-sharing practices from outside Africa, as well as studies on data sharing on topics other than health-related data, will be excluded.

**Methods**

The proposed scoping review will use the JBI scoping review methodological framework.25

**Search strategy**

A three-step search strategy will be used. A preliminary search of MEDLINE (PubMed), Scopus, LILAC, and Web of Science will be conducted. An examination of the texts found in the titles and abstracts of extracted papers and of the index terms used to describe the articles will be done. Key terms will be decided through discussion between the authors.26

A second search utilizing the keywords and index terms identified from the preliminary search will be conducted across all applicable databases. The additional keywords, sources, and search terms discovered during the search will be included in the search strategy. The search will involve the use of keywords or text words such as: “data sharing,” “data access,” “open science,” “data management,” “open access,” “data ethics,” “datasets,” “data management activities,” “repository,” “health,” “medical,” “African countries,” “Africa,” and “human.” Also index terms like “information dissemination” and “data systems” will be used for the search. The search strategies will be designed categorically for each database using the relevant index and free-text terms. A complete search strategy for MEDLINE (PubMed) is presented in Appendix I.

In the final step, the reference lists of eligible articles will be searched (by OLO and DAS) for possible additional references. The titles and abstracts of all articles found in the search will be sorted and complete text versions of eligible articles will be obtained. Where necessary, authors of articles will be contacted for further information.

**Source selection**

The major databases for health care subjects that will be searched are MEDLINE (PubMed), Web of Science, LILAC, and Scopus. A manual search of the
references of included articles will also be done to discover potentially relevant articles based on the eligibility criteria. Relevant conference proceedings found in the previously mentioned databases will also be considered for inclusion in the review. The systematic search for articles will be carried out and the searches will be repeated prior to writing up the study findings to extract more current studies for inclusion. Any discrepancies related to source selection will be resolved by consensus.

Study selection
Following the search, all results from the databases and hand searches will be exported into the JBI System for the Unified Management, Assessment and Review of Information (JBI SUMARI; JBI, Adelaide, Australia) for screening. Duplicates will be removed before each entry is screened for eligibility. Study selection will be done in two stages. First, titles and abstracts will be screened against the inclusion criteria by two independent reviewers (OLO and DAS). Then, all potentially relevant full-text articles will be retrieved and screened for inclusion in the final review. The reasons for exclusion of full-text studies that do not meet the inclusion criteria will be reported in the review, and any disagreements will be resolved by consensus.

Data extraction
Two independent reviewers (OLO and DS) will retrieve data using the data extraction form created for this review (Appendix II), and the other authors (BO and AW) will check the relevance of the extracted data. Any inconsistencies will be resolved through discussion until an agreement is reached. The data extracted for this review will encompass details such as the year of publication, country of origin, aim/purpose of the study, types of data, study population and sample size, perceptions, barriers and facilitators, ethical considerations, legal, institutional and funding-related aspects, as well as the author-reported recommendations to mitigate the challenges of data sharing in the articles. The tabular summary will be detailed in a narrative summary, conveying the objectives of the scoping review.

Data presentation
Following PRISMA-ScR guidance, the results of the search will be reported in full in the final scoping review report and presented in a PRISMA flow diagram. The results of the scoping review will be presented in tables and synthesized into relevant charts and graphs, to facilitate the readers’ ability to understand and utilize the findings. The data-sharing practices, country of origin, types of data, aim/purpose of the study, study population and sample size, data-sharing platforms, perceptions, barriers and facilitators, ethical considerations, legal, institutional and funding-related aspects, as well as the author-reported recommendations to mitigate the challenges of data sharing in the articles will be reported. The tabular summary will be detailed in a narrative summary, conveying the objectives of the scoping review.

Acknowledgments
The current scoping review protocol will contribute toward a postdoctoral project on data-sharing practices for OLO.

References


Appendix I: Search strategy

MEDLINE (PubMed)
Search conducted on July 1, 2021

(A similar strategy will be used in other databases. However, due to the different index terms featured in different databases, the currently presented index terms as MeSH terms, may not be available in all databases)


Filters: from 2011/1/1 - 2021/6/30

A total of 2138 studies were found.
Appendix II: Data extraction instrument

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year of publication</th>
<th>Country of origin (where the study was published or conducted)</th>
<th>Aims/purpose</th>
<th>Study design</th>
<th>Study population and sample size</th>
<th>Data sharing practices (types of data, data sharing platform, perceptions)</th>
<th>Facilitators for data sharing</th>
<th>Barriers to data sharing</th>
<th>Ethical, legal, institutional, and funding-related aspects</th>
<th>Author-reported recommendations to improve health data sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>