frontiers | Research Topics

Open science in Africa

Edited by
Katie Wilson, Rania Mohamed Hassan Baleela, Heila Pienaar,
Martie Van Deventer and Dale Peters

Published in
Frontiers in Research Metrics and Analytics

ISSN 1664-8714
ISBN 978-2-8325-3250-8
DOI 10.3389/978-2-8325-3250-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Open science in Africa

Topic editors

Katie Wilson — Curtin University, Australia
Rania Mohamed Hassan Baleela — University of Khartoum, Sudan
Heila Pienaar — University of Pretoria, South Africa
Martie Van Deventer — University of Pretoria, South Africa
Dale Peters — University of Cape Town, South Africa

Citation

Wilson, K., Baleela, R. M. H., Pienaar, H., Van Deventer, M., Peters, D., eds. (2023). *Open science in Africa*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3250-8

Table of contents

frontiers | Frontiers in Research Metrics and Analytics

TYPE Editorial
PUBLISHED 14 June 2023
DOI 10.3389/frma.2023.1233867

OPEN ACCESS

EDITED AND REVIEWED BY
Dietmar Wolfram,
University of Wisconsin–Milwaukee,
United States

*CORRESPONDENCE
Heila Pienaar
heila.pienaar@up.ac.za

RECEIVED 02 June 2023
ACCEPTED 06 June 2023
PUBLISHED 14 June 2023

CITATION
Pienaar H (2023) Editorial: Open science in Africa. *Front. Res. Metr. Anal.* 8:1233867. doi: 10.3389/frma.2023.1233867

Editorial: Open science in Africa

Heila Pienaar*

Department of Information Science, University of Pretoria, Pretoria, South Africa

KEYWORDS

editorial, open science, Africa, open data, open access, policy, barriers and facilitative factors

Editorial on the Research Topic
Open science in Africa

The ongoing transition toward Open Science (OS) is increasing transparency and collaboration in the research enterprise. This Research Topic aims to investigate the transition to OS in Africa, including the concerns and advantages of OS for researchers and stakeholders. It also explores the role of new technologies and infrastructure in implementing OA and bridging the knowledge divide between countries. In this editorial, we provide an overview of eight articles that shed light on various aspects of open science, data sharing, and the challenges and opportunities they present in the African context. These articles highlight the importance of policymakers, institutions, and researchers working together to foster a culture of open science and to address the existing barriers to data accessibility on the African continent.

The article by Okafor et al. focuses on the adoption of open science (OS) practices in Africa, considering the limitations and prospects for its institutionalization. The authors emphasize the significance of science access for the advancement of scientific research and the development of the next generation of scientists in Africa. They highlight the global resurgence of discussions around open science due to the COVID-19 pandemic, particularly in resource-poor settings like Africa where OS practices are currently limited. Overall, the review article serves as an advocacy strategy and informative guide for policymakers and stakeholders involved in promoting and integrating open science practices in Africa. It highlights the importance of overcoming barriers and fostering a supportive environment for open science to thrive on the continent (Okafor et al.).

The next article, "*Rethinking the a in FAIR data: issues of data access and accessibility in research*" by Shanahan and Bezuidenhout, raises concerns about the assumptions of accessibility in FAIR (Findable, Accessible, Interoperable, and Reusable) data principles. The authors emphasize that access to FAIR data resources can be influenced by geopolitical factors, exacerbating existing access inequities. They stress the need for increased awareness and consideration of these issues in FAIR implementation (Shanahan and Bezuidenhout).

The article, "*Open science in Africa: what policymakers should consider*" by Chiware and Skelly, underlines the importance of African governments and institutions embracing open science principles and building research infrastructures that align with the global open science movement. The authors highlight the significance of OS policy frameworks and provide insights for policymakers, aiming to guide similar initiatives in Africa (Skelly and Chiware).

"*African researchers do not think differently about open data*" by Skelly and Chiware, explores African researchers' attitudes toward open data and demonstrates that their perspectives are not significantly different from their international counterparts. This finding emphasizes the need for policymakers and institutions to understand and address researchers' concerns and expectations regarding data sharing and the open data ecosystem (Skelly and Chiware).

In "*Open access and its potential impact on public health—a South African perspective*", Strydom et al. examine the impact of open access on public health in South Africa. They highlight the benefits of open science and discuss financial implications and potential solutions for reducing publication costs for researchers and institutions. The authors also address privacy concerns and the role of data protection legislation in medical research and data reuse (Strydom et al.).

Hey's article, "*Open science and big data in South Africa*", focuses on the challenges and opportunities presented by "Big Scientific Data" in South Africa, particularly in the context of the Square Kilometer Array project and the Multi-Purpose Reactor. The author highlights the importance of open science policies and the FAIR principles in managing and making such data accessible, proposing the use of semantic markup and emphasizing the role of interdisciplinary teams in research data management (Hey).

Chigwada's "*Feasibility of a national open data policy in Zimbabwe*" explores the potential for implementing a national open data policy in Zimbabwe. The study assesses the readiness of the country in terms of open data activities, highlighting the need for advocacy, awareness creation, and collaboration among stakeholders to craft and enact a national open data policy. The author emphasizes the value of government and research data for driving research and innovation (Chigwada).

"*Building awareness and capacity of bioinformatics and open science skills in Kenya: a sensitize, train, hack, and collaborate model*" by Karega et al., presents a framework for promoting bioinformatics and open science skills in Kenya. The authors showcase the Sensitize-Train-Hack-Collaborate/Community model, which combines awareness-building, training, collaborative projects, and community engagement to empower researchers with the necessary skills and tools in open science and bioinformatics (Karega et al.).

These articles collectively underscore the importance of open science, data accessibility, and policy development in Africa. They highlight the need for increased awareness, capacity building, and interdisciplinary collaborations to overcome challenges and leverage the potential of open science.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Research Metrics and Analytics

REVIEW
published: 15 April 2022
doi: 10.3389/frma.2022.855198

Institutionalizing Open Science in Africa: Limitations and Prospects

Izuchukwu Azuka Okafor[1,2,3], Smart Ikechukwu Mbagwu[1,4], Terkuma Chia[5], Zuwati Hasim[6], Echezona Ejike Udokanma[7] and Karthik Chandran[8]*

[1] Department of Anatomy, Faculty of Basic Medical Sciences, College of Health Sciences, Nnamdi Azikiwe University, Nnewi, Nigeria, [2] Department of Obstetrics and Gynaecology, College of Medicine, University of Ibadan, Ibadan, Nigeria, [3] Pan African University of Life and Earth Science Institute (Including Health and Agriculture), University of Ibadan, Ibadan, Nigeria, [4] Beth Israel Deaconess Medical Centre, Harvard Medical School, Boston, MA, United States, [5] Department of Anatomy, Faculty of Basic Medical Science, College of Health Sciences, Nile University, Abuja, Nigeria, [6] Department of Language and Literacy Education, Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia, [7] Birmingham Business School, University of Birmingham, Birmingham, United Kingdom, [8] Department of Automation, Shanghai Jiao Tong University, Shanghai, China

The advancement of scientific research and raising the next-generation scientists in Africa depend largely on science access. The COVID-19 pandemic has caused discussions around open science (OS) to reemerge globally, especially in resource-poor settings like Africa, where the practice of OS is low. The authors highlighted the elements, benefits, and existing initiatives of OS in Africa. More importantly, the article critically appraised the challenges, opportunities, and future considerations of OS in Africa. Addressing challenges of funding and leadership at different levels of educational, research, and government parastatals may be pivotal in charting a new course for OS in Africa. This review serves as an advocacy strategy and an informative guide to policymaking and institutionalization of OS in Africa.

Keywords: open science, Africa, advocacy, institutions, engagement

OPEN ACCESS

Edited by:
Heila Pienaar,
University of Pretoria, South Africa

Reviewed by:
Adriana Bankston,
University of California Office of the President, United States
Rania Mohamed Hassan Baleela,
University of Khartoum, Sudan

****Correspondence:***
Izuchukwu Azuka Okafor
okaforizuchukwu33@gmail.com;
iza.okafor@unizik.edu.ng

Specialty section:
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

Received: *14 January 2022*
Accepted: *02 March 2022*
Published: *15 April 2022*

Citation:
Okafor IA, Mbagwu SI, Chia T, Hasim Z, Udokanma EE and Chandran K (2022) Institutionalizing Open Science in Africa: Limitations and Prospects. Front. Res. Metr. Anal. 7:855198. doi: 10.3389/frma.2022.855198

INTRODUCTION

Open science (OS) is a movement focusing on making science more open, accessible, effective, democratic, and transparent to society, notwithstanding the level of education (1). Suffice to say that OS, as an inclusive science, potentially closes the science, technological and innovation divide between and within nations. According to the final report of the United Nations Educational, Scientific and Cultural Organization (UNESCO) on OS, twelve elements of OS exist, including open data, open infrastructure, open access (OA), open hardware, open laboratories, open-source, open innovation, open notebook, open evaluation, open educational resources (OERs), crowd funding, and citizen science (UNESCO, 2021). UNESCO has described these elements of OS in their recent recommendations (UNESCO, 2021). The recommendations has posited that none of the elements of OS should be neglected in implementing OS strategies and all the components should work in synergy to produce a more effective and scalable OS system. These elements have been summarized in **Figure 1**.

Access to science holds the key to strengthening health systems, advancing scientific research, and efficiently training Africa's next-generation scientists. However, this access is limited mainly due to inadequate funding of science in Africa, with poor funding of 0.1–0.5% gross domestic product (GDP) for science and technology in many African countries as against UNESCO's recommendation of at least 1% GDP (Christie, 2019; Krishna, 2020). Since the COVID-19 pandemic, the need for OS has re-emerged in Africa and other

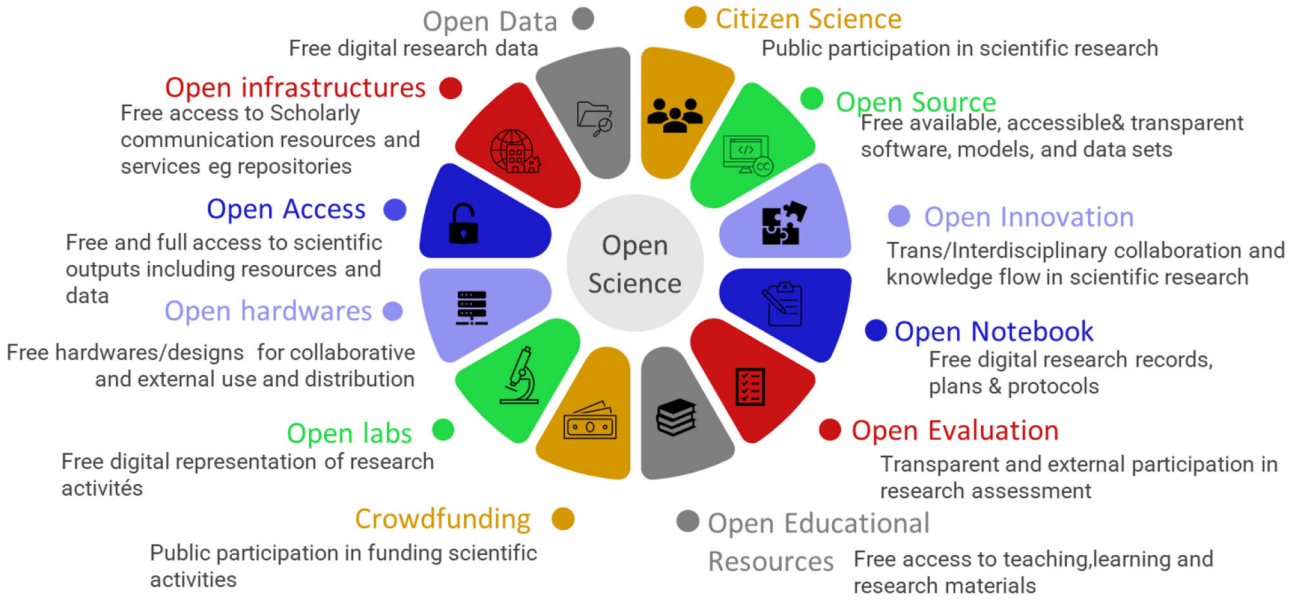


FIGURE 1 | The elements of open science as adapted from the UNESCO recommendation 2021.

resource-poor settings because of its critical role in pandemic preparedness and response. The inequality in OS practices and the associated consequences in Africa compared to the developed countries became more evident with the planning and implementation of COVID-19 pandemic responses (Havemann et al., 2020).

OS as a phenomenon is either misconstrued, neglected, or not yet institutionalized (Krishna, 2020). The late sixteenth and early seventeenth century saw the emergence of OS (David, 2004). Modern OS was started by the global adoption of the institutional scientific journal. Consequently, England established the Royal Society in 1660 while France established the French Academy of Sciences in 1666 (David, 2004). The 1990s OS movement began in the United States of America as a springboard to its global spread (Venith, 2015). Following the 2015 competitiveness council, the European Research Ministers developed European OS Agenda (Heise and Pearce, 2020). This has seen the commencement of diverse OS and OA projects and initiatives like OpenAIRE, RECODE, and OpenScienceLink (Heise and Pearce, 2020). Africa has witnessed some OS projects, including Africa Open Science Platform (AOSP), DataFirst, and OA for Africa. Further, Library Support for Embedded NREN Services and E-infrastructure (LIBSENSE), which supports OS and research in Africa, was launched in 2017 with diverse regional workshops conducted (**Table 1**) (Kuchma, 2022). Despite these efforts, only three African nations (Gabon, Mauritius, and Namibia) gave written feedback to the first draft of the Recommendation on OS out of the 47 UNESCO Africa member states (UNESCO, 2021). These questions the political will of the African member states to fully institutionalize OS in their countries. Kenya, Ghana, and Morocco first embraced the OS movement. While French-speaking Sub-Saharan African nations have shown hesitation in adopting this movement, Burkina Faso and Sierra Leone led OS adoption in this region. This movement notwithstanding, it took the current COVID-19 to strongly re-establish the need for OS (UNESCO, 2021). Following the globally adopted public health measures in containing COVID-19, there is a need for real-time data on the rate of infection, mortality, and emerging variants in nations. This is pertinent in constantly evaluating how nations are faring in disease containment, treatment, and lessons drawn for worse hit nations. This strategy re-instated the need for OS globally, especially in Africa where it has been underutilized. In this critical time, lessons may be drawn from most African nations with minimal infection and mortality rates; however, the dearth of OS may have been a limitation.

OS is undoubtedly beneficial but embodies diverse challenges in Africa, may involve some level of restriction in research flexibility, time cost, and poor or non-existent incentive structure (Allen and Mehler, 2019). Also, the language barrier plagues OS in Africa as most of the available OS platforms are English, which poses a challenge to science communication (Mwelwa et al., 2020). Some of the solutions to the challenges of OS in Africa have been discussed at different levels by experts and stakeholders, which may include proper funding, stable internet, science infrastructure, leadership, policy development, proper monitoring, and evaluation. In this exploratory review, we analyse the benefits, challenges, and opportunities of OS in Africa. Furthermore, we conduct chronological profiling of the OS platforms and initiatives in Africa. This study serves as an evidence-based informative guide to facilitate advocacy strategies, policymaking, and institutionalization of OS in Africa.

BENEFITS OF OPEN ACCESS TO SCIENCE

The aim of OS is to let anyone access the results of a scientific research or publicly funded research data for knowledge, reuse

TABLE 1 | Some open science initiatives/platforms in Africa.

Projects/initiatives	Hosts/country	Type	Focus	Period/year of establishment	Remarks
Africanfossils.org (https://africanfossils.org/)	Partnership by Autodesk, Turkana Basin Institute, and the National Museums of Kenya, Stony Brook University and the National Geographic Society.	Virtual lab for fossil collections	• To host a collection of 3D models of significant fossils and artifacts for researchers students and interested audience	2014	Promotes the increase in knowledge to the public on prehistoric times
African Virtual University Project (AVU) (https://avu.org/avuweb/)	Pan African Intergovernmental Organization with headquarters in Kenya	eLearning Network	• To provide education in the area of Science, Renewable Energy, Food Security and ICT, etc.	1997	Provides educational training to 18 participating African countries
African Journals Online (https://www.ajol.info/index.php/ajol)	South Africa	Digital Repository For African research	• To increase global and continental online access, awareness, quality and use of African-published, peer-reviewed research.	1998	Currently hosts 535 Journals with 274 Open Access Journals
The Scholarly Communication in Africa programme (SCAP) (http://www.cilt.uct.ac.za/cilt/scap)	Centre for Educational Technology and the Research Office at the University of Cape Town. In close collaboration with the Southern African Regional Universities' Association (SARUA),	Training funded by Canadian International Development Research Centre (IDRC)	• To increase African universities' contribution to regional and global knowledge production.	2010–2014	Promoted the visibility of African researchers, creation of repositories and exploration of affordable business models for the open online publication of scholarly materials
Open Access for Africa (https://umb.libguides.com/OAA)	UNESCO and the Network of African Science Academies (NASAC), Royal Netherlands Academy of Arts and Sciences, Kenya National Academy of Sciences, African Academy of Sciences, and Kenyan Ministry of Education, Sciences and Technology.	Advocacy	• Provision of expert intervention for research and development in Africa.	2015 (29–30 January)	UNESCO encouraged the establishment of training centers for capacity building in the area of open Access philosophies and systems.
African Digital Research Repositories (https://www.internationalafricaninstitute.org/repositories)	International African Institute (IAI), London and AfricArXiv	Digital repository	• Improve the discoverability of African research and publications • Enhance the interoperability of existing and emerging African repositories • Identify ways through which digital scholarly search engines can enhance the discoverability of African research	2016	Promotes research-based knowledge from African repositories
Electronic Publishing (https://codesria.org/spip.php?rubrique257andlang=en)	Council for the Development of Social Science Research in Africa (CODESRIA), Dakar, Senegal.	Advocacy	• To discuss opportunities and challenges to the Open Science movement in the region.	2016 (March 30–April 1)	Dakar Declaration on Open Science in Africa to promote and support Open Science across Afsrica.

(Continued)

TABLE 1 | Continued

Projects/initiatives	Hosts/country	Type	Focus	Period/year of establishment	Remarks
The African Open Science Platform (AOSP) (https://council.science/current/news/the-national-research-foundation-of-south-africa-to-host-the-african-open-science-platform/)	National Research Council of South Africa supported by South Africa's Department of Science and Innovation (DSI), key institutions in Africa, and the International Science Council (ISC).	Advocacy	• To provide current landscape of data/science initiatives in Africa • To create a Pan-African open science community. • To promote the formation of a national open science fora.	2016 (Operational kick off in 2020)	Encourages increased commitment to Open Science
LIBSENSE (Library Support for Embedded NREN Services and e-infrastructure) (https://spaces.wacren.net/display/LIBSENSE/Home)	WACREN—West and Central African Research and Education Network in partnership with different organizations.	Repositories	• Advancing open Science in Africa through strengthening and expanding services at the institutional, national and regional level.	2017	Promotes the availability and adoption of indigenous open science services and infrastructures in Africa
AfricArXiv (https://info.africarxiv.org/)		Digital archive for African research,	• Provide open access to research information • Highlight, display and promote African journals and African research output and expertise • Provide collaboration among African scientists locally and globally. • Fill the gaps where institutional repository systems are missing	2018	Provides platform for preprints, accepted manuscripts (post-prints), and published articles of African scientists. Provide collaboration among African scientists.
The H3ABionet project (https://www.h3abionet.org/)	South Africa	Bioinformatics Network	• Education and training • Development of Bioinformatics tools and services • Scientific engagement and communications	2019	Provides Support for research in genomic sciences
African Academy of Sciences (AAS) Open Research (https://aasopenresearch.org/)	Headquarters is located in Kenya	Repository	• For publication and peer review of research articles majorly supported by AAS and The Alliance for Accelerating Excellence in Science in Africa (AESA)	2019	Provides scholarly impact while promoting reproducibility and transparency
DataFirst (https://www.datafirst.uct.ac.za/)	South Africa	Data Repository	• Provides a repository of data for South Africa • Provides training and research on data quality and usage	2020	Promotes access to open research data infrastructure especially in South Africa

or innovation purposes. Looking at the various benefits of open data and OS, in January 2021, the Organization for Economic Co-operation and Development (OECD) council, while promoting OS, adopted a revised "Recommendation on Access to Research Data from Public Funding." The revised recommendation aimed to enhance access to scientific data in order to address global challenges, and at the same time, to advocate for protection of specific data. This is clearly outlined in the European Commission Recommendation 2018/790 of April 25 2018 (European Commission, Directorate-General for Communications Networks, Content and Technology, 2018):

> *research data that results from publicly funded research becomes and stays findable, accessible, interoperable and re-usable ('FAIR principles') within a secure and trusted environment, through digital infrastructures (including those federated within the European Open Science Cloud (EOSC), where relevant), unless this is not possible or is incompatible with the further exploitation of the research results ('as open as possible, as closed as necessary'). This could be for reasons, in particular, of privacy, trade secrets, national security, legitimate commercial interests and to intellectual property rights of third parties. Any data, know-how and/or information whatever its form or nature which is held by private parties in a joint public/private partnership prior to the research action should not be affected by these policies or national action plans. (L134/15).*

A recent study highlighted the need for shared information to all, citing an example of what the world is currently facing with the COVID-19 pandemic (Paic, 2021). The authors hinted that scientists and researchers around the globe came together and shared their knowledge on the full genome of the coronavirus that could provide a basis for understanding the symptoms, finding ways of treatment, and producing vaccines that may protect people from the virus. The establishment of OA journals to share studies related to the virus was indeed constructive and a way forward in getting all researchers to solve the global issue. Paic emphasized that "the sharing of research data can help accelerate the fight against pandemics and other global emergencies" (Paic, 2021). Evidences have emerged on how data sharing and management both locally and globally has helped to fight the spread of COVID-19 (Budd et al., 2020; Gao et al., 2020). In agreement with this fact, we opine that the relative success being recorded against the COVID-19 pandemic is due to the complete adoption of the different aspects of OS (**Figure 1**) and the global synergy in doing this. Indeed, there are various benefits to OA or OS. These benefits have been summarized into themes (**Figure 2**) and further elaborated in the sections below.

Networking and Collaboration

The first benefit of OS is its ability to promote networking and collaboration among researchers or between researchers and research funders or with other stakeholders. Bezuidenhout et al. (2020) surveyed data sharing by the low/middle-income country scientists belonging to the New Partnership for Africa's Development (NEPAD)-Southern African Network for the Biosciences (NEPAD-SANBio). These scientists believed that data sharing (known as open data) could give them opportunities to build networks and collaboration. For example, in open data, researchers could access and share questionnaires, data, and metadata that could be re-used. Harding (2016), through his work on global health innovation technology models, identified the critical role of an open-source platform that is to enable "re-usable clinical intelligence that can be shared and redistributed in the context of clinical innovation before, during, and after care is delivered" (p. 4). The form of collaboration among the global health community is evident, through the development of mHealth (involving healthcare data exchange *via* mobile phone technology), for effective patient engagement. For this purpose,

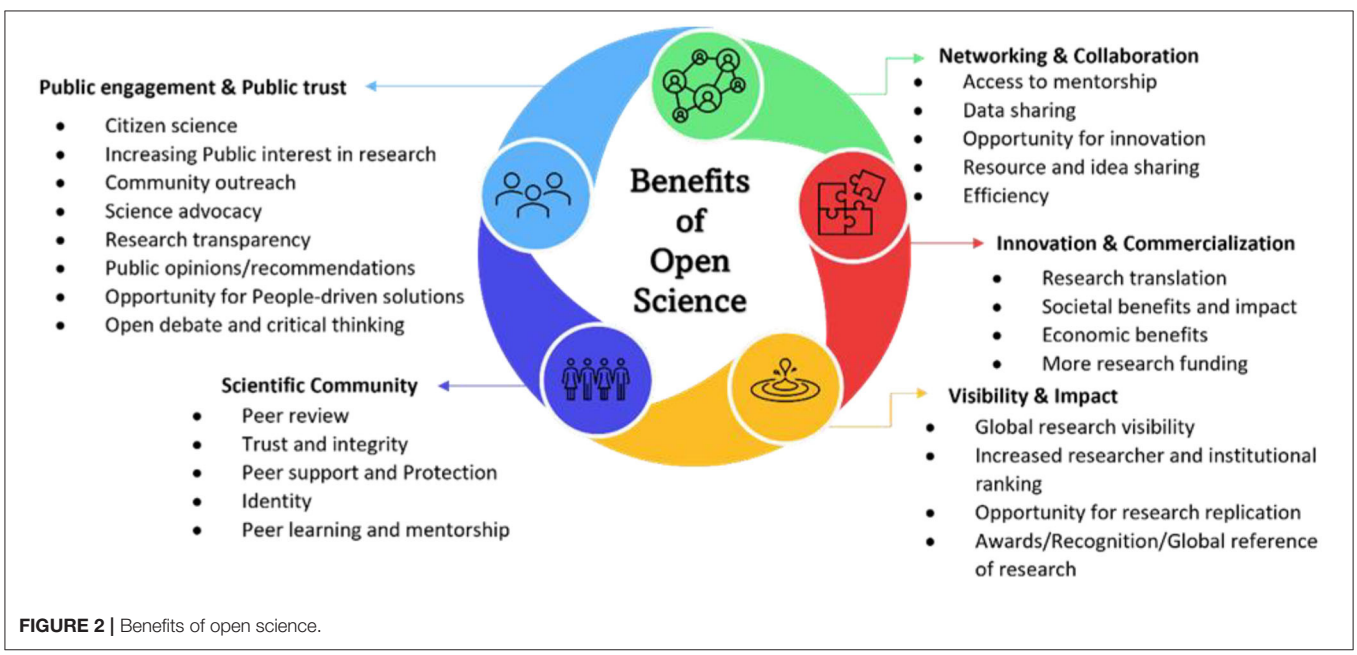


FIGURE 2 | Benefits of open science.

collaboration occurs in the form of peer-to-peer clinical support. Also, with the mHealth platform, the OS provides a virtual collaboration setting that engages research scientists in quickly sharing knowledge for clinical innovation with their global network. In OA, Besancon and coworkers (Besançon et al., 2021) assert that it allows for peer-reviewing through the open-review principle, where peer-review reports are made open to the public. This is seen as a form of collaboration to help in improving and maintaining the quality of research reports. Another OS element that promotes collaboration and networking is the "open innovation" that promotes interdisciplinary research. Researchers from various fields and even the stakeholders could come together in providing input, designing, producing, and delivering the expected outcomes from the research objectives.

Public Engagement and Public Trust

OS permits the sharing and transferring of knowledge and scientific data into meaningful information to the general public by giving access to the software and the datasets through OA and open-source platforms. Hudson et al. (2020) opine that "public funders anticipate that research will lead to public benefit" (p. 377). Hence, it is appropriate to engage the public by giving access to the research data and research findings for the purpose of knowledge sharing and dissemination. OS promotes better science-society relationship where engagement is not just referring to the general public but to the various interested parties. This is part of the citizen science initiative where public participation in scientific research is encouraged. Also, the return of meaningful results, from the research carried out, to the public or the funders is considered as part of the social benefits of OS. The transparency in data sharing will also increase public trust in the research conducted. This will be achieved through open evaluation where external involvement for research assessment is made possible. Data and research information that are shared through open system have the chance to be peer-reviewed, annotated, recommended, refuted, discussed, read and taught (Priem et al., 2012). Consequently, this will expand the value of research (Fleming et al., 2021). It will also serve as a means of enhancing understanding, data checking and data confirmation for accuracy (Exley et al., 2015; Shea, 2015) which subsequently establishes the reliability and credibility of the research results that could best be achieved through direct information disclosure in research publications that are accessible to the public, the stakeholders, and the funders (Lakomý et al., 2019). Science communication and public engagement through open-science initiative is important for it allows open debate, initiates critical thinking and allows correcting misinformation from the media (Eagleman, 2013). The public can only become aware of the processes and the complexities of conducting research through the open sharing of knowledge (Lakomý et al., 2019). Science engagements and advocacies must be community-driven and possibly incentivized to increase public trust and participation in science-based decision making processes. In valuing scientific pursuits, public engagements in a form of crowd funding (i.e., contributions from the public) for scientific activities are sought after. Notably, OS is a way of communicating science to the public aiming at developing understanding, changing attitudes, and developing interest toward science, at the same time promoting literacy in science.

Visibility and Impact

OS in the forms of OA and open data repositories helps in contributing to the visibility of research and in turn leads to greater impact of the research to the scientific community or the society at large. The visibility of research works does not only benefit individual or group of researchers but also their affiliations. The visibility of research works *via* OA helps to showcase and promote the scientists, their affiliated institutions as well as the fund providers. Furthermore, it serves for promotional purposes and also helps institutions to fulfill the requirement of global ranking assessment by publication citations (Momeni et al., 2021). There is an increased probability for others to read and cite a scientific publication with OA, thus enhancing knowledge sharing. One of the impacts from research visibility is that the research could be replicated in other or related contexts. According to Adeyemo and Jamogha (2021), institutional repositories have the role toward enhancing institutional visibility and supporting "scholarly communication among the academic community" (p. 3) especially where the shift from physical print to digital sources has made their work globally accessible. Visibility of research also impacts on the quality of the research being published as scholars are more likely to ensure their work meets certain ethical standards and is worthy of publication and global reference. Advocacy for OA must not jeopardize rigor and quality of publications. Hence, stakeholders must ensure to maintain and keep improving the on current policies that keep the quality of OA publications in check. Institutional repository can be an indicator of institutional quality by displaying works that are of public value (Crow, 2002). Momeni et al. (2021) in their study on the impact of changing publication model from closed to OA, found that the impact factor increases after flipping the journal publications from closed-access to open-access. Their findings reiterated the results of earlier studies (Busch, 2014; Bautista-Puig et al., 2020; Adeyemo and Jamogha, 2021). However, the increase on the impact factor varies across scientific fields (Momeni et al., 2021). In addition, the visibility of research also allows for OERs that could be used for teaching or training purposes which will definitely give a significant impact on teaching, learning, and research development through shared knowledge and practices (Das, 2011; Stagg, 2014; Guo et al., 2015; Bliss and Blessinger, 2016).

Scientific Community

OS is instrumental in the development of the scientific community. As part of the OS principles, research should be accessible, transparent, re-usable and reproducible. Under this philosophy, a growing scientific community from various disciplines is expected. Data and research information that are available *via* open-source platforms could facilitate scientific collaboration and promote discussions among experts in their respective fields. For example, Taylor and coworkers (Taylor et al., 2017) looked into how OS could benefit Modeling and Simulation (MS) researchers. They asserted that there are various

forms of OS that serve as artifacts for MS research (i.e., published research articles, model or simulation program and its execution environment, software, experimentation schema, data, etc.). These artifacts "would be available openly and in a discoverable form" (p. 544), allowing for reproducibility by adopting good open data. To add, OS gives room for verification of research data by other scientists. Beck et al. (2020) who reviewed the benefits of OS on bioassessment, mentioned that "open data products can increase efficiency of the individual researcher and a collective research team by encouraging collaborators to adopt an OS workflow" (p. 6). Additionally, OS and OA initiatives will enhance crowdsourcing for reusability of research data, techniques or methods, by relevant or interested research community of intra- or inter-disciplines. Hetu et al. (2019) based on their study on the impact of open genomic projects, indicated that large-scale databases should be widely accessible to allow advancement of genomic medicine and capacity building in research and development particularly for developing countries where genomic research skills is still growing. They believed that through this accessible databases and OS, researchers or scientists, across the globe could together orient selected projects. They found that international OS project (genomics research, in their case) can make impact on capacity building (of scientific community) through training of researchers, development of research infrastructures, and building of expertise. By having this scientific community through the OS initiative, it helps not only in the capacity building but also in accelerating research (Besançon et al., 2021; Ewers et al., 2021; Kadakia et al., 2021). Kadakia et al. (2021) mentioned that OS "promotes standard processes for sharing protocols and registering studies, reporting and disseminating results, sharing data, biospecimens and code" (p. 1) that allows such data "to be findable, accessible, interoperable, and re-usable to permit independent scrutiny, replication, and follow-on investigations" (p. 1). All these will help the scientific community to accelerate their research. For instance, the COVID-19 pandemic has created urgency for National Institutes of Health (NIH) to set up a research platform that is intended for researchers to share research tools, metadata, and their reports. Similarly, some academic publishers have come together to support preprints to expedite knowledge transfer (Puebla, 2020; Fraser et al., 2021; Hayashi, 2021), and open-access policies to encourage the sharing of information that enable researchers to learn and synthesize from the emerging evidence (Kadakia et al., 2021).

Innovation and Commercialization

OS initiatives assist in innovating and commercializing research protocols and outputs. Valuable data gathered in the accessible pools of open system, particularly for the scientists, investigators, consultants and researchers, are useful for reproducibility and reusability of research (McKiernan et al., 2016). The reproducibility and reusability are not limited to the data or metadata shared but also to the information related to the research process and procedures, methods and approaches, protocols, models, policies, systems and technologies, cases, etc. These qualities of reusability and reproducibility often times lead to innovation (Kedron et al., 2021). Although Capps (2021) raised a concern over OS becoming a contingency for irresponsible innovation and research misconduct, which necessitates the need for appropriate policies and sanctions to safeguard the future of OS in Africa. Howbeit, we still could not deny the benefits it has to offer with regards to innovation and commercialisation. Again, the recent example from the COVID-19 pandemic that has accelerated the OS practice saw to a cooperative and collaborative work that led to vaccine development. These vaccines have been commercialized with different brand names from Pfizer, Astra Zeneca, Sinovac, Moderna, and Sinopharm. However, "the open commons demand stronger normative principles to support innovative use of new scientific knowledge, but also requires obligations to use it ethically" (Capps, 2021) through open innovation (Wendzel et al., 2017), following this, Granados Moreno et al. (2019) added that OS will expedite innovation by means of partnership and collaborative process that could lead to optimal innovation with minimal economic burden through partnership agreements.

It is evident that all aspects of OS (**Figure 1**) provide several benefits not only to the researchers but also the public. Among the benefits discussed in this section are networking and collaboration opportunities, building public engagement and public trust, promoting visibility and impact through OA platforms, establishing scientific community, as well as enhancing innovation and commercialization. With these benefits, OS is worth being supported through relevant policies especially in growing economies like Africa.

CHALLENGES OF OPEN SCIENCE IN AFRICA

The primary objective of OS is to increase the value and reliability of scientific output, increasing efficacy and spurring discovery and innovation (Nosek et al., 2015; Heuritsch, 2020). Multiple mechanisms are employed to achieve this objective including deliberate institutional policies, infrastructure and relationships that promote OA publications, open data and scientific resources as well as removal of restrictive intellectual and other proprietary rights (Ali-Khan et al., 2018). The requirements to drive these mechanisms are enormous and laden with several challenges. The challenges hampering the adoption and development of OS are even more pronounced in resource-limited environments like Africa (Mwelwa et al., 2020; Mwangi et al., 2021). Directly or indirectly, many of these barriers may be associated with lack of adequate funding for education/research which is reported to fall below expectations in many African countries (Teferra and Altbachl, 2004).

Due to insufficient funding, several African researchers lack state-of-the-art facilities which are available to their counterparts elsewhere (Kokwaro and Kariuki, 2001; Yusuf et al., 2014). This directly impacts OS which relies heavily on technology and skills. African researchers are unable to undertake OS projects, having to work with limited resources, thereby reducing the propensity for quality and credibility which are the hallmarks of OS. The cost of disseminating research is another challenge as in most cases such cost have to be incurred personally by the researchers due

to lack of grants which could have covered research publications (Ahinon and Havemann, 2018). This has informed the decision by several publishers to give varying amounts of discounts on article processing charges or outright waivers to researchers from low and middle income countries. However, these seems not to be enough as the prices after discount are still unaffordable since these charges are in foreign currencies which have higher values than most African currencies (Ezema, 2011). Paucity of research sponsorship sets the stage for several other impediments to OS in Africa such that if addressed could revolutionize OS on the continent.

Relatedly, lack of infrastructure, availability of tools, and processes that aid OS infer low skills to champion OS among African researchers. Authors have noted the low internet penetration in Africa compared to the developed countries, and even when available, the speed is typically slower (Steiner et al., 2005; Les Cottrell, 2013). Cyber infrastructure for instance enables investigators to cope with the data volume, provide effective data interfaces and visualization and utilizes more powerful algorithms to extract more information from these data sets (Ramachandran et al., 2021). Consequently, availability of infrastructure and capacity building on research design, data entry and the use of cyber platforms is indispensable to the advancement of OS in Africa.

Lack of deliberate policies and legal frameworks promoting OS from governments, institutions and funders prevents the advancement of OS in Africa (Onie, 2020). In the European Union for example, there is a well-coordinated policy and programmes on OS (European Union, 2017). Policies must be balanced and focused on how to navigate some of the potential barriers of OS policies such as privacy, trade secrets, national security, legitimate commercial interests and intellectual property rights of third parties. The availability of policies together with the provision of funds has improved OS in Europe and has placed it as one of the leading regions for OS (Leonelli et al., 2018). Relevant policies with strong legal frameworks can be institutionalized in Africa and tailored to meet the localized realities of the challenges OS face in Africa.

The lack of OS awareness is an additional hindrance to OS among African researchers (Teferra and Altbachl, 2004; McKiernan et al., 2016; Mwelwa et al., 2020). Many are unfamiliar with OS and/or its ramifications; as such they have not bothered to develop the requisite skills that enable the practice of OS. Some African researchers who are well informed about OS are reluctant to practice it due to the lack of incentives for OS practices (Allen and Mehler, 2019). Typically, researchers employing traditional methods get results quicker and publish faster unlike OS methods which take longer due to the complexities of transparency. With the pressure for academics to "publish or perish", traditional researchers have greater possibilities for faster career progression because of the time demand of OS practices (Allen and Mehler, 2019). With low motivation and less recognition given to OS practitioners in Africa, researchers are further discouraged to fully practice OS. These researchers are usually keen to access quicker routes to publish their research to further their careers and in many cases encouraging unethical and illegal practices to get desired results. Such practices embolden corruption in the academia in a direct antagonism to the integrity and quality which OS seeks to entrench (Nosek et al., 2015; Heuritsch, 2020).

OS requires partnerships and collaborations between stakeholders including government, academic and research institutions, research funders, researchers, libraries, publishers, information and communication technology experts and end users of the research outputs (Kennedy and Ruttenberg, 2019). In the absence of these alliances, advancing OS in Africa remains challenging because these interdependencies are complementary in achieving OS (Krishna, 2020).

OPPORTUNITIES AND FUTURE CONSIDERATIONS

The Strides Toward OS in Africa

The OA movement in Africa is gradually growing. As at 2015, over 500 OA journals published in Africa are captured in the African Journals Online (AJOL) and Directory of Open Access Journals (DOAJ). Meanwhile, 125 OA digital repositories in Africa are registered in the Directory of Open Access Repositories (OpenDOAR) while 18 OA policies from the region are listed in the Registry of Open Access Repository Policies and Mandates (ROARMAP). Ever since, a lot more African researchers also publish in international OA journals. Some African institutions are on the lead and have taken initiatives to boost OA movement in Africa. For instance, Stellenbosch University offers on-site trainings and shares valuable materials online to new OA repositories while the Academy of Science of South Africa (ASSAf) in partnership with UNESCO Cluster Office in Southern Africa offers training to OA journal publishers in the region (Academy of Science of South Africa, 2022).

International organizations like Electronic Information for Libraries (EIFL) and International Network for Advancing Science and Policy (INASP) support libraries in the region while the Irish African Partnership for Research Capacity Building (IAPRCB), joins several universities in Ireland, Malawi, Mozambique, Tanzania, and Uganda to develop a harmonized approach to research capacity building through its OA repository. The International OA Week has been made annual events in some African research institutions and helps in raising consciousness among academic communities of the region about the benefits of OA.

In addition, the Southern African Regional Universities Association (SARUA), representing 66 public universities in Southern Africa, published a research report on Opening Access to Knowledge in Southern Africa, recommending OA as a potential strategy for Africa. To some extent, all major stakeholders—researchers, research administrators, policy makers, journal editors, publishers, librarians, OA experts, students and general public—have started to realize the benefits of OA and seem to be making efforts, albeit little, to implement OA projects in the region.

UNESCO had shown their readiness to work with African countries willing to work toward national OA policy and also called for training centers to build capacity and expertise on OA ideas and structures. In 2015, UNESCO, Network of African

Science Academies (NASAC), Royal Netherlands Academy of Arts and Sciences, African Academy of Sciences (AAS), Kenya National Academy of Sciences, and Kenyan Ministry of Education, Sciences and Technology jointly hosted a consultative meeting on OA for Africa in Kenya bringing together about 45 top policy makers and expert representatives of 20 countries of Africa providing intervention for research and development in Africa.

The recent landmarks on OSs created the opportunity for OS discussions in Africa in recent times. The Council for the Development of Social Science Research in Africa (CODESRIA) hosted her conference in 2016 titled "Electronic Publishing: OA Movement and the Future of Africa's Knowledge" in Dakar, Senegal where opportunities and challenges to the OS movement in Africa were discussed. This led to a Dakar Declaration on OS in Africa where all the signatories agreed to promote and support OS across Africa by organizing events on OS which will target both students and researchers. To foster scientific innovation and capacity to contribute to global scientific research output, African countries should be provided with virtual high-tech laboratories under an open license in addition to the standard OA materials such as course materials, textbooks, multimedia applications etc. This trend is beginning to emerge as there are online scientific laboratory initiatives in Africa that will boost the commitment to OS as an academic practice. Eighteen African governments have established the African Virtual University Project (AVU), a leading eLearning Network in Africa with the vision to meaningfully increase access to quality higher education and training through the innovative use of information communication technologies (ICTs) covering several scientific disciplines. Similarly, Africanfossils.org has also been established as a free online virtual lab for scholars to explore and interact with fossil collections under the Creative Commons Attribution-Non Commercial-Share Alike License (Canton, 2021).

The LIBSENSE initiative launched in 2017 has been building a community of OS practitioners and making headway for the adoption of OS services and structures in Africa. The initiative was led by West and Central African Research and Education Network (WACREN) in partnership with a number of other organizations with the aim to advance OS in Africa through strengthening and expanding services at the institutional, national and regional level (Abbott et al., 2020; WACREN, 2021). In 2021, a virtual workshop was co-organized by LIBSENSE partners, EIFL, WACREN, UbuntuNet Alliance, the Arab States Regional Education Network (ASREN), Confederation of Open Access Repositories (COAR), AfricaConnect 3, OpenAIRE, and GÉANT, which showcased the current national level activities on OS policies, repositories, community building and coordination in 15 African countries. As a follow up, three working groups were set up, on (1) OS policies, governance and leadership; (2) OS infrastructure—OA journals, repositories for publications and data and open discovery services; and (3) capacity building—communities of practice and training. In addition, this workshop set up region-specific and language-specific discussions in Arabic (North Africa) and French (West and Central Africa) working groups who were charged with co-developing guidelines, support and training materials (Abbott et al., 2020; WACREN, 2021).

LIBSENSE also started a series of open community calls, titled "Co-designing collaborative free and open source-OA publishing infrastructures", organized by WACREN, EIFL, and the Coko Foundation, with African journals and books editors and publishers, researchers, librarians and tool builders. In these calls, they discussed needs and tools for OA scholarly publishing in Africa; what open source tools and services for publishing books, journals and textbooks are currently in use, and the training and support needs (Abbott et al., 2020; WACREN, 2021; Kuchma, 2022). LIBSENSE recognizes that OS in Africa, with respect to diversity and sustainable development, can be best realized through localized, yet interoperable, infrastructures—rather than being subcontracted to private industry or external organizations. These services will be able to more directly answer to the necessities of African research communities, and also contribute to building local capacity and knowledge around OS (Abbott et al., 2020). It is important to note that the UNESCO OS Partnership has put OS on the national agenda of several African governments. Taking advantage of this strategic opportunity, LIBSENSE has begun to work with several African countries that are committed to advance OS policies, infrastructures and services to develop African National OS Roadmaps that can then be piloted in other African countries (Abbott et al., 2020).

The above discussions illustrate a trend in OS efforts in recent times. The high tendency to work in isolation by African scientists and scientific organizations has implications in the effectiveness and efficiency of science systems, thus limiting the needed collaboration to address the complex problems of OS in Africa (Bezuidenhout et al., 2017). The recent endorsement of the Africa Continental Free Trade zone Agreement among member states of the African Union (AU) and the AU's efforts toward actualizing United Nations Sustainable Development Goals, could afford the stimulus needed for development of an inclusive and strong OS initiative in Africa (UNSDGS, 2015). We hereby identify and discuss the opportunities and future considerations for OS in Africa below.

Globally, OS movement has been witnessing an unprecedented increase in its embrace. Although weakly implemented in Africa according to the UNESCO report in 2015 (Mwelwa et al., 2020), there is a gradual and steady adoption of this movement currently with variation between the English and French-speaking African countries (Ahinon and Havemann, 2018). Several OS initiatives have been established in Africa or for African researchers under the various elements that constitute the concept of OS. For example Open Science in Haiti and Francophone Africa (SOHA) project and African Open Science Platform (AOSP) are under OA (advocacy and publishing) initiatives while Africa Open Science and Hardware (AfricaOSH) is an example of an initiative or platform for OS hardware. We highlighted in this review some major OS platforms or initiatives in Africa (**Table 1**).

Open Science Policy and Policy Makers

Government institutions in Africa have adopted an open government charter that requires them to open some of their data assets, in a way that many national statistical offices now collaborate internationally in developing open data practices

(Nordling, 2015). Despite the 18 OA policies from Africa registered in the Registry of OA Repository Policies and Mandates, Africa is still in shortage of functional policies and policy making bodies on OS. The areas of OS where policies are required include funding, research data management, IP, and copyright (Mwelwa et al., 2020), It is important that IP protection is well-balanced to protect the rights of originators without reducing the chances of innovation (Mwelwa et al., 2020). Again, Africa has been criticized for poor performances with implementation of some global policies. African states accept policies as a working document but usually do not drive the implementation due to the lack of strong systems, stable leadership and integrity. Irrespective of reports of suboptimal success with several other African policies, it is necessary that Africa makes an evidence-based attempt to respond to OS campaigns with a dependable and unifying policy. There is lack of key policies that are necessary for the advocacy and industrialization of OS in Africa both at the national and institutional levels. The enactment of regulatory policies or regional framework for OS is paramount, from the acquisition of data, data usage, management, publication, translation, and re-use.

There are key organizations in Africa (both private and government-owned) that work in the field of OS and have some form of internal frameworks to fulfill their set objectives (**Table 1**). However, some of these organizations may not be able to single-handedly unite Africa on an operational model for OS in Africa. The question to ask is, "do these existing organizations in Africa have such potential and capacity to drive the conversations around a holistic OS policy for Africa?" The African Union Commission (AUC) through her department of Science, Technology and Innovation had made a recent attempt to bring experts and stakeholders together through the African Regional Multi-stakeholder Meeting on OS in 2020 (UNESCO, 2020). While this is a good first step, it is imperative that this be sustained. It is not clear if there is a roadmap, timeline, and strategy for achieving the set goals declared in the regional meeting. More so, there may be a need to involve or get technical support from global OS platforms like International Science Council (ISC), UNESCO and other collaborators in future conversations to share their experiences for early identification of potential pitfalls. There is a need to set up a special caretaker committee, for example within the AUC to organize, supervise, implement, monitor and lead the development of OS policies for Africa. However, a strong leadership system within the AUC and African countries is needed to drive the sustainability of the above-mentioned initiatives.

The capacity for Africans to fully exploit the opportunities presented by digital revolution that could drive innovation and development on the continent would be greatly enhanced by a strong, multi-state OS system (Boulton et al., 2020). In this digital age, Africa needs to explore digital technology for the continent to economically benefit from the fourth industrial revolution (Ndung'u and Signé, 2020). OS stakeholders in Africa need to take advantage of Africa's increasing interest in Internet and other digital technologies (Kende, 2021). Hence, technology must be at the forefront of any policy being developed by Africa on OS. A good place to start in the consideration of OERs is to have clear and effective policies on IP right and copyright. For example, concerning intellectual capital, a clear policy would plainly lay out the corresponding rights of the institution, its employees, students, or contractors, who are involved either directly or indirectly with sharing materials. In policy negotiations, it is important to consider the relative benefits of making flexible copyright policies that spontaneously apply open licenses to contents except for serious reasons to retain all-rights reserved copyright over such contents. Concurrently, any of such policies should still make it easy for anyone to invoke all-rights reserved copyright whenever necessary. Human resource policies should be developed where non-existent or reformed by updating of costing/resourcing and performance management systems so that African scientists can be rewarded for: (1) time spent in educational resources development, (2) using resource-based learning (when it is more effective), (3) using available materials with similar content (when it is more cost-effective) than producing a new one, and (4) sharing intellectual capital through global networks to improve resources, personal and institution's profile.

The AU's Agenda 2063 (AU, 2015) and the United Nations Sustainable Development Goals (UNSDGs, 2015) will be far from being met in Africa if OS is not prioritized through a working policy for Africa. Again, some pre-existing policies, habits and processes that had earlier been helpful are now obvious inhibitors of innovation, and need to change (Mwelwa et al., 2020). Africa needs to be involved critically in the discussions of private sector monopolization of scientific data by the ISC (International Science Council, 2019) as they are major preys to this. African policy makers must bear in mind the critical areas to focus in their policy review or engagements. These have been well-outlined by Mwelwa et al. (2020):

> "*managing, curating and using large and diverse data volumes, developing the incentives, methods and standards for data sharing, maintaining security against malign interventions, ensuring the preservation of ethical standards, developing the systems and software to undertake all these tasks and keeping abreast of the rapidly evolving state of the art in data science*".

Scientists need customized research systems to do robust and transparent science especially in Africa where young researchers are working to build OS practices from the scratch. The needs of African science communities are different from those that are part of more developed research systems. Hence, there is need to drive sustainable policies and their implementation that will reflect each country's needs and ensure consistent growth of OS in Africa (Onie, 2020). An important question to ask regarding OS policy development is whether OS policy should be institutional specific or whether this could be standardized across African institutions.

Funding for Open Science

Funding for OS in Africa remains untraceable, unstable or unsustainable. There is neither funding documented for the African region nor African organizations documented as funders

in the UNESCO's global OA portal funding mandates for OA for world regions. There is a need to apply funding toward retooling universities for research. For example, Nigerian librarians struggle to acquire current scholarly literature and modern technology for their libraries because of constant budget cuts (Okere, 2020). Dedicated and reliable funds are necessary to sustain scientific research in higher institutions. Funding is necessary for the provision of support staff, investment in research travels and establishment of procedures for data collection, grants management, and ethics. These provisions would enable effective teaching and research in science.

It is unclear how many African institutions have self-owned OERs and repositories (Bezuidenhout et al., 2020). These resources are either occasionally accessible or non-existent in most African institutions perhaps because of the low funding for African universities and research institutions (Bezuidenhout et al., 2020). It should become compulsory to establish universities with these resources being considered as part of its accreditation criteria to create a systematic funding for OS in African institutions.

Openness in research demands an e-infrastructure that will expedite information sharing. Therefore, there is a need for accessibility of reliable internet connection that can enable data sharing, especially for large datasets usually seen in bioinformatics and imaging research. In addition to a reliable internet, storage and high-performance computing capabilities are also essential (Maphosa, 2019). In a poor-resource setting like Africa, e-infrastructure to support OS is inadequate (Maphosa, 2019). However, the development of National Research and Education Networks (NRENs) in Africa is advancing speedily (Foley, 2016). NRENs provide internet infrastructure and services to support research and educational activities within a country (Dyer, 2009). The ICT infrastructure provided by NRENs is key in facilitating OS practices, starting with the open scholarly literature through institutional repositories in African universities and research institutions. Again, NRENs offer OERs and facilitate research collaborations between local and international scientists (Foley, 2016).

The major funders for OS in Africa include African Development Bank, African Union, research and health ministries, heads of state, Gates Foundation, Chan Zuckerberg Initiative, World Bank, and others who have given considerable support to African-led projects and networks that reinforce research. Examples include the AOSP (funded by the National Research Council of South Africa) and the Alliance for Accelerating Excellence in Science in Africa (AESA)—partnership of AAS, NEPAD Agency with US$5.5 million initial seed funding from the Bill and Melinda Gates Foundation, the Wellcome Trust and the UK Department for International Development (UkDFID). These funders provide not only funds for future infrastructural development, but also expertise, and contacts to international expertise and national/regional governance stakeholders (Havemann et al., 2020).

In the face of inadequate funding for science in most African countries, it is important that countries utilize free open source hardware (FOSH) to be able to avert the cost implications of scientific equipment or to reduce them drastically (Maia Chagas, 2018). Maia Chagas (2018) has documented all the notable FOSHs for scientific research and education with the associated links. Also, it is important that OS actors in Africa: (1) identify ongoing projects and funders addressing infrastructural reform on OS, (2) approach funders and governments directly, and (3) create a record of grants on OS, to co-apply and consolidate for overlapping activities to maximize funding opportunities. Governments and international bodies can adopt the performance based financing for OS initiatives at any level to be able to attract and sustain effective and efficient initiatives for African countries (Sieleunou et al., 2017). It will be waste of scarce resources for lower-income African countries to fund research without expected scientific rigor and integrity. Thus, funding policies should not just be targeted at increasing output but also intended to improve relevance, transparency, and scientific rigor, especially if research outputs are geared toward being useful for decision-making in Africa. Governments should provide the motivation and training resources needed for people to imbibe the policy changes. It is believed that when there is an increased impact of OS funding through increased innovation or productivity, countries will be more willing to commit more to financing OS initiatives. Overall, "investments should generate a virtuous cycle in which long-term changes in research output yield more government and international funding" (Onie, 2020).

Advocacy and Incentives

The sustainable culture of OS can thrive in Africa if there is adequate institutional or individual advocacy plans toward OS. First, we need to start a global advocacy for the relevance of the scientific content emanating from Africa (Pennisi, 2021). The exclusion and neglect of science done in Africa and by Africans in the global scientific decisions and policies contributes to the disinterest of African researchers from OS practices (Mwelwa et al., 2020). Such unconscious bias affects the understanding of the natural world, and makes it more difficult for researchers from Africa to operate effectively (Harris et al., 2017). Getting included in systems where OS is already established will boost the uptake level of OS by African researchers. Following the authors' experience and peer reports, there are none or few Africans in the editorial boards of notable open access science journals. This inequality may contribute to the understanding of science-based issues in Africa, and their acceptance for publications (Onie, 2020). Again, it is a common experience with the authors and their peers on increased non-acceptance of scientific articles from African researchers by top scientific journals on the bases of being insignificant to the wider readership or simply because of using a traditional method, usually still in use in Africa (Onie, 2020). A study presented US researchers with identical abstracts and observed that the researchers were more likely to recommend the article to a peer if its authors were listed as being from the United Kingdom than if they were from Malawi (Harris et al., 2017). Advocacies against these biases will spike the confidence of African researchers toward OS practices. Journals should take the lead in reducing under-representation, while maintaining scientific rigor, and authors should explicitly describe their study populations ahead of time, and not generalize

their findings beyond the study population without any good justification. Double-blind reviews have been used to tackle the positive bias experienced by prestigious institutions or authors (2015) while the use of Open reviews could reduce potential bias against studies from African researchers (Carroll et al., 2017; Onie, 2020). Open publication systems will give the reputable non-African journal publishers the opportunity to share ideas on publication standards especially relating to OS policy which African researchers and publishers could adopt. More recently, we have also experienced special issues focused on the under-represented populations. The sustenance of such cultures will not only encourage equity and remove bias from global interpretation of scientific data, but also give more opportunity for OA publications amongst African scientists.

Strong advocacies against some academic practices which discourage open and rigorous science in Africa is necessary to achieve a quickened OS culture in African institutions. Many African institutions judge their faculty members according to Western standards, including publishing in "prestigious" journals. The pressure to publish at all costs to meet certain promotional criteria is one of the biggest challenges to creating credible scientific output from Africa (Rawat and Meena, 2014).

Again, the use of metrics for faculty appraisals should only be applied if they seem useful to the overall goal for science—knowledge accumulation for the greater societal good. For example, AAS open research platform does not utilize the impact factor metric system characteristic of many journals as they describe it as flawed. The AAS Open Research model is part of advancement in scientific publishing that berates the use of such measures. Individual articles published in the AAS platform displays article-level metrics as and when applicable, such as Altmetrics; PubMed citations for articles that have passed peer review; and the number of views and PDF downloads on AAS Open Research and in PubMed Central.

It is proper to introduce an incentive system to encourage more African researchers to easily adopt OS. Some incentives that should be supported at all levels includes provision of research grants, publication grants, travel grant, training grants and also rewarding OS practice during promotional assessments. However, it is essential to know that the best ideas to improve science today may become less useful in the future. For example, China recently stopped its cash-for-publication system after realizing its impact on the quality of publications. This is a proof that sustainable change to good behavior cannot be achieved when people are indirectly incentivized to do the opposite (Mallapaty, 2020). The reward system should encourage research cultures that can guard against harmful practices and lay down a good strategy for OS advancement in Africa (Mallapaty, 2020).

Collaboration and Networking

Collaborative research in sciences is relatively low in Africa and needs to improve (Onyancha and Maluleka, 2011; Pouris and Ho, 2014). Scientists from growing research cultures like Africa should be encouraged to join societies and conferences hosted in countries with more openness in science, to create the mutual scientific exchange necessary for OS behaviors. There is a need to acknowledge and confront the isolation of African scientists from opportunities for international collaboration. Inadequate funding and travel restrictions in many countries in Africa reduce the opportunities for networking, international collaboration and lead to more isolation of African researchers (Ochola and Gitau, 2009; Ranganath, 2017; Kasprowicz et al., 2020; Marincola and Kariuki, 2020). It is necessary that efforts to make space for African researchers be focused on empowerment and based on mutual respect, rather than taking control of their systems (Minasny et al., 2020). Strong partnership and collaboration policies must be in place to ensure collaborations between African researchers and researchers from the developed countries are true partnerships (Minasny et al., 2020). More collaboration amongst African researchers with OS requirements should be initiated and incentivized by OS stakeholders. The multilingual nature of Africa creates opportunity for collaborations that are not limited by language bias. Such initiatives should be sponsored by scientific societies and other stakeholders of OS at all levels to encourage the deepening of OS practices and entrenching the OS culture amongst African scientists.

Training and Capacity Building

The question to ask as Africa embraces advancements in OS is "who needs to be trained?" This should be informed by research and not by assumption. We posit that every length and breadth of the stakeholder chain in African OS system may require some form of training or the other (Onie, 2020). Having a few persons or organization which shows some degree of expertise may not be enough for the wholesome intervention needed in Africa. Stakeholders must first assess the training needs of open system drivers in Africa and her beneficiaries to ensure its effective implementation, usage and replication. Awareness must be created on OS tools and skills for OS practice to become easily adoptable. The European Commission OS Skills Working Group recommended that researchers should be sensitized and trained on OS practices that ease the practice of OS such as open research, OA, open education, open data, open peer review, and citizen science. Capacity building to enable effective use of OER should involve: (*1) supporting policy-makers and heads of institutions to understand the key elements necessary to create supportive policy environments, develop materials, use technology, and conduct research; (2) identifying best-practice examples of use of OER and facilitating institutional visits, so that participants have an opportunity not only to observe effective use of OER in practice but also to start developing support networks and communities of practice* (Organisation for Economic Co-operation Development, 2007).

The essential skills needed for institutions to effectively utilize OER include:

- OER advocacy and promotion skills.
- Knowledge on content licensing legal framework.
- Skill in business model development, course development and programme design.
- Technical know-how and network management.
- Expertise in monitoring and evaluation.
- Skills in effective curation and sharing OERs.

- Research and communication skills for information sharing (Organisation for Economic Co-operation Development, 2007).

To aid training in OS, free online resources should be readily available. In recent times, there has been a surge in written materials, YouTube videos, and mass open online courses in different languages (Onie, 2020). Beyond training on OER, African scientists need competency training on specific areas of science especially areas where expertise is generally lacking or where scientific infrastructure is limited to improve their global participation in those areas. Training in good scientific practices will position African scientists to be more critical and adopt practices that improve the integrity of their work. It will also allow them to add their diverse voices to the on-going multifaceted debates on OS, paying attention to the benefit of science to the society, locally and globally (Onie, 2020). Training should also boost scientists' career trajectories especially now that institutions are beginning to seek for evidence of OS practice as criterion for recruitment (Onie, 2020). Institutionalizing OS in Africa is critical to capacity building for the next generation scientists in the region. OS is needed not just for OS-related capacity building but also for easy access to technological and scientific skills which are lacking in Africa, that are necessary to drive the scientific innovations and development in the region.

STUDY LIMITATIONS

This study might have missed out some important articles on OS published in other languages other than English, considering the multilingual nature of the African continent. Again, we could not access a few non OA articles at country or organizational level which may also be useful to our evidence synthesis. The evidences available in literature may not be reflective of the true situation in some countries of Africa or may have since changed especially as there are a few or no information on open science practices in many African countries. However, we carefully executed the literature selection of this exploratory review and ensured contextual interpretation of our findings to enhance the usefulness of this study as an advocacy tool for OS in Africa.

CONCLUSION

In this exploratory review, we critically analyzed a reemerging issue in Africa—OS—occasioned by the COVID-19 pandemic. We highlighted OS benefits, OS platforms, challenges, and opportunities in Africa. OS offers several benefits to the development of science in Africa including but not limited to sharing of scientific data. It provides opportunities for all to work together at the same time building understanding of research procedures, practices, and findings. OS stakeholders need to promote, utilize, and upscale the OS platforms and initiative highlighted in this study and only then can they benefit from OS as themed above: networking and collaboration, public engagement and public trust, visibility and impact, scientific community and, innovation and commercialization. Some of the major challenges that plague OS in Africa were highlighted in this review but the lack of funding for science seems more critical as it directly impact other challenges. We have created a conceptual framework for creating OS solutions in Africa following the evidences generated from the literature on the challenges of OS in Africa (**Figure 3**). This is a two-prong framework that shows the non-sequential overlap and interdependence of funding and leadership as the pivot for creating thriving OS systems in Africa. As discussed above, funding is key to delivering many OS initiatives and strategies; and these OS solutions are to be steered by dependable leadership system built across education, research and African governments at different levels (**Figure 3**). It is important to note that there are inequalities in the practice of OS in Africa even amongst African countries, individual researchers or organizations as a result of varying challenges. Hence, there is need to conduct both qualitative and quantitative research to understand organizational and individual perspectives of OS practice especially concerning challenges and choices of OS practice. This will enable the creation of a more sustainable advocacy or implementation that works for each country. Policy making in OS must take into consideration the context of each country's challenges to maximize opportunities in its implementation. OS still remains a significant contributor in solving global problems, and thus a potent channel for Africa's development. Hence, institutionalizing OS in Africa should be

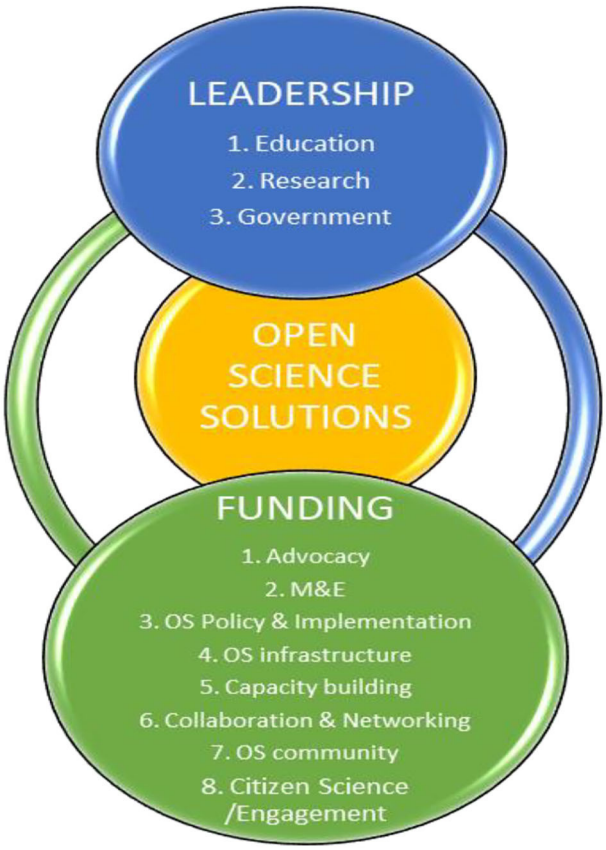


FIGURE 3 | Conceptual framework for creating open science solutions in Africa. M&E, monitoring and evaluation; OS, open science.

on the forefront of science stakeholders in Africa more than ever before, especially due to the current pandemic.

AUTHOR CONTRIBUTIONS

IO and SM: conceptualization. IO, SM, TC, ZH, EU, and KC: methodology, data curation, visualization, validation, resources, writing—original draft, and writing—review and editing. IO: project administration. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We appreciate Dr. Adibi M. Nor (Senior Lecturer, University of Malaya, Malaysia) and Mr. Umeasalugo Kosisochukwu (Ph.D. Candidate, Ludwig Maximilians University, Germany) for their comments and peer review feedback on the first draft of the manuscript. We acknowledge Scholarly Publishing and Academic Resources Coalition (SPARC), the Right to Research Coalition (RRC), and Max Planck Society for cohosting OpenCon 2017 conference on November 11-13 in Berlin, Germany, where the corresponding author first conceived this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frma.2022.855198/full#supplementary-material

REFERENCES

(2015). Nature journals offer double-blind review. *Nature* 518, 274. doi: 10.1038/518274b

Abbott, P., Appiah, K., and Oaiya, O. (2020). *Barriers and Enablers to Open Access Repository (OAR) Development and Management in African HLIs: Research from the LIBSENSE OAR Workshops in the UA, WACREN and ASREN Regions*. West and Central African Research and Education Network. Available online at: https://zenodo.org/record/3884974/files/WACREN%202020%20Full%20Paper_Final.pdf?download=1 (accessed January 09, 2022).

Academy of Science of South Africa (2022). *Scholarly Publishing Programme*. Available online at: https://www.assaf.org.za/index.php/programmes/scholarly-publishing-programme (accessed January 09, 2022).

Adeyemo, O. O., and Jamogha, E. (2021). *Institutional Repository as a Catalyst for Enhanced University Visibility: The Case of Obafemi Awolowo University*. Covenant Journal of Library and Information Science, 4(1). Covenant University, Ota, Nigeria. Available online at: https://journals.covenantuniversity.edu.ng/index.php/cjlis/article/view/2654

Ahinon, J., and Havemann, J. (2018). *Open Science in Africa - Challenges, Opportunities and Perspectives*. Zenodo. doi: 10.5281/zenodo.1492745

Ali-Khan, S. E., Jean, A., and Gold, E. R. (2018). Identifying the challenges in implementing open science. *MNI Open Res.* 2, 5. doi: 10.12688/mniopenres.12805.1

Allen, C., and Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biol.* 17, e3000246. doi: 10.1371/journal.pbio.3000246

Bautista-Puig, N., Lopez-Illescas, C., Moya-Anegon, F., de, Guerrero-Bote, V., and Moed, H. F. (2020). Do journals flipping to gold open access show an OA citation or publication advantage? *Scientometrics* 124, 2551–2575. doi: 10.1007/s11192-020-03546-x

Beck, M. W., O'Hara, C., Stewart Lowndes, J. S., D., Mazor, R., Theroux, S., et al. (2020). The importance of open science for biological assessment of aquatic environments. *PeerJ* 8, e9539. doi: 10.7717/peerj.9539

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., et al. (2021). Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med. Res. Methodol.* 21, 117. doi: 10.1186/s12874-021-01304-y

Bezuidenhout, L., Havemann, J., Kitchen, S., Mutiis, A., and de Owango, J. (2020). *African Digital Research Repositories: Mapping the Landscape*. doi: 10.5281/ZENODO.3732273

Bezuidenhout, L., Kelly, A. H., Leonelli, S., and Rappert, B. (2017). '$100 is not much to you': open science and neglected accessibilities for scientific research in Africa. *Crit. Public Health* 27, 39–49. doi: 10.1080/09581596.2016.1252832

Bliss, T. J., and Blessinger, P. (2016). *Open Education: International Perspectives in Higher Education*. Cambridge: Open Book Publishers.

Boulton, G., Loucoubar, C., Mwelwa, J., Muliaro, J. M., Ozor, N., and Bolo, M. (2020). *Open Science in Research and Innovation for Development in Africa*. African Technology Policy Studies Network (ATPS).

Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., et al. (2020). Digital technologies in the public-health response to COVID-19. *Nat. Med.* 26, 1183–1192. doi: 10.1038/s41591-020-1011-4

Busch, S. (2014). *The careers of converts - how a transfer to BioMed Central affects the Impact Factors of established journals*. Available online at: https://blogs.biomedcentral.com/bmcblog/2014/01/15/the-careers-of-converts-how-a-transfer-to-biomed-central-affects-the-impact-factors-of-established-journals/

Canton, H. (ed.). (2021). *The Europa Directory of International Organizations: 2021*. Abingdon, OX, New York, NY: Routledge,.

Capps, B. (2021). Where does open science lead us during a pandemic? A public good argument to prioritize rights in the open commons. *Camb. Q. Healthc Ethics* 30, 11–24. doi: 10.1017/S0963180120000456

Carroll, H. A., Toumpakari, Z., Johnson, L., and Betts, J. A. (2017). The perceived feasibility of methods to reduce publication bias. *PLoS ONE* 12, e0186472. doi: 10.1371/journal.pone.0186472

Christie, R. (2019). African science: Better but still inadequate. *S. Afr. J. Sci.* 115(7/8), Art. #6233. doi: 10.17159/sajs.2019/6233

Crow, R. (2002). *The Case for Institutional Repositories: A SPARC Position Paper*. Washington, DC: SPARC.

Das, A. K. (2011). Emergence of open educational resources (OER) in India and its impact on lifelong learning. *Library Hi Tech News* 28, 10–15. doi: 10.1108/07419051111163848

David, P. A. (2004). Understanding the emergence of 'open science' institutions: functionalist economics in historical context. *Indus. Corporate Change* 13, 571–589. doi: 10.1093/icc/dth023

Eagleman, D. M. (2013). Why public dissemination of science matters: a manifesto. *J. Neurosci.* 33, 12147–12149. doi: 10.1523/JNEUROSCI.2556-13.2013

European Commission, Directorate-General for Communications Networks, Content and Technology (2018). *Commission Recommendation*.

European Union (2017). *Implementing Open Science: Strategies, Experiences and Models*. Euro Available online at: https://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fec.europa.eu%2Fresearch-and-innovation%2Fsites%2Fdefault%2Ffiles%2Frio%2Freport%2FMLE%252520OS_4th%252520thematic%252520report.pdf&clen=912741&chunk=true

Ewers, M., Ioannidis, J. P., and Plesnila, N. (2021). Access to data from clinical trials in the COVID-19 crisis: open, flexible, and time-sensitive. *J. Clin. Epidemiol.* 130, 143–146. doi: 10.1016/j.jclinepi.2020.10.008

Exley, K., Cano, N., Aerts, D., Biot, P., Casteleyn, L., Kolossa-Gehring, M., et al. (2015). Communication in a Human biomonitoring study: focus group work, public engagement and lessons learnt in 17 European countries. *Environ. Res.* 141, 31–41. doi: 10.1016/j.envres.2014.12.003

Ezema, I. J. (2011). Building open access institutional repositories for global visibility of Nigerian scholarly publication. *Library Rev.* 60, 473–485. doi: 10.1108/00242531111147198

Fleming, J. I., Wilson, S. E., Hart, S. A., Therrien, W. J., and Cook, B. G. (2021). Open accessibility in education research: enhancing the credibility,

equity, impact, and efficiency of research. *Educ. Psychol.* 56, 110–121. doi: 10.1080/00461520.2021.1897593

Foley, M. (2016). *The Role and Status of National Research and Education Networks in Africa*. Washington, DC: World Bank.

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., et al. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biol.* 19, e3000959. doi: 10.1371/journal.pbio.3000959

Gao, F., Tao, L., Huang, Y., and Shu, Z. (2020). Management and data sharing of COVID-19 pandemic information. *Biopreserv. Biobank.* 18, 570–580. doi: 10.1089/bio.2020.0134

Granados Moreno, P., Ali-Khan, S. E., Capps, B., Caulfield, T., Chalaud, D., Edwards, A., et al. (2019). Open science precision medicine in Canada: points to consider. *Facets* 4, 1–19. doi: 10.1139/facets-2018-0034

Guo, Y., Zhang, M., Bonk, C. J., and Li, Y. (2015). Chinese faculty members' open educational resources (OER) usage status and the barriers to OER development and usage. *Int. J. Emerg. Technol. Learn.* 10, 59. doi: 10.3991/ijet.v10i5.4819

Harding, K. (2016). Global health innovation technology models. *Nanobiomedicine* 3, 7. doi: 10.5772/62921

Harris, M., Marti, J., Watt, H., Bhatti, Y., Macinko, J., and Darzi, A. W. (2017). Explicit bias toward high-income-country research: a randomized, blinded, crossover experiment of English clinicians. *Health Affairs* 36, 1997–2004. doi: 10.1377/hlthaff.2017.0773

Havemann, J., Bezuidenhout, L., Achampong, J., Akligoh, H., Ayodele,O., Hussein, S., et al. (2020). *Harnessing the Open Science Infrastructure for an Efficient African Response to COVID-19*. Zenodo. doi: 10.5281/ZENODO.3733768

Hayashi, K. (2021). How could COVID-19 change scholarly communication to a new normal in the open science paradigm? *Patterns* 2, 100191. doi: 10.1016/j.patter.2020.100191

Heise, C., and Pearce, J. M. (2020). From open access to open science: the path from scientific reality to open scientific communication. *SAGE Open* 10, 215824402091590. doi: 10.1177/2158244020915900

Hetu, M., Koutouki, K., and Joly, Y. (2019). Genomics for all: international open science genomics projects and capacity building in the developing world. *Front. Genet.* 10, 95. doi: 10.3389/fgene.2019.00095

Heuritsch, J. (2020). "Knowledge utilization and open science policies: noble aims that ensure quality research or "ordering discoveries like a pizza"?" in *IAF Symposium on Future Space Astronomy and Solar-System Science Missions 2019* (Red Hook, NY: Curran Associates, Inc.).

Hudson, M., Garrison, N. A., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., et al. (2020). Rights, interests and expectations: indigenous perspectives on unrestricted access to genomic data. *Nat. Rev. Genet.* 21, 377–384. doi: 10.1038/s41576-020-0228-x

Kadakia, K. T., Beckman, A. L., Ross, J. S., and Krumholz, H. M. (2021). Leveraging open science to accelerate research. *N. Engl. J. Med.* 384, e61. doi: 10.1056/NEJMp2034518

Kasprowicz, V. O., Chopera, D., Waddilove, K. D., Brockman, M. A., Gilmour, J., Hunter, E., et al. (2020). African-led health research and capacity building- is it working? *BMC Public Health* 20, 1104. doi: 10.1186/s12889-020-08875-3

Kedron, P., Li, W., Fotheringham, S., and Goodchild, M. (2021). Reproducibility and replicability: opportunities and challenges for geospatial research. *Int. J. Geograph. Inform. Sci.* 35, 427–445. doi: 10.1080/13658816.2020.1802032

Kende, M. (2021). *Promoting the African Internet Economy - Internet Society*. Available online at: https://www.internetsociety.org/resources/doc/2017/africa-internet-economy/ (accessed January 09, 2022).

Kennedy, M. L., and Ruttenberg, J. (2019). *Implementing Open Science Principles through Research Partnerships*. Session 82 - Library Integration into the Research Lifecycle: a STEM Perspective - Academic and Research Libraries and Health and Biosciences, Athens. Available online at: http://library.ifla.org/id/eprint/2574/

Kokwaro, G., and Kariuki, S. (2001). Medical research in Africa: problems and some solutions. *Malawi Med. J.* 13, 40.

Krishna, V. V. (2020). Open science and its enemies: challenges for a sustainable science–society social contract. *J. Open Innov. Technol. Mark. Complex* 6, 61. doi: 10.3390/joitmc6030061

Kuchma, I. (2022). *Open science practices in Africa | EIFL*. Available online at: https://www.eifl.net/blogs/open-science-practices-africa (accessed January 09, 2022).

Lakomý, M., Hlavová, R., and Machackova, H. (2019). Open science and the science-society relationship. *Society* 56, 246–255. doi: 10.1007/s12115-019-00361-w

Leonelli, S., Rappert, B., and Bezuidenhout, L. (2018). Introduction: Open Data and Africa. *Data Sci. J.* 17, 5. doi: 10.5334/dsj-2018-005

Les Cottrell, R. (2013). Pinging Africa - a decadelong quest aims to pinpoint the internet bottlenecks holding Africa back. *IEEE Spectr.* 50, 54–59. doi: 10.1109/MSPEC.2013.6420148

Maia Chagas, A. (2018). Haves and have nots must find a better way: the case for open scientific hardware. *PLoS Biol.* 16, e3000014. doi: 10.1371/journal.pbio.3000014

Mallapaty, S. (2020). China bans cash rewards for publishing papers. *Nature* 579, 18. doi: 10.1038/d41586-020-00574-8

Maphosa, S. (2019). The academy of science of south africa and science diplomacy. *S. Afr. J. Sci.* 115, 9-10. doi: 10.17159/sajs.2019/a0314

Marincola, E., and Kariuki, T. (2020). Quality research in Africa and why it is important. *ACS Omega* 5, 24155–24157. doi: 10.1021/acsomega.0c04327

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., et al. (2016). How open science helps researchers succeed. *eLife* 5, 16800. doi: 10.7554/eLife.16800

Minasny, B., Fiantis, D., Mulyanto, B., Sulaeman, Y., and Widyatmanti, W. (2020). Global soil science research collaboration in the 21st century: time to end helicopter research. *Geoderma* 373, 114299. doi: 10.1016/j.geoderma.2020.114299

Momeni, F., Mayr, P., Fraser, N., and Peters, I. (2021). What happens when a journal converts to open access? A bibliometric analysis. *Scientometrics* 126, 9811–9827. doi: 10.1007/s11192-021-03972-5

Mwangi, K. W., Mainye, N., Ouso, D. O., Esoh, K., Muraya, A. W., Mwangi, C. K., et al. (2021). Open science in Kenya: where are we? *Front. Res. Metr. Anal.* 6, 669675. doi: 10.3389/frma.2021.669675

Mwelwa, J., Boulton, G., Wafula, J. M., and Loucoubar, C. (2020). Developing open science in Africa: barriers, solutions and opportunities. *Codata* 19, 31. doi: 10.5334/dsj-2020-031

Ndung'u, N., and Signé, L. (2020). "The fourth industrial revolution and digitization will transform africa into a global powerhouse," in *Foresight Africa: Top Priorities for the Continent 2020-2030* (Washington, DC), 61–66.

Nordling, L. (2015). Research: Africa's fight for equality. *Nature* 521, 24–25. doi: 10.1038/521024a

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science* 348, 1422–1425. doi: 10.1126/science.aab2374

Ochola, L. I., and Gitau, E. (2009). Challenges in retaining research scientists beyond the doctoral level in Kenya. *PLoS Negl. Trop. Dis.* 3, e345. doi: 10.1371/journal.pntd.0000345

Okere, O. O. (2020). *Swimming Against the tide: Oluchi Ojinamma Okere Considers Nigerian Academic Libraries and Their Struggle for Relevance*. Available online at: https://www.researchinformation.info/analysis-opinion/open-book-swimming-against-tide

Onie, S. (2020). Redesign open science for Asia, Africa and Latin America. *Nature* 587, 35–37. doi: 10.1038/d41586-020-03052-3

Onyancha, O. B., and Maluleka, J. R. (2011). Knowledge production through collaborative research in sub-Saharan Africa: how much do countries contribute to each other's knowledge output and citation impact? *Scientometrics* 87, 315–336. doi: 10.1007/s11192-010-0330-5

Organisation for Economic Co-operation and Development (2007). *Giving Knowledge for Free: The Emergence of Open Educational Resources*. Paris: OECD.

Paic (2021). *Making Data for Science as Open as Possible to Address Global Challenges*. Available online at: https://oecd-innovation-blog.com/2021/01/20/oecd-recommendation-access-research-data-public-funding-update-covid-19/

Pennisi, E. (2021). *Africans have begun to study their continent's rich human diversity-but what comes after current grants end?* Available online at: https://www.science.org/content/article/africans-begin-take-reins-research-their-own-genomes

Pouris, A., and Ho, Y.-S. +6(2014). Research emphasis and collaboration in Africa. *Scientometrics* 98, 2169–2184. doi: 10.1007/s11192-013-1156-8

Priem, J., Groth, P., and Taraborelli, D. (2012). The altmetrics collection. *PLoS ONE* 7, e48753. doi: 10.1371/journal.pone.0048753

Puebla, I. (2020). Preprints: a tool and a vehicle towards greater reproducibility in the life sciences. *J. Reprod. Neurosci.* 2, 1465. doi: 10.31885/jrn.2.2021.1465

Ramachandran, R., Bugbee, K., and Murphy, K. (2021). From open data to open science. *Earth Space Sci.* 8, e2020EA001562. doi: 10.1029/2020EA001562

Ranganath, M. (2017). *Three Reasons for the African Research Gap – And How to Close It*. William Davidson Institute.

Rawat, S., and Meena, S. (2014). Publish or perish: Where are we heading? *J. Res. Med. Sci.* 19, 87–89.

Shea, N. A. (2015). Examining the nexus of science communication and science education: a content analysis of genetics news articles. *J. Res. Sci. Teach.* 52, 397–409. doi: 10.1002/tea.21193

Sieleunou, I., Turcotte-Tremblay, A.-M., Yumo, H. A., Kouokam, E., Fotso, J.-C. T., Tamga, D. M., et al. (2017). Transferring the purchasing role from international to national organizations during the scale-up phase of performance-based financing in Cameroon. *Health Syst. Reform* 3, 91–104. doi: 10.1080/23288604.2017.1291218

Stagg, A. (2014). OER adoption: a continuum for practice. *RUSC. Univ. Know. Soc.* 11, 151. doi: 10.7238/rusc.v11i3.2102

Steiner, R., Tirivayi, A., Tirivayi, N., Jensen, M., Hamilton, P., and Buechler, J. (2005). *PAREN, Promoting African Research and Education Networking*. Ottawa, ON: IDRC.

Taylor, S. J. E., Anagnostou, A., Fabiyi, A., Currie, C., Monks, T., Barbera, R., et al. (eds.). (2017). "Open science: approaches and benefits for modeling and simulation," in *2017 Winter Simulation Conference (WSC)* (Las Vegas, NV).

Teferra, D., and Altbachl, P. G. (2004). African higher education: challenges for the 21st century. *Higher Educ.* 47, 21–50. doi: 10.1023/B:HIGH.0000009822.49980.30

UNESCO (2020). *Online Regional Consultation for Africa on the First Draft of the UNESCO Recommendation on Open Science*. Available online at https://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fen.unesco.org%2Fsites%2Fdefault%2Ffiles%2Freport_afr_regional_discussion_os_dec_2020.pdf&clen=399968&chunk=true

UNESCO (2021). *Draft Recommendation on Open Science.* Available online at: https://en.unesco.org/science-sustainable-future/open-science/recommendation (accessed 09 January 2022).

Venith, A. (2015). *Online Knowledge Portals for Development in Africa Scoping* Study.

WACREN (2021). *Communique*. Available online at: https://wacren.net/en/news/libsense-national-open-science-roadmaps-debuts-wacren-2021/ (accessed 09 January 2022).

Wendzel, S., Caviglione, L., Mazurczyk, W., and Lalande, J.-F. (2017). Network information hiding and Science 2.0: can it be a match? *Int. J. Electron. Telecommun.* 63, 217–222. doi: 10.1515/eletel-2017-0029

Yusuf, S., Baden, T., and Prieto-Godino, L. L. (2014). Bridging the Gap: establishing the necessary infrastructure and knowledge for teaching and research in neuroscience in Africa. *Metab. Brain Dis.* 29, 217–220. doi: 10.1007/s11011-013-9443-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Research Metrics and Analytics

TYPE Original Research
PUBLISHED 15 July 2022
DOI 10.3389/frma.2022.950212

OPEN ACCESS

EDITED BY
Katie Wilson,
Curtin University, Australia

REVIEWED BY
Joseph Wafula,
Jomo Kenyatta University of Agriculture and Technology, Kenya
Michelle Willmers,
University of Cape Town, South Africa

*CORRESPONDENCE
Lara Skelly
lara.skelly@gmail.com

SPECIALTY SECTION
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

RECEIVED 22 May 2022
ACCEPTED 27 June 2022
PUBLISHED 15 July 2022

CITATION
Skelly L and Chiware ERT (2022) African researchers do not think differently about Open Data. *Front. Res. Metr. Anal.* 7:950212. doi: 10.3389/frma.2022.950212

African researchers do not think differently about Open Data

Lara Skelly [1,2]* and Elisha R. T. Chiware [3]

[1]Loughborough University Library, Loughborough University, Loughborough, United Kingdom, [2]Stellenbosch Business School, Stellenbosch University, Stellenbosch, South Africa, [3]Cape Peninsula University of Technology Library, Cape Peninsula University of Technology, Cape Town, South Africa

A key motivation for Open Science is accessibility. For researchers in resource-poor economies, this translates into access to the methods, data and publications that will foster scientific research and discovery in such communities and environments. Attitudes toward Open Science are in flux, and there is a growing awareness of the roles and responsibilities that researchers have to one another in this regard. This paper explores how African researchers approach issues relating to Open Data by reporting on the *State of Open Data Report* data. Focusing on the attitudes toward Open Data, this paper reports on how African researchers view (i) data sharing, (ii) the use of shared data, and (iii) the Open Data ecosystem. The findings show that, although the attitudes of African researchers have changed over time, they are not very different from those held by their international counterparts. These findings will aid policymakers, as well as academic and research institutions, in highlighting the areas of future growth for Open Data in Africa.

KEYWORDS

Open Data, Africa, Open Science, attitudes, longitudinal

Introduction

Scientific research in the 21st century has been scaled to new practices where scientists work more collaboratively and in data-intensive environments (Tenopir et al., 2011). Advances in technology have enabled this increased scale, as well as the consistently high levels of investments and development in research infrastructures. The emergence of Open Science principles and the insistence by governments and leading global research funders to make publicly funded research more open and accessible for the public good is advancing data sharing in and across research domains. This collaborative and multidisciplinary nature of scientific research leads to significant changes in how research is conducted and, more importantly, how research data is managed and preserved. New 'best practice' procedures and resources are being enabled by the adoption of the Findable, Accessible, Interoperable and Reusable (FAIR) practices, which include data accessibility, discovery, reuse, preservation and, particularly, data sharing (Wilkinson et al., 2016). The other major driver of global Open Data sharing practices among scientific communities is the realization of the importance of community involvement through citizen science, especially in data collection and the utilization of research outputs in those communities.

The Royal Society (2012) points out that the "publication of scientific theories and the experimental and observational data on which they are based, permits others to identify errors, to support, reject or refine theories and to reuse data for further understanding and knowledge". Bird and Frey (2013) also emphasize that "Open Science entails the sharing of more than mere fact". A fundamental benefit of openly sharing data is safeguarding resources, especially time and money, while gaining knowledge leverage. Without systematic data processing that assigns the attribution of intellectual property to those who created data, scientists will always be skeptical about openly sharing data. The lack of clarity around attribution, quality, and responsibility erodes trust in sharing and complicates processes. In addition, data, due to its complexity and variability, if not properly managed, may not be discovered.

Consequently, research data management is crucial in making data systematically and logically storable, discoverable and accessible, and curated with appropriate metadata. Funding agencies, national governments and entities supporting research are enforcing a range of requirements to enable the long-term preservation and sharing of research data by encouraging both research publications and results that are "open" and readily accessible for the benefit of humankind and knowledge development. Ramsay (2022) emphasizes the important point that "data sharing is essential to the advancement of science", and Tenopir et al. (2011) also state that data sharing is a valuable part of the scientific method allowing for verification of results and extending research from prior results.

Various scientific research disciplines have, over the years, developed their own systems and protocols for sharing data both in and out of the laboratory. There has never been a mandatory approach to how data should be shared. Over the last decade and going into the future, the scientific community is witnessing a deluge of data and, consequentially, the development of systems to support the management of large data sets. In Chemistry, researchers are said to be lagging in recognizing the importance and value of curating their data and information for the purposes of exchanging it (Bird and Frey, 2013). Bird and Frey (2013) further point out that "the growth and complexity of datasets produced have encouraged the expansion of e-Research, and stimulated the development of methodologies for managing, organizing, and analyzing 'big data'". The growing e-Science and e-Research practices now underpin scientific research projects. Furthermore, calls from research funders for open access practices to research outputs and, more importantly, Open Data management approaches are driving research disciplines to find new solutions to share data.

In 2011, Tenopir et al. (2011) showed that in various researchers' practices, "barriers to effective data sharing and preservation were deeply rooted in the practices and culture of the research process as well as the researchers themselves". By 2015, Tenopir et al. (2015), continuing on the same research that tracks researchers' data sharing practices, reported a shift in behavior. Their results point to (i) the increased acceptance of and willingness to engage in data sharing and (ii) an increase in actual data-sharing behaviors. Tenopir et al. (2015) further noted increased perceptions about the risk associated with data sharing and that specific barriers to data sharing persisted. It was also reported that there are differences across age groups, with younger respondents feeling more favorably toward data sharing and reuse yet making less of their data and research available or "open" than older respondents. The generational differences could be attributed to fear of competition in career development and promotion. Geographic differences were also noted to exist and were understood in terms of collectivist and individualist cultural differences.

Through the *State of Open Data Reports*, the levels of data sharing and usage have been monitored since 2016 in over 190 countries. The *2021 Report* goes beyond the usual metrics and includes new topics on "What motivates researchers to share data and the perceived discoverability and credibility of data shared openly". The key findings of this report are that (i) there is more concern about sharing datasets than ever before, (ii) there is more familiarity and compliance with the FAIR data principles than ever before, and (iii) repositories, publishers, and institutional libraries have a pivotal role to play in helping make data openly available.

In Africa, several researchers have explored the data sharing practices among researchers and have identified a number of barriers and opportunities. These barriers to data sharing in African research institutions are mainly associated with (i) the lack of policy and guideline frameworks at institutional and national levels, (ii) limited funding and (iii) inadequate infrastructures (Bezuidenhout, 2017; Bezuidenhout and Chakauya, 2018; Bangani and Moyo, 2019; Chiware, 2020; Abebe et al., 2021). The research in Africa and other developing and low-income environments on data sharing, has mainly focused on data sharing in public health and medical research.

Therefore, the purpose of this paper was to explore whether the attitudes of African researchers are different from those in other environments concerning data sharing and its practices. The paper analyzed data from *The State of Open Data Reports* (Hyndman, 2018) to answer these questions. More specifically, the analysis was centered around researchers' attitudes toward the following three areas: (i) the sharing of their own data, (ii) the shared data of others, and (iii) the Open Data ecosystem in place to enable wider data sharing. It was found that the attitudes of African researchers have changed over time, but they are not very different from their international counterparts.

Literature review

Africa provides a complex context within which the topic of Open Data has been explored and written up in the literature. This context can be viewed from an individual, policy, or

resource perspective. These contexts inform the researchers' attitudes toward sharing their own data, using shared data, and the wider Open Data ecosystem.

From the individual context perspective, an exploration of South African and Kenyan biochemistry researchers' data sharing perspectives revealed that individual perceptions of research environments were highly influential in shaping data sharing practices. Current low perceptions could be addressed through discussions on incentives and approaches that will improve the overall weak research environment (Bezuidenhout, 2017). In a study entitled "Narratives and counternarratives on Data Sharing in Africa", Abebe et al. (2021) argue that the many narratives emerging on data sharing among African scientists are somewhat distorting the full complexity of the African data sharing landscape where obstacles, issues, and challenges of data sharing on the continent are multifaceted. Anane-Sarpong et al. (2020) also believe that much more empirical research that engages stakeholders in data sharing in African research is still required to further and better understand the multitude of Open Science and Open Data challenges.

Understandably, there are serious concerns due to unclear institutional and national policy frameworks that currently guide research data sharing in many African countries, where the approaches are said to be in a piecemeal fashion and are further complicated by an uneven global *Open Data* sharing framework that is dictated by national and regional interests (Bezuidenhout, 2017). Stein (2020) points out that data sharing is an important aspect of science and is now required by most funding bodies and journals. Abebe et al. (2021) see entrenched and unbalanced historical power dynamics, trust issues, the need to better understand the African context, and the disregard of African generated research at play.

The national regulatory guidelines of many African nations do not explicitly allow for broad genetic data sharing, for example, and there is a need to reconsider these policies and propose creative solutions. In African genomic data sharing, Ramsay (2022) argues that despite a steady increase in data in international repositories, very little is coming from Africa and that the current analysis of genome data from Africa has yielded over 3 million novel and previously undocumented variants that could benefit the global community. Anane-Sarpong et al. (2020) also established that in the broader health sector, the impediments to data sharing include, among others, (i) risks faced by under-resourced researchers and their institutions, which have no capacity to quickly generate data produced into new knowledge, (ii) the lack of integrated guidelines and support mechanisms to address risk and reward researchers, and (iii) the general lack of confidence in existing protective safeguards.

From a resource perspective, Alter and Vardigan (2015) established that the many barriers to data sharing among researchers in low and middle-income countries are around (i) informed consent, (ii) data management, (iii) data dissemination, and (iv) the validation of research contributions. Ramsay (2022) also points to the challenges in the research ecosystems, including brain drain, lack of opportunities for young researchers, and limited resources. Denny et al. (2015) suggest that "for data sharing to be effective and sustainable, multiple social and ethical requirements need to be met and that an effective model of data sharing will be one in which considered judgments will need to be made about how best to achieve scientific progress, minimize risks of harm, promote fairness and reciprocity, and build and sustain trust".

The growth and development of data repositories are still limited in African institutions. However, those receiving international funding and who are publishing in international journals might be facing new requirements that ultimately require them to openly share data through different disciplinary and publishers' platforms. Bezuidenhout and Chakauya (2018) further elaborate on the "hidden concerns of sharing research data by low/middle-income country scientists" by pointing to the uneven landscape in which these scientists are equated when compared to their counterparts in developed environments that have far more developed and stable infrastructures. Despite these barriers, there is a growing interest among lower and middle-income scientists in data sharing (Bezuidenhout and Chakauya, 2018).

These contexts inform researchers' attitudes. Recent global surveys show that attitudes toward data sharing, data use, and data reuse are primarily positive, and that behavior does not always support these attitudes. Assistance through (i) Data Managers or Data Librarians, (ii) readily available data repositories for both long-term and short-term storage, and (iii) educational programs for both awareness and to help engender good data practices are needed (Tenopir et al., 2020). Bangani and Moyo (2019) observe that in South African universities, researchers preferred to use data produced by others and were not open to sharing their own data.

A study by Thoegersen and Borlund (2021) on researcher attitudes toward data sharing in public data repositories shows a need for greater clarity and consistency in using the term "*data sharing*" in future studies to better understand the use of this term, its phenomenon and to allow for cross-study comparison. In reviewing social scientists' data-sharing behaviors, Kim and Adler (2015) established that personal motivations and norms of data-sharing supported data sharing practices. However, institutional pressures by funding agencies, journals and data repositories required encouragement to facilitate social scientist data-sharing behaviors.

Methods

This article uses data collected for *The State of Open Data Report* in the 2017 and 2021 waves. The study is a longitudinal study, funded by the International Research Center, with the support of the Open Data for Development (OD4D) Network,

and conducted by Digital Science, SpringerNature and Figshare. The data for the various waves, including the associated questionnaires and reports, are freely available (Hyndman, 2018).

The survey covers various topics relating to Open Data, including attitudes and experiences of researchers toward data sharing across multiple research domains. Relevant to this article are the attitudes of the African researchers. Most questions are closed, categorical-type questions, limiting the statistical tests that could be performed. To conduct a comparative analysis, simple percentages were used and evaluated.

The sample is collated on the question, "Which country/territory are you located in" and, given the international fluidity of researchers, it is conceivable that those responding that they are located currently in Africa would not necessarily all originate from Africa. Among those who originate from Africa and are presently still in Africa, these researchers would likely have had international exposure through travel or collaboration with others. However, for ease of reference, the respondents who gave "Africa" as their continent of location will be termed "African researcher".

The first year that the *State of Open Data Report* was prepared was 2016. In that year, no responses were received from the African continent. The following year, 2017, saw 151 responses. Some of the questions asked in that year were repeated in 2021, providing a longitudinal insight over 5 years.

This study on whether African researchers view Open Data sharing differently from their non-African counterparts was framed within a three-layered framework: (i) attitudes to sharing one's own data, (ii) attitudes to the use of shared data, and (iii) attitudes to the broader sharing ecosystem. The framework can also be viewed within the overall research landscape in Africa as to whether the correct policies, incentives and infrastructures exist to enable positive attitudes toward the advancement of science on the continent and globally. The study on attitudes toward data sharing in Africa can also be viewed in the context of the conclusions from Baždarić et al. (2021) that this has to be framed within the broader understanding and appreciation of the principles and practices of Open Science.

Kim and Adler (2015) explored data-sharing behaviors on individual motivations, institutional pressures, and pressures using a combination of new institutional theory and the theory of planned behavior to develop a model that explains and predicts data sharing behavior. The suggested framework for this study is closely related to the new institutional theory and the theory of planned behavior. The exploration was on whether African researchers think differently about Open Data considering their personal positions and the conditions within their institutions.

Results

Research on data sharing practices globally and in Africa has been ongoing. Several outputs already provide details on how new approaches are being adopted or shunned globally. It is important to note what Abebe et al. (2021) terms narratives and counternarratives on African data sharing. To a large extent, recent literature on the subject has tended to ignore the African reality that is driven by (i) the lack of both developed research infrastructures and (ii) coordinated institutional and national policy frameworks on data sharing and the broader Open Science environments.

Given this wealth of information that is already available on the attitudes of African researchers toward Open Data, this article contributes by using new data to either confirm or refute existing findings.

Data comparison of attitudes between africa and the rest of the world

Attitudes to sharing one's own data

African researchers are no different from their counterparts in other regions when comparing their reported comfort levels relating to how other researchers might reuse their data. Replication, reanalysis, reinterpretation, isolated reuse, and combination reuse are globally accepted as expected forms of reusing data.

African researchers share much of the same challenges that other researchers experience. The uncertainty of sharing rights, ethical considerations and other permissions are of universal concern. African researchers are less concerned with the lack of time, size of datasets, organizing data and the risk of being 'scooped', while they are more unsure about which repository to use.

Motivations to share data are no different on the African continent than elsewhere. Public benefit and increased impact are given as strong circumstances that would motivate the sharing of data, which speaks to the altruistic attitudes of researchers. Additional key motivators are the recognitions, whether through full data citation, citation of research papers or co-authorships. Non-African researchers report considerably more frequently that the citation of research papers is a possible motivator (61% of non-African researchers, while only 29% of African researchers). Non-African researchers give the citation of research papers as the primary motivator for sharing data, while African researchers reflect that public benefit is what motivates them.

Attitudes to using Open Data

Researchers throughout the surveyed countries reported that shared data offers benefits: it (i) fosters collaboration,

(ii) validates findings, (iii) complements existing data and (iv) avoids the duplication of efforts. Whether these benefits are experienced, or just a perception of potential benefits is unclear. Regardless, the survey participants indicated that those who share their data do not receive sufficient recognition for their data. If recognition is a primary motivator, as reported in the previous section, then policies that follow those climates that foster recognition would also increase the sharing of research data.

A substantial majority of the respondents (79%) indicated that shared data added to the credibility of the research. As public benefit is a strong motivator, spreading the perception of shared data that is linked to research credibility would likely increase data sharing practices.

When participants were asked how they determined the quality of the shared data, many factors were considered and shown to be relevant, including (i) the reputation of the source of the data, (ii) the associated peer-reviewed article, and (iii) the availability of visualizations that are consistent with the data. Non-African researchers reported that clear dataset descriptions, which provide sufficient context, are a strong indicator of quality (84%), an opinion that was not as universally held by the African respondents: only 41% of them reported this to be the case. All researchers agreed that datasets that are easy to find are more likely to be viewed as credible.

Attitudes to the Open Data ecosystem

There is global agreement reported from the study participants that national mandates to make research data openly available to access, reuse, repurpose and redistribute would be welcomed since a mere 7% of the respondents disagreed. There is also support for funders to mandate data sharing as part of their grant awards, with 53% being in favor and 27% being against such a policy.

When asked whether making research articles open access should be common scholarly practice, the view was strongly affirmative: 93% of the African respondents and 87% of the non-African respondents agreed. The adoption of Open Data practices appears to have slightly less traction where 84% of the African respondents and 80% of the non-African respondents agreed.

Data comparison of African attitudes over time

The participants who responded in 2017 were from varied countries, with most responses originating from the more research-active countries: Egypt, Nigeria and South Africa. More than 60% came from universities, a distribution that is essentially unchanged in 2021.

Attitudes to sharing one's own data

In the 2017 questionnaire, fewer options were given to respondents in answer to the question of *what would motivate them*. Only a quarter of the respondents gave *being cited* as a motivator, while the *ease of sharing data* and *freedom of information request* was much stronger. It would seem that African researchers have become more aware of the factors motivating the sharing of data in the past 5 years, which speaks to a greater awareness of the Open Science movement. As in the 2021 data, respondents of the 2017 questionnaire gave *public benefit* as a strong motivator. Furthermore, *co-authorship credit* as a motivator has increased considerably over the past 5 years.

Attitudes to using Open Data

Only one question relating to the attitudes toward using Open Data was asked in 2017: *How do you think the data shared by others have or could benefit you?* The data comparison between the 2017 and 2021 responses shows that African researchers appear to have a greater awareness of the personal benefits of Open Data. The ambiguous phrasing of the question makes it unclear if the attitudes to the benefits resulted from a personal experience or simply a reflection of the current narrative among researchers.

Attitudes to the Open Data ecosystem

Support for a national mandate has increased considerably. In 2017, 33% of the respondents took a neutral position, and 54% supported the statement relating to national mandates for Open Data. In 2021, those taking the neutral position had dropped to only 6%, with an overwhelming 87% supporting a national mandate.

As the number of respondents doubled over that time, one can surmise that there has been an increased interest in issues relating to Open Data in the past 5 years. This increased awareness and interest is reflected in the other questions asked.

Discussion

This study focused on whether African researchers think differently about Open Data compared with their global peers, and whether the thinking of African researchers have changed over time. The findings reported here illustrate that, in general, research stakeholders are supportive of data sharing, with African researchers' practices and experiences not very different from their international counterparts, despite the policy differences. It is clearly highlighted that African researchers' attitudes toward data sharing have changed positively over the past decade. It is affirming to know that the impact that the global Open Science movement's message on the benefits of data sharing is spreading across the globe and to the

different research communities. The changing positive attitudes by African researchers toward data sharing could be attributed to the new requirements by journal publishers and research funders that data outputs must be visible to knowledge consumers.

The results further indicate that by 2021 African researchers were more supportive of their national governments' mandates to share data within and across research domains. There are growing efforts across the continent to formulate Open Science policy frameworks both at national and institutional levels. The United Nations Educational and Scientific Organisation has the *UNESCO Recommendation on Open Science*, a framework for member nations to develop their Open Science policy frameworks. Other donors and regional organizations like the Electronic Information for Libraries and the West and Central African Research and Education Network have also supported the development of national mandates that can assist researchers in participating in data sharing activities. In Southern Africa, especially South Africa, the Open Science environment has developed rapidly. South Africa has also moved a step further through a partnership with the European Union's *Europe Open Science Cloud* to develop its own South African Open Science Cloud framework, further supporting the about-to-be finalized South African Open Science Framework.

This study also highlights several issues that require attention from African academic and research institutions to support researchers' data sharing practices that will enable best science advancement and societal engagement. There should be clear national and institutional policy frameworks to enable good data sharing practices to take root among African researchers. The position of international research funders, journal publishers, and inter-institutional and country collaborations need to be spelt out in future policies to ensure equitable data custodianship in African generated research. Given the attitudes of African researchers toward such policies, the time is right to put them into place.

The African academic and research communities should not be left out and lose focus of the potential benefits of Open Data for reproducibility and efficiency in research, especially in poorly resourced environments that require more collaborative use of infrastructure and resources. The potential gains for further and faster benefits in advancing science and knowledge production are all too evident in coordinated data sharing activities.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: Figshare repository https://doi.org/10.6084/m9.figshare.c.4046897.v6.

Author contributions

LS took the lead in the conceptualization, methods, results, and final editing. EC took the lead in introduction, literature review, and conclusion. Both authors contributed to the article and approved the submitted version.

Funding

This work has been partially funded by Stellenbosch Business School.

Acknowledgments

Thanks to Patrick Horrigan and Linda Kelly for assistance with editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abebe, R., Aruleba, K., Birhane, A., Kingsley, S., Obaido, G., Remy, S. L., et al. (2021). "Narratives and counternarratives on data sharing in Africa," in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM), 329–341. doi: 10.1145/3442188.3445897

Alter, G. C., and Vardigan, M. (2015). Addressing global data sharing challenges. *J. Empir. Res. Hum. Res. Ethics* 10, 317–323. doi: 10.1177/1556264615591561

Anane-Sarpong, E., Wangmo, T., and Tanner, M. (2020). Ethical principles for promoting health research data sharing with sub-Saharan Africa. *Dev. World Bioeth.* 20, 86–95. doi: 10.1111/dewb.12233

Bangani, S., and Moyo, M. (2019). Data sharing practices among researchers at South African Universities. *Data Sci. J.* 18, 28. doi: 10.5334/dsj-2019-028

Baždarić, K., Vrkić, I., Arh, E., Mavrinac, M., Gligora Marković, M., Bilić-Zulle, L., et al. (2021). Attitudes and practices of open data, preprinting, and peer-review—A cross sectional study on Croatian scientists. *PLoS ONE* 16, e0244529. doi: 10.1371/journal.pone.0244529

Bezuidenhout, L. (2017). Technology transfer and true transformation: implications for open data. *Data Sci. J*. 16, 26. doi: 10.5334/dsj-2017-026

Bezuidenhout, L., and Chakauya, E. (2018). Hidden concerns of sharing research data by low/middle-income country scientists. *Glob. Bioeth*. 29, 39–54. doi: 10.1080/11287462.2018.1441780

Bird, C. L., and Frey, J. G. (2013). Chemical information matters: an e-Research perspective on information and data sharing in the chemical sciences. *Chem. Soc. Rev*. 42, 6754. doi: 10.1039/c3cs60050e

Chiware, E. R. T. (2020). Open research data in African academic and research libraries: a literature analysis. *Libr. Manag*. 41, 383–399. doi: 10.1108/LM-02-2020-0027

Denny, S. G., Silaigwana, B., Wassenaar, D., Bull, S., and Parker, M. (2015). Developing ethical practices for public health research data sharing in South Africa. *J. Empir. Res. Hum. Res. Ethics* 10, 290–301. doi: 10.1177/1556264615592386

Hyndman, A. (2018). *State of Open Data. Figshare. Collection [Internet]*. Available from: https://doi.org/10.6084/m9.figshare.c.4046897.v6 (accessed July 1, 2022).

Kim, Y., and Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *Int. J. Inf. Manage*. 35, 408–418. doi: 10.1016/j.ijinfomgt.2015.04.007

Ramsay, M. (2022). African genomic data sharing and the struggle for equitable benefit. *Patterns* 3, 100412. doi: 10.1016/j.patter.2021.100412

Stein, C. M. (2020). Challenges of genetic data sharing in African studies. *Trends Genet*. 36, 895–896. doi: 10.1016/j.tig.2020.07.010

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE* 6, e21101. doi: 10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., et al. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE* 10, e0134826. doi: 10.1371/journal.pone.0134826

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., et al. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE* 15, e0229003. doi: 10.1371/journal.pone.0229003

The Royal Society (2012). *Science as an Open Enterprise: The Royal Society Science Policy Centre Report*. London.

Thoegersen, J. L., and Borlund, P. (2021). Researcher attitudes toward data sharing in public data repositories: a meta-evaluation of studies on researcher data sharing. *J. Doc*. 78, 1–17. doi: 10.1108/JD-01-2021-0015

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J,. Appleton, G., Axton, M., Baak, A, et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3, 160018. doi: 10.1038/sdata.2016.18

frontiers | Frontiers in Research Metrics and Analytics

ORIGINAL RESEARCH
published: 27 July 2022
doi: 10.3389/frma.2022.912456

Rethinking the A in FAIR Data: Issues of Data Access and Accessibility in Research

Hugh Shanahan[1]* *and Louise Bezuidenhout*[2]

[1] Department of Computer Science, Royal Holloway University of London, London, United Kingdom, [2] DANS (Data Archiving and Networked Services), The Hague, Netherlands

The FAIR data principles are rapidly becoming a standard through which to assess responsible and reproducible research. In contrast to the requirements associated with the Interoperability principle, the requirements associated with the Accessibility principle are often assumed to be relatively straightforward to implement. Indeed, a variety of different tools assessing FAIR rely on the data being deposited in a trustworthy digital repository. In this paper we note that there is an implicit assumption that access to a repository is independent of where the user is geographically located. Using a virtual personal network (VPN) service we find that access to a set of web sites that underpin Open Science is variable from a set of 14 countries; either through connectivity issues (i.e., connections to download HTML being dropped) or through direct blocking (i.e., web servers sending 403 error codes). Many of the countries included in this study are already marginalized from Open Science discussions due to political issues or infrastructural challenges. This study clearly indicates that access to FAIR data resources is influenced by a range of geo-political factors. Given the volatile nature of politics and the slow pace of infrastructural investment, this is likely to continue to be an issue and indeed may grow. We propose that it is essential for discussions and implementations of FAIR to include awareness of these issues of accessibility. Without this awareness, the expansion of FAIR data may unintentionally reinforce current access inequities and research inequalities around the globe.

Keywords: FAIR, Open Science, data, low/middle income countries, accessibility

OPEN ACCESS

Edited by:
Rania Mohamed Hassan Baleela, University of Khartoum, Sudan

Reviewed by:
Gerit Pfuhl, UiT the Arctic University of Norway, Norway
Keith Russell, Monash University, Australia
Markus Stocker, Technische Informationsbibliothek (TIB), Germany

***Correspondence:**
Hugh Shanahan
hugh.shanahan@rhul.ac.uk

Specialty section:
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

Received: *05 April 2022*
Accepted: *07 June 2022*
Published: *27 July 2022*

Citation:
Shanahan H and Bezuidenhout L (2022) Rethinking the A in FAIR Data: Issues of Data Access and Accessibility in Research. Front. Res. Metr. Anal. 7:912456. doi: 10.3389/frma.2022.912456

FAIR DATA: THE NEW CORNERSTONE OF RESPONSIBLE RESEARCH

Since their conception in January 2016, the FAIR data principles (Wilkinson et al., 2016) have rapidly gained traction and widespread global acceptance. The FAIR data principles were first published under FORCE11[1] and advocate for the Findability, Accessibility, Interoperability and Reusability of research data and scholarly digital objects more generally. FAIR consists of 15 requirements grouped under the four categories. These requirements serve to guide the actions of data publishers, stewards and other stakeholders to enable responsible data sharing. Central to

[1]https://www.force11.org/group/fairgroup/fairprinciples (accessed April 4, 2021).

the concept of FAIR is its application "to both human-driven and machine-driven activities," with a goal of machine-actionability to the highest degree possible or appropriate. The widespread uptake of the FAIR principles has led to a plethora of diverse activities, including infrastructure development, disciplinary standard setting and ontology creation, and capacity building in data stewardship (Gaiarin et al., 2021). There has been very recent further development in other principles such as the TRUST principles on how repositories should be run (Lin et al., 2020). An analysis of the other principles is beyond the scope of this paper.

The FAIR data standards are an important element of the Open Research ecosystem. Indeed, Open Data, FAIR, and research data management (RDM) are three overlapping but distinct concepts, each emphasizing different aspects of handling and sharing research data (Higman et al., 2019). Higman et al. (2019, p. 1) clarify this relationship in the following way: "FAIR and open both focus on data sharing, ensuring content is made available in ways that promote access and reuse. Data management by contrast is about the stewardship of data from the point of conception onwards. It makes no assumptions about access, but is essential if data are to be meaningful to others."

Within Open Research, FAIR, Open, and RDM are central not only to practical discussions on infrastructure evolution, but also underpin motivational and aspirational discourse. The ethical drivers of equitable access, transparency as well as the elimination of financial barriers to research outputs play an important role in the evolving aspiration of a "global knowledge commons." This concept, first introduced by Hess and Ostrom, refers to information, data, and content that is collectively owned and managed by a community of users, particularly over the Internet. Key to the structure of the commons is shared access to digital resources (Hess and Ostrom, 2006, 2007), which emphasizes the reusability—and thus the FAIRness—of data.

While the FAIR data principles have gained rapid acceptance and support, the processes, practices, technical implementation and infrastructures necessary to make data FAIR continue to evolve. It is recognized that realizing a FAIR ecosystem will involve developing key data services that are needed to support FAIR. These include "services that provide persistent identifiers, metadata specifications, stewardship and repositories, actionable policies and Data Management Plans. Registries are needed to catalog the different services" (Collins et al., 2018, p. 8). The challenges of embedding FAIR data practices within research thus include both the technical challenges of creating FAIR-enabling data infrastructures and the need for education and capacity building within research communities.

Accessibility as a FAIR Principle

The FAIR accessibility principle can be understood as requiring that data are stored properly—for long term—so that it can easily be accessed and/or downloaded with well-defined access conditions. At a minimum, this principle requires access to the metadata. The principle makes a number of requirements of the metadata that accompanies data, including that (A1) (meta)data are retrievable by their identifier using a standardized communications protocol. This includes that (A1.1) the protocol is open, free and universally implementable and that (A1.2) the protocol allows for an authentication and authorization procedure where necessary. It also requires that (A2) metadata are accessible, even when the data are no longer available. In practice this requires that the metadata accompanying the data be understandable to humans and machines, are registered or indexed in a searchable resource and are deposited in a trusted repository (Wilkinson et al., 2016).

As can be seen from the requirements, the FAIR accessibility requirements are highly dependent on the availability of trusted digital repositories and FAIR-oriented curation processes. At the moment the international repository landscape is rapidly evolving, and considerable efforts are being made to promote certification processes to promote FAIR data practices. Indeed, a recent collaboration between the FAIRsFAIR research consortium[2] and CoreTrustSeal[3] has worked to integrate FAIR-enabling assessment into the CoreTrustSeal certification of repositories. Integral to this work is the recognition that: "the FAIR Principles are clarified through indicators and evaluated through (ideally automated) tests against digital objects"[4].

In response to the recognized need for more automated tests for FAIR, a number of data assessment methods and tools have been developed to assign "FAIR scores" to datasets based on a number of criteria. These include automated tools such as F-UJI[5], FAIR-Enough[6] and FAIR-Checker[7], as well as manual and educational tools such as the ARDC FAIR self-assessment tool[8], and FAIR Aware[9].

The use of these different assessment tools offers researchers an opportunity to test the FAIR-ness of their data. The scores returned by these tools do vary, according to the criteria included in their design. Nonetheless, regardless of the tool the assessments of accessibility are largely interlinked to the existence of repository and curation infrastructures. For instance, F-UJI scores the accessibility principle on three criteria, namely:

A1-01: metadata contains access level and access conditions of the data
A1-02: metadata is accessible through a standardized communication protocol
A1-03: data is accessible through a standardized communication protocol

When these accessibility requirements are scrutinized, however, it becomes apparent that the FAIR scores returned for any database aim to provide an objective view of access. Within

[2]https://fairsfair.eu/ (accessed April 4, 2021).
[3]https://www.coretrustseal.org/ (accessed April 4, 2021).
[4]https://www.coretrustseal.org/why-certification/coretrustsealfair-statement-of-cooperation-support/ (accessed April 4, 2021).
[5]www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool (accessed April 4, 2021).
[6]www.fair-enough.semanticscience.org (accessed April 4, 2021).
[7]www.fair-checker.france-bioinformatique.fr/base_metrics (accessed April 4, 2021).
[8]www.adrc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/ (accessed April 4, 2021).
[9]www.fairsfair.eu/fair-aware (accessed April 4, 2021).

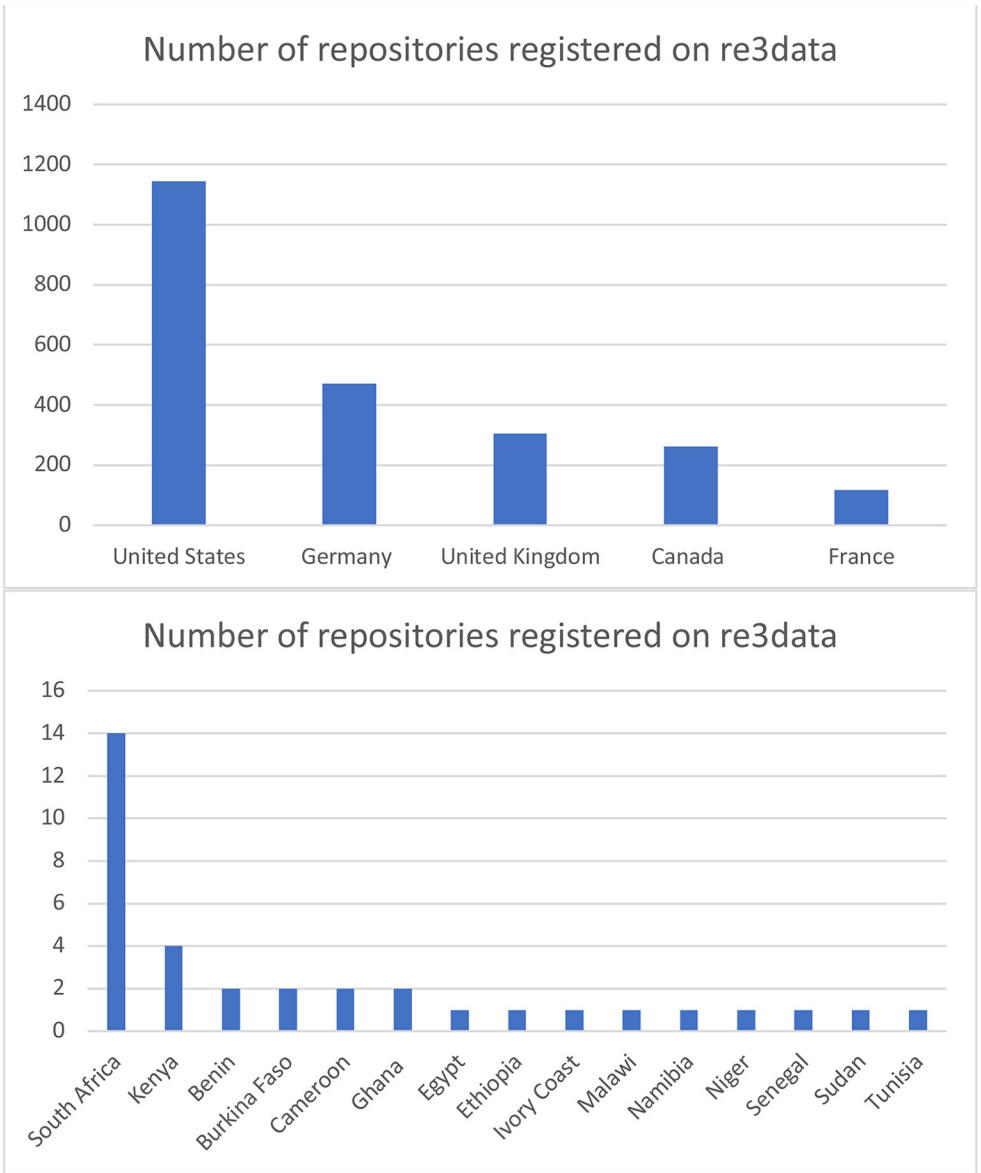


FIGURE 1 | Distribution of repositories within Re3Data according to geographic location in 2022. Available online at: https://www.re3data.org/browse/by-country/ (accessed April 4, 2021).

this aim, however, there is an implicit assumption that access is considered solely in relation to the structure of the metadata or data and is independent of the user attempting to access those resources. As a result, the scores cannot be taken to measure the actual accessibility of data or metadata from a *user* perspective. This observation is linked to the realization that depending on the geographic location of the user request the availability of the metadata/data may vary considerably which raises considerable questions. Most pertinent, it becomes important to question whether assigning a FAIR accessibility score to metadata/data could create a false sense of access that undermines existing discussions about inequity in Open Science (Bezuidenhout et al., 2017b; Ross-Hellauer et al., 2022).

ACCESSIBILITY AND INFRASTRUCTURES

The European Commission Open Science Monitor tracks trends for open access, collaborative and transparent research across countries and disciplines[10]. The most recent version included a breakdown of the geographic location of the trusted data repositories included in the re3data catalog[11]. As is evident in **Figure 1** below, a high number of the data repositories (2,299) registered on re3data reside in just five countries: USA, Germany,

[10]https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor_en (accessed April 4, 2021).
[11]www.re3data.org (accessed April 4, 2021).

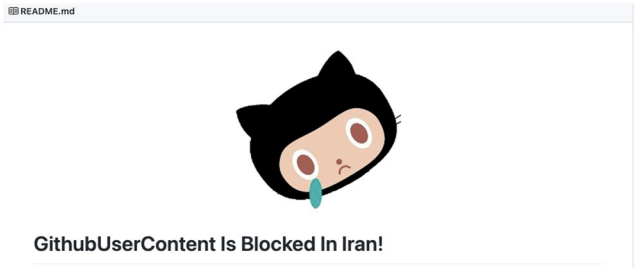


FIGURE 2 | User message returned from a GitHub access request by Iranian users in 2020. Available online at: https://github.com/pi0/github-is-blocked-in-iran (accessed April 4, 2021).

UK, Canada, and France. Similarly, many other high-income countries (HICs) host multiple repositories.

In contrast, the whole of the African continent has 35 repositories registered on re3data. Aside from Kenya (4) and South Africa (14), all other countries host either one or two[12]. This unequal global distribution of repositories contributes to the accessibility concerns outlined in the section above. These concerns group around two key issues, namely geopolitical and infrastructural access problems. These concerns are discussed in more detail below.

Geoblocking and Access Restrictions

Geoblocking is a term used to describe the intentional blocking or restriction of access to websites, apps or other internet content depending on the geographic location of the users. Geoblocking is commonly used in commercial applications to segment customers geographically, and often goes largely unnoticed within the general research community. Indeed, the 2018 ban on geoblocking between Member States of the European Union has even further lowered the visibility of this topic[13].

In recent years, however, a small number of academic studies have drawn attention to the impact of geoblocking practices on research (Bezuidenhout et al., 2019; Bezuidenhout and Havemann, 2021). A key observation from these studies is that the Open Science ecosystem is increasingly being populated by diverse actors and many commercial companies are offering key services to the research community. As commercial companies, these actors are subject to the financial legislation of the country in which they are registered. For commercial companies registered in the USA, for instance, this means that they are prohibited from transacting with customers/users residing within countries against which the USA holds financial sanctions. As evidenced in **Figure 2** below, the US financial sanctions in place against Iran means that Iranian researchers were unable to access GitHub, a key Open Science tool until 2021[14]. Researchers in Syria and Crimea continue to experience access blocks to GitHub.

Geoblocking on web sites bars access to web sites on the basis of the country a user is sending a request for a web page (identified with a Uniform Resource Location, URL) through their web browser. It is important to note that even when a web site blocks a user both the browser and web site will for the most part exchange some data (unless a web site is not responsive if it is down or because there are connectivity issues). Web standards for geolocation exist[15] but they are based on the user providing additional data such as their longitude and latitude. A simpler method for identifying a user's country is through the Internet Protocol (IP) address the browser is sending the request from. Web server software, such as Apache HTTP server,[16] allow web developers to control access to either a particular directory or whole site with a suitably configured file to block access through this IP based approach. In this case, a web browser will receive from the web site a specific error code, namely a 403 code[17]. As repositories will have limited resources, it is unlikely that they will develop more sophisticated approaches to geoblock users.

Similarly, the difficulties of conducting financial transactions from countries under sanctions makes it extremely difficult for researchers within these sanctioned countries to engage with key research activities. These include publishing in academic journals, paying membership fees to academic societies or membership-dependent resources, and buying key software/hardware to refine their datasets (Adam, 2019; Bezuidenhout et al., 2019). These issues of access also extend other key data repositories and collections. Through discussions within dual-use and biosecurity communities it is certain that access to datasets and data tools can be restricted according to the geographic location of the users. This includes, for example, reports of the USA blocking Chinese supercomputer groups[18], and the USA restricting access to climate change data due to security concerns[19].

In addition to sanctions-related geoblocking, there is also an increasing trend for national governments to use access restrictions as a means of political control. As demonstrated in **Figure 3** below, a number of African countries have recently experienced limitations on freedom of press and access to information[20]. Similar limits have been reported from other countries across the globe[21]. While not directed at academia, it is evident that these shutdowns can have a significant impact on research within these countries.

[12]https://www.re3data.org/browse/by-country/ (accessed April 4, 2021).
[13]https://digital-strategy.ec.europa.eu/en/news/geo-blocking-regulation-questions-and-answers (accessed April 4, 2021).
[14]https://github.blog/2021-01-05-advancing-developer-freedom-github-is-fully-available-in-iran/ (accessed April 4, 2021).
[15]https://www.w3.org/TR/geolocation/ (accessed April 4, 2021).
[16]https://httpd.apache.org (accessed April 4, 2021).
[17]https://developer.mozilla.org/en-US/docs/Web/HTTP/Status#client_error_responses (accessed April 4, 2021).
[18]https://www.bbc.com/news/business-56685136 (accessed April 4, 2021).
[19]https://www.stripes.com/news/us/navy-pulls-plug-on-climate-task-force-after-pentagon-deems-climate-change-a-national-security-issue-1.596184#.XWYgtXfMVhg.twitter (accessed April 4, 2021).
[20]https://blogs.lse.ac.uk/africaatlse/2020/05/28/popular-support-for-media-freedom-press-africa-complicated-picture/ and https://www.bbc.com/news/world-africa-59958417 (accessed April 4, 2021).
[21]https://twitter.com/Meenwhile/status/1265711539140440064 and https://netblocks.org/ and https://www.bbc.com/news/world-asia-55923486 (accessed April 4, 2021).

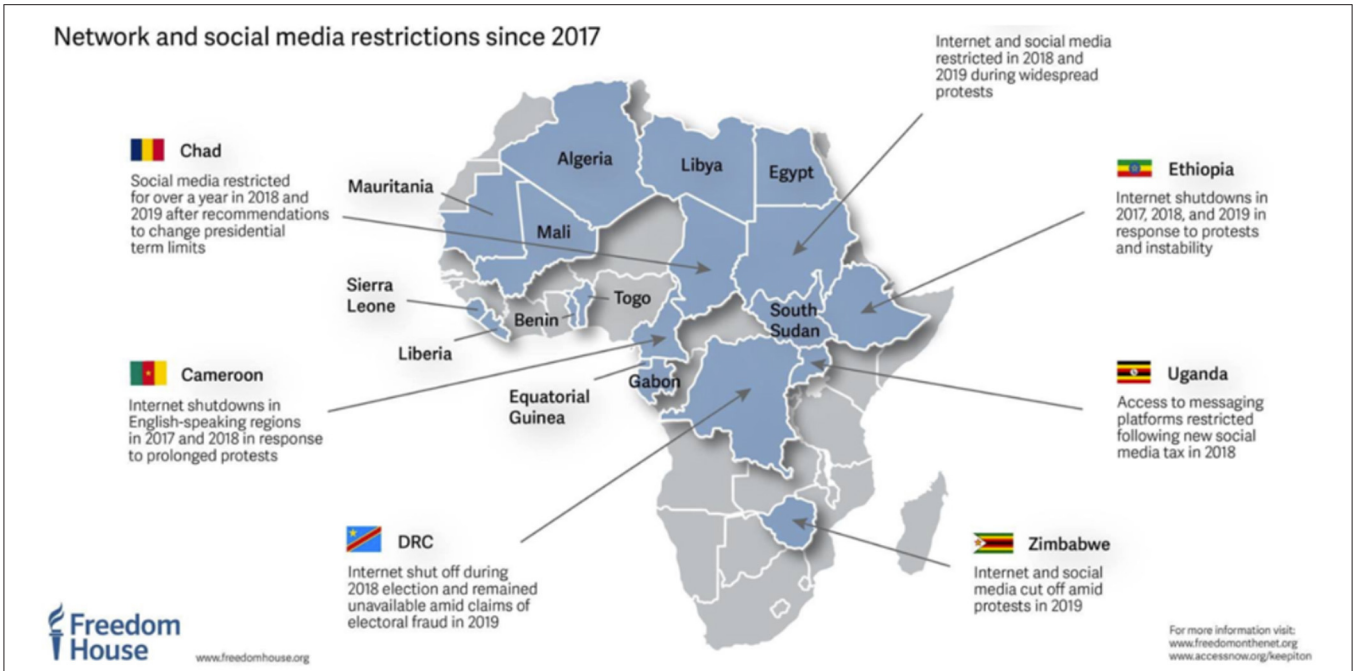


FIGURE 3 | Internet restrictions within Africa between 2017 and 2019 from Freedom House data Available online at: https://twitter.com/freedomhouse/status/1138135355931201536/photo/1 (accessed April 4, 2021). .

Time-Outs and Last-Mile Connection Issues

While the challenges of overt access restriction are becoming increasingly visible, there are a range of other access issues that are widespread, pernicious and regularly overlooked. These relate to poor connectivity that hampers researchers across the globe. These "last mile" challenges refer to the inaccessibility of online data due to a range of issues including low bandwidth, unstable connectivity, power outages and the cost of data (Bezuidenhout et al., 2017a).

Unstable connectivity can mean that when a web browser makes a request for a web page a sufficient period of time will pass where the browser does not get a response and hence creates a time-out error. This in particular can affect large data downloads or downloads of multiple smaller files.

The work-from-home requirements of the COVID-19 pandemic presented additional challenges to researchers working in many low/middle-income countries (LMICs) due to data transmission costs. For many researchers, landline or fibreoptic connectivity is not possible in their home context, meaning that they relied on mobile data for connectivity. A recent study on mobile data costs demonstrated that the three countries with the most expensive mobile data per 1GB are all in Africa. These were Malawi ($27.41), Benin ($27.22), and Chad ($23.33)[22]. These data costs, when put into perspective with the average salaries of researchers and postgraduate students in these countries, makes accessing datasets or engaging with online collaborations prohibitively expensive.

QUANTIFYING THE ABSENCE OF ACCESSIBILITY FROM A USER PERSPECTIVE

While the introduction presents a range of concerns relating to the accessibility of data, it is difficult to advocate for action on these issues without quantifiable data. To date, much of the evidence presented in support of these concerns relies on small qualitative studies or anecdotal evidence. In order to address the paucity of data, the authors set out to quantify the level of difficulty associated with getting access this paper simulates access from a range of countries (high, middle and low income; some under sanction and others not) by using web proxies of a set of web sites that are key for Open Science.

Rather than focusing on the content of the pages downloaded, this study set out to test whether sites were downloaded at all and what error codes were returned if there was a failure to download. The results thus do not distinguish between access time-out or blocking. The results were recorded and compared with the results from the other countries in the study.

Methodology

The analysis is based on two steps. In the first instance a set of web sites were selected that are used in Open Science. Once that list was collated, proxies were set up for a set of countries to examine

[22]https://www.visualcapitalist.com/cost-of-mobile-data-worldwide/ (accessed April 4, 2021).

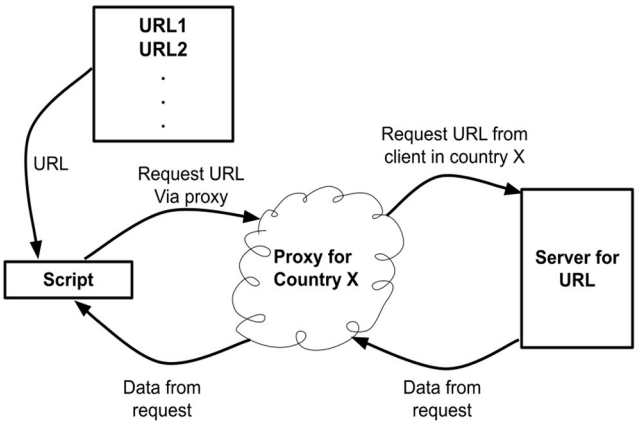


FIGURE 4 | Schematic of software used to download data.

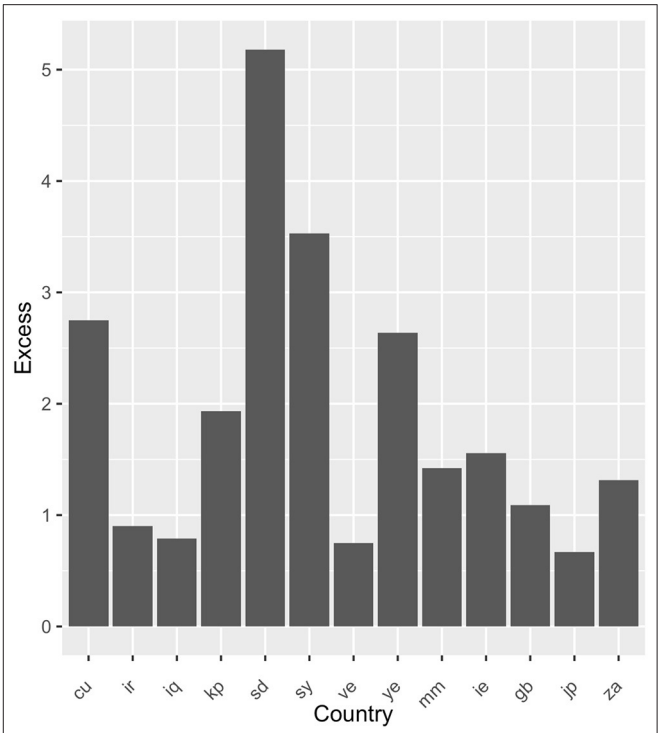


FIGURE 5 | Ratio of number of URLs that did not return HTML in a country to the number of URLs that did not return HTML for us.

access to those sites from the set of countries[23]. All of the software developed for this project and accompanying data can be found on the repository Zenodo (Shanahan and Bezuidenhout, 2022).

Selection of Suitable Web Sites

Two sets of web sites were collated. In the first instance a curated set of 254 web sites from 101 tools + JISC list of open science tools (Bezuidenhout and Havemann, 2021). This lists key sites such as github.com, bioarxiv.com and osf.io. A second set of web sites was collated from Re3Data which is a registry of research data repositories. A script was run to download all the web sites listed in re3data in June 2020. The URLs and Re3Data IDs for 2527 sites were downloaded using this approach. Hence 2,781 URLs were collated for this study.

Proxies

Fourteen countries were selected to download the above total list of URLs. These countries are Cuba (cu), the United Kingdom (gb), Ireland (ie), Iraq (iq), Iran (ir), Japan (jp), North Korea (kp), Myanmar (mm), Sudan (sd), Syria (sy), the United States of America (us), Venezuela (ve), Yemen (ye), and South Africa (za). Their corresponding ISO 3166 standard two letter code[24] are listed in parenthesis and in this paper these codes will be used.

The proxy service provided by the company Bright Data[25] (previously referred to as Luminati) was used to provide clients in each of the above countries. A schematic of the software can be found in **Figure 4**. Using the API of Bright Data for each country a request was made to download the URLs in the above list. The User-Agent string for the HTTP request was set to correspond to the most up to date chrome browser[26]. Download data was captured and stored in individual JSON files. This data includes the HTTP response status codes, the HTML downloaded if access was successful and error codes if access was unsuccessful. The data was gathered in August 2020.

RESULTS

Two types of data collected during the study are presented below. The first attempts to identify the effect of data not being returned because of connectivity issues. In particular we contrast sites where access is unsuccessful in specific countries with the same sites being successful in downloading in other countries. This can be tracked with a timeout error code in the former and a return code of 200 in the latter (which indicates a successful download)[27]. The second looks for potential cases where sites are blocking access for users in specific countries and allowing access for others. In this case one tracks cases which return a code of 403 (described previously) and a 200 code.

Access to Sites Is Variable Across Countries

For each country the number of sites that did not return any HTML was noted. It is noted that 284 sites consistently could not be downloaded from any country. For each pair of countries (c1,c2c2) the number of sites that failed to return HTML when downloading from c1 but did return HTML when downloaded from c2 was computed. This is referred to as N(c1,c2c2).

[23]The methodology was based on a paper by McDonald et al. that focused on geoblocking (McDonald et al., 2018).

[24]https://www.iso.org/iso-3166-country-codes.html (accessed April 4, 2021).

[25]https://brightdata.com (accessed April 4, 2021).

[26]Chrome/84.0.4147.89 (accessed April 4, 2021).

[27]https://developer.mozilla.org/en-US/docs/Web/HTTP/Status#successful_responses (accessed April 4, 2021).

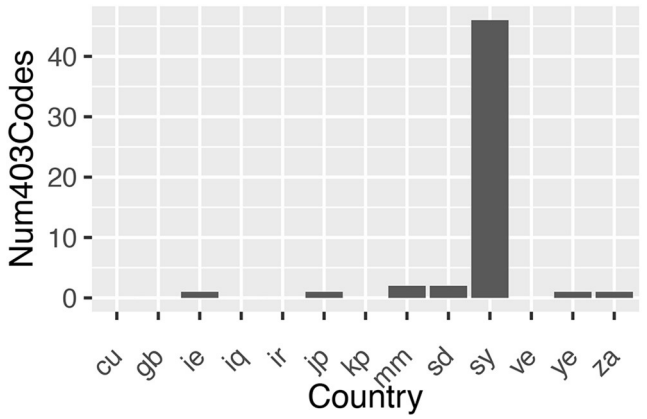


FIGURE 6 | Number of URLs that returned a 403 HTTP response status code when the same URL returned a 200 HTTP response status code from US.

Assuming that the us will have near-universal access **Figure 5** plots the ratio N(c1,c2c2)/N(c2,c1c1) where c1 are the other countries in the list and c2 = us. If two countries are able to access precisely the same sites then the ratio should be one. If c1 can access more sites than c2 then the ratio is <1 and correspondingly if c2 can access more sites than c1 then the ratio should be >1. The results are summarized in **Figure 5**.

From the above results we find that countries have a variable level of access to the URLs. In particular, Cuba, Sudan, Syria, and Yemen (cu, sd, sy, and ye) are much more likely to be unable to download the URLs. Countries such as Ireland, Japan. and the United Kingdom (ie, jp, and gb) give a range of excess values indicating that the spread of values for Iran, Iraq, Venezuela, and Myanmar (ir, iq, ve, and mm) are not significant. This corresponds to cases where connectivity is poor.

Specific Blocking of Countries Appears to Exist

If a web server understands a request to access a URL from a client but refuses to authorize it then it returns a HTTP response status code of 403, as opposed to a response code of 200 if the request is successful. Using US again as a control, for each country the set of URLs which return a 403 response code for that country and returns a 200 response code for US were collated. The number found for each country are plotted in **Figure 6**. The URLs these correspond to are listed in the **Appendix**. As evidenced in **Figure 6**, the significant increase in 403 status codes for Syria (sy) suggests possible geoblocking.

DISCUSSION

The approach of using proxies to test the access to data is a useful tool for exploring accessibility from a FAIR perspective. Specific geoblocking of sites is harder to detect as some sites may be directly blocked (posting a 403 code) but others, such as Github, may list a web page but may return different content indicating that access to the site is blocked. These preliminary data clearly demonstrate that more research is urgently needed in order to problematize this issue and provide data to inform future Open Science policies. Nonetheless, even this preliminary data raises important issues relating to the accessibility described and defined through the FAIR data principles. These include the observations that:

1. Accessible in-country doesn't mean accessible in all countries
2. User-experiences of FAIR data may vary considerably—as may scores when testing from places that return 403s
3. Discussions on FAIR accessibility cannot be de-coupled from broader discussions on access to Open Data.

Accessible In-country Doesn't Mean Accessible in All Countries

Figure 1 above illustrates the geographic distribution of trusted digital repositories. It clearly demonstrates that there is a significant bias toward repositories located in HICs. This bias is unsurprising, as HICs continue to dominate global research and development (R&D) expenditure (69.3% in 2013) as well as host the majority of researchers[28]. Because the majority of the repositories, as well as the bulk of their users, are located in HICs, it is possible that this has implications for the "FAIRness" of their design.

The criteria associated with the FAIR accessibility principle means that it is possible for data to be considered accessible without all researchers being able to query data/metadata from their geographic locations. This means that the identifier used to query the database does not return the appropriate data/metadata. It is important to note that this lack of return is likely not related to the standardized communications protocols in place, but rather due to additional barriers in place at various points in the data journey. This draws attention to the possibility that discussions of data accessibility need to be expanded beyond metadata and query protocols to consider a broader range of barriers embedded within the digital landscape.

When considering an expanded discussion around data accessibility it is important to note that there are likely no "quick fixes." The use of VPNs has been suggested as a tool for bypassing geopolitical barriers to data, as a means of virtually locating the user request in a different country. However, advocating for the use of VPNs as a means of integrating into the current data landscape must raise concerns. Some countries with repressive governments have outlawed VPNs as a means of maintaining control over information flows[29]. Furthermore, national governments have also been reported to engage in VPN blocking as a means of censorship and control[30]. Requiring researchers to use VPNs as a means of engaging with the current Open Science infrastructure can thus place them in positions of personal risk and can thus not be viewed as a viable alternative to the current problem. VPNs will also not fix overall connectivity issues.

[28]https://en.unesco.org/node/252279 (accessed April 4, 2021).
[29]https://protonvpn.com/blog/are-vpns-illegal/ (accessed April 4, 2021).
[30]https://en.wikipedia.org/wiki/VPN_blocking (accessed April 4, 2021).

User-Experiences of FAIR Data May Vary Considerably

Even when there are no barriers to accessing the data stored in trusted digital repositories, the evidence presented in this paper suggest that user experiences of interacting with FAIR data may vary considerably around the world. Understanding FAIR from a user perspective is important not only as a means of improving downstream service provision, but also as a means of community engagement. The success of the FAIR principles is contingent on the engagement of researcher communities, and their subsequent adoption of FAIR research practices. If some user communities continue to struggle to access and re-use FAIR data it is possible that this may affect the levels of community engagement and support. Such concerns follow on from similar observations from studies on support for Open Data practices (Bezuidenhout et al., 2017b)[31]. This lack of engagement could lead to a lag in the adoption of FAIR data practices and exacerbate the existing under-representation of certain user communities within the FAIR landscape.

Discussions on FAIR Accessibility Cannot Be De-coupled From Broader Discussions on Access to Open Data

A central element of current data discussions is the statement that while not all data can be open (e.g., some research data, such as medical data, needs to remain private, and access-controlled), all data must be FAIR. This coupling of Open and FAIR has been used by governments, funders, and institutions to strengthen their commitment to Open Science. As highlighted by Higman, Bangert, and Jones FAIR principles "are being applied in various contexts; the European Commission has put the FAIR principles at the heart of their research data pilot alongside open data. Beyond Europe, the American Geophysical Union (AGU) has a project on Enabling FAIR Data and the Australian Research Data Commons (ARDC) supports a FAIR programme" (Higman et al., 2019, p. 1). Funded researchers are increasingly expected to ensure that the data produced in their research are FAIR, regardless of whether it will be Open.

Within Open Data/Open Science discussions there is a growing recognition that the so-called "digital divide" continues to slow down the evolution of the global research ecosystem. Indeed, infrastructural challenges are regularly mentioned in relation to Open Science in LMICs (CODATA Coordinated Expert Group, 2020) and highlight the need for large-scale infrastructural investment. In contrast, however, similar discussions about local infrastructure are not a priority in FAIR discussions. Not addressing the impact of infrastructures on FAIR-ifying data has a number of consequences. It either suggests that making data accessible is not influenced by the infrastructure available to researchers, or provides the impression that nothing can be done *at the moment* by individual researchers until the research infrastructures evolve.

[31] https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en#funderspolicies (accessed April 4, 2021).

FINAL COMMENTS

It is recognized that data FAIRness is a "moving target" and as infrastructure, practices and processes continue to develop so too will the requirements of what is regarded as being sufficiently FAIR. This awareness reflects the nature of the FAIR principles, namely that they are aspirational (i.e., they are not a set of well-defined technical standards) and do not strictly define how to achieve a state of "FAIRness." As described by Wilkinson and colleagues, the FAIR data principles "describe a continuum of features, attributes, and behaviors that will move a digital resource closer to that goal" (Wilkinson et al., 2018, p. 1). This ambiguity, they suggest, has led to a wide range of interpretations of what constitutes a FAIR resource[32].

The ambiguity of what FAIRness constitutes can be thought of on many different levels, but underpins the non-absoluteness of the concept. This paper advocates for the further discussion on how the FAIR principles are translated into action. In contrast to current discussions that focus on the interpretation of the FAIR principles from a disciplinary perspective, this paper emphasizes the urgent need for discussions on the variability introduced by *geographic* and *geo-political factors*. In particular, the paper advocates for a critical reflection on the "frames of reference" used as a basis for discussions on what constitutes "as FAIR as possible for the present." The use of the accessibility principle to illustrate these points is important, as findability and accessibility are widely considered to be the "easier" of the FAIR principles to achieve.

A brief survey of the current geo-political climate around the world suggests that issues relating to accessibility that are raised in this paper are poised to get worse if nothing is done. The current war in Ukraine and the proposed sanctions on Russia by NATO nations suggest that issues of geoblocking might be exacerbated going forward[33]. Issues of access and time-outs are becoming more frequent due to a growing trend of using internet access to control civil unrest. Moreover, the investment in information and communication technologies in LMICs, while growing, will continue to present challenges for decades to come. Researchers in these regions are unlikely to experience a "level playing field" of connectivity with their HIC colleagues for decades.

Bringing these often-overlooked issues together highlights how current FAIR discussions on data accessibility often fail to recognize pressing challenges experienced by many researchers around the world. To date, there has been little recognition of these issues, let alone discussion of responsibilities for addressing these issues. It is anticipated that any attempts to rectify the current situation will require a joint effort from the international

[32] An analogy can be drawn with physical infrastructure. A rope bridge and suspension bridge both enable traversal of a river. One may regard the latter as a better implementation overall for doing (more stable, able to carry greater loads) but the former may well be an excellent starting point as it is low cost and can be set up quickly.

[33] https://www.bbc.com/news/world-europe-60125659 (accessed April 4, 2021).

research community, national governments and international data organizations.

As the research landscape continues to evolve through the creation of national and regional Open Science Clouds, these issues are timely. The evolution of FAIR discussions to include principles such as TRUST should serve to further foreground these issues. Indeed, the TRUST principles commit repositories to Monitoring and identifying evolving community expectations and responding as required to meet these changing needs (Lin et al., 2020, p. 3). Recognizing the issues of accessibility means that understandings of what constitutes a "community" need to be critically unpacked. Indeed, the considerable heterogeneity of research communities around the world, and the challenges that they face, needs to be better addressed within FAIR/TRUST discussions, as well as integrated into the technical design of the evolving Open Science landscape.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: 10.5281/zenodo.6411335.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Adam, D. (2019). *Science Under Maximum Pressure in Iran. The Scientist Magazine®*. The Scientist. Available online at: https://www.the-scientist.com/news-opinion/science-under-maximum-pressure-in-iran-66425#.XYOLAuY_hiw.twitter (accessed March 23, 2022).

Bezuidenhout, L., and Havemann, J. (2021). The varying openness of digital open science tools. *F1000* 9:1292. doi: 10.12688/f1000research.26615.2

Bezuidenhout, L., Karrar, O., Lezaun, J., and Nobes, A. (2019). Economic sanctions and academia: overlooked impact and long-term consequences. *PLoS ONE* 14:d6ztj. doi: 10.31730/osf.io/d6ztj

Bezuidenhout, L., Kelly, A. H., Leonelli, S., and Rappert, B. (2017a). '$100 Is Not Much To You': Open Science and neglected accessibilities for scientific research in Africa. *Crit. Public Health* 27:1252832. doi: 10.1080/09581596.2016.1252832

Bezuidenhout, L., Leonelli, S., Kelly, A. H., and Rappert, B. (2017b). Beyond the digital divide: towards a situated approach to open data. *Sci. Public Policy* 44, 464–475. doi: 10.1093/scipol/scw036

CODATA Coordinated Expert Group, Berkman, P. A., Brase, J., Hartshorn, R., Hodson, S., Hugo, W., Leonelli, S., et al. (2020). *Open Science for a Global Transformation: CODATA coordinated submission to the UNESCO Open Science Consultation*. Paris: Zenodo. doi: 10.5281/zenodo.3935461

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., et al. (2018). *Turning Fair Into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. Brussels.

Gaiarin, S. P., Meneses, R., Gallas, S. M., Cepinskas, L., L'Hours, H., von Stein, I., et al. (2021). *D5.8 Pan-European Uptake Final Report (1.0)*. Zenodo. doi: 10.5281/zenodo.5786729

Hess, C., and Ostrom, E. (2006). "Understanding Knowledge as a Commons," in eds C. Hess and E. Ostrom (London: MIT Press). Available online at: https://mitpress.mit.edu/books/understanding-knowledge-commons (accessed March 3, 2022). doi: 10.7551/mitpress/6980.001.0001

Hess, C., and Ostrom, E. (2007). "Introduction: an overview of the knowledge commons," in *Understanding Knowledge as a Commons: From Theory to Practice*, eds C Hess and E Ostrom (Boston: MIT Press), 3–26.

Higman, R., Bangert, D., and Jones, S. (2019). Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights* 32:uksg.468. doi: 10.1629/uksg.468

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., et al. (2020). The TRUST Principles for digital repositories. *Sci. Data* 7:7. doi: 10.1038/s41597-020-0486-7

McDonald, A., Bernhard, M., Valenta, L., VanderSlot, B., Scott, W., Sulluvan, N., et al. (2018). "403 forbidden: a global view of CDN geoblocking," in *ACM Reference Format: Ensafi*, 15. doi: 10.1145/3278532.3278552

Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessl, A., Klebel, T., and Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: a scoping review. *Royal Soc. Open Sci.* 9:211032. doi: 10.1098/rsos.211032

Shanahan, H., and Bezuidenhout, L. (2022). *hughshanahan/Re3DataDownloadAnalysis: V1.0 (Open Science)*. Zenodo.

Wilkinson, M. D., Dumontier, M., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Wilkinson, M. D., Sansone, S-A., Schultes, E., Doorn, P., da Silva Santos, L. O. B., and Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. *Sci. Data* 5, 1–4. doi: 10.1038/sdata.2018.118

Conflict of Interest: LB was employed by DANS (Data Archiving and Networked Services).

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

TABLE A1 | URLs that returned a 403 response code for a given country (returning 200 for us).

Country code	URLs receiving a 403 code
ie	www.hon.ch/HONmedia/
jp	https://www.census.gov/geo/maps-data/data/tiger-line.html
mm	• https://www.census.gov/ • https://dnr.mo.gov/geology/geosrv/
sd	• https://data.mendeley.com/ • https://www.mendeley.com/
sy	• figshare.com • forschung.deutsche-rentenversicherung.de/FdzPortalWeb • ds.data.jma.go.jp/gmd/wdcgg • www.nothobranchius.info • geodata.grid.unep.ch • geocommons.com • geogratis.cgdi.gc.ca • cbeo.communitymodeling.org • henke.lbl.gov/optical_constants • hubblesite.org/gallery • www.mitime.org/mirage • iobis.org • jaspar.genereg.net • kpbc.umk.pl • wdcpc.org • nfdp.ccfm.org • neuromorpho.org • nfdp.ccfm.org • patterns.projects.cis.ksu.edu • qed.econ.queensu.ca/jae • ccdb.wishartlab.com • repository.up.ac.za • www.kadonis.org • wwwdev.ebi.ac.uk/pride • portal.clarin.nl • vsso.cssdc.ac.cn • clarin.informatik.uni-leipzig.de/repo • wals.info • wdc.dlr.de • sedac.ciesin.columbia.edu • simbad.u-strasbg.fr/simbad • slgo.ca/en • westnile.ca.gov • uk-air.defra.gov.uk/data • www.bgs.ac.uk/services/ngdc • www.bmrb.wisc.edu • www.brc.ac.uk • www.caida.org/data • www.calsurv.org • www.cdc.gov/DataStatistics • www.forestdata.cnttp://www.chemspider.com • www.chemsynthesis.com • climate.weather.gc.ca • www.crystallography.net/cod • www.afdc.energy.gov
ye	www.runmycode.org
za	https://www.census.gov/

frontiers | Frontiers in Research Metrics and Analytics

TYPE Original Research
PUBLISHED 12 August 2022
DOI 10.3389/frma.2022.985999

OPEN ACCESS

EDITED BY
Katie Wilson,
Curtin University, Australia

REVIEWED BY
Elvin Shava,
University of Johannesburg,
South Africa
Tshepho Mosweu,
University of Botswana, Botswana

*CORRESPONDENCE
Josiline Chigwada
josyphiri@gmail.com

SPECIALTY SECTION
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

RECEIVED 04 July 2022
ACCEPTED 28 July 2022
PUBLISHED 12 August 2022

CITATION
Chigwada J (2022) Feasibility of a national open data policy in Zimbabwe.
Front. Res. Metr. Anal. 7:985999.
doi: 10.3389/frma.2022.985999

Feasibility of a national open data policy in Zimbabwe

Josiline Chigwada*

Library Department, Chinhoyi University of Technology, Chinhoyi, Zimbabwe

A study on the feasibility of a national open data policy in Zimbabwe was done to document open government data globally and in Zimbabwe. The study showcases the benefits of open government data and the opportunities and challenges toward the development of a national open data policy. Web content analysis and document analysis were used to collect data concerning the readiness of the country in implementing open data activities. The open data barometer was used to gather qualitative data which is essential in assessing the preparedness of the country in opening up government and research data. Content analysis was used to analyse the data which was presented thematically based on the objectives of the study. The findings indicated that the Government of Zimbabwe has endorsed a couple of open data frameworks though some projects are done by non-governmental organizations. The major challenge is implementation of these conventions and commitment to make the data accessible. The results indicated that open data must be made available and accessible within Zimbabwe as a matter of national policy. The author recommends the need for advocacy and continuous awareness creation among the stakeholders so that a national open data policy can be crafted and enacted. The enactment of a national open data policy would guide the use of and access to government data and research data which is valuable in research.

KEYWORDS

open government data (OGD), open data (OD), OECD open data initiatives, open data policy, open data barometer, ZIMSTAT, research data management (RDM)

Introduction

Open Government Data (OGD) initiatives started in mid-2000 at central and local government levels whereby open government data portals were developed in Organisation of Economic Cooperation and Development (OECD) and non OECD countries (Ubaldi, 2013). This is a movement to open up government data by putting it into the public domain to create good conditions for social inclusivity and democracy. There is also a possibility of economic growth through the creation of new products and services using public data by individuals, private enterprises and civil society organizations. The advocates of open government data focus on data in government databases and they deal with the legal and technical issues of access, use and reuse of the databases. The government creates a lot of data through the various ministries as they conduct their day to day activities and this data is valuable to researchers and decision makers as a way of solving the challenges that are being faced in various countries.

In Zimbabwe, the concept of opening up government data was spearheaded by non-governmental organizations and the Open Data Inventory (ODIN) ranks Zimbabwe at number 97 in 2020 with an overall score of 48 (Open Data Watch, 2020). In order for government data to be accessible and useful, there is need for policy backup and this calls the need for a national open data policy in Zimbabwe. It is against this background that the study seeks to answer the following objectives:

- To examine the open data landscape in Zimbabwe.
- To analyze open government and open government data in Zimbabwe.
- To identify the opportunities and challenges that can be used to inform the open data policy for the country.

Problem statement

The open government data movement is concerned with the release of large amounts of government data in various formats and conditions that allow re-use of the data. The development of open science in Zimbabwe is still in its infancy stage as pointed out by Chigwada et al. (2017). The study revealed that there is no open data repository in Zimbabwe and researchers do not want to share their data in fear of intellectual property rights and data abuse. Chiparausha and Chigwada (2019) concurred and added that data is not accessible in Zimbabwe. Government ministries are not willing to share their data because of confidentiality issues and lack of trust among the researchers in fear of abuse of the data. This can also be worsened by the existence of harsh legislation such as the Access to Information and Protection of Privacy Act (AIPPA) of 2002 and Public Order and Security Act (POSA) which could negatively impact open data and information dissemination in Zimbabwe. During the open data day commemorations in March 2019, researchers lamented over the inaccessibility of government data in Zimbabwe where they pointed out that one has to pay large sums of money to access data for research purposes (Open Knowledge Foundation, 2019). The current economic situation in Zimbabwe can also affect the development of open government data initiatives in Zimbabwe since people cannot purchase modern software and equipment to improve the digitization processes. There are also political factors that affect the proper dissemination of information and data which allow some centers to hide damaging government data, and the lack of accountability in data sharing leading to some institutions not releasing nor accounting for their data. The National Open Data Policy can assist in solving some of these challenges. It is against this background that a study on the feasibility of a national policy on open data in Zimbabwe was done to showcase the open data landscape in Zimbabwe pointing out the opportunities and challenges that can be used to inform policy development in the country. This is a background study that was done which would be followed by other studies which would strategise how the policy can be enacted and implemented and how awareness can be created among the policy makers and the general populace. The open data subject is still new in the country meaning that a lot of awareness creation is essential to conscientise the stakeholders on the need and the importance of a national open data policy in Zimbabwe.

Literature review

Open government and open government data globally

Advocates in United States of America (USA) published a set of open government data principles in 2008 leading to the advent of the term Open Government Data (OGD) (Laboutkova, 2015; Hodess, 2019; Tauberer, 2022). The term can be easily understood by defining government data and open data. Government data is regarded as data that is produced by public bodies by using public funds collected through taxes while open data is the data that can be used, reused, distributed freely (OECD, 2022). Open government data initiatives include budget information, business information, registers, patent, and trademark information, public tender databases, geographic information, legal information, meteorological information, social data, and transport information (Ubaldi, 2013; Attard et al., 2015; OECD, 2022). There is a lot of data which is produced by public institutions and opening up this data increases the transparency and accountability to citizens. There is a project called OECD open government data which has a mandate of progressing international efforts on open government data impact assessment (OECD, 2022). Therefore, the open government data principles stipulate that government data shall be considered open if it is complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary, and license-free (Tauberer, 2014; Gong, 2016). The Sunlight Foundation added two more principles on top of the eight which are permanence and usage costs (Ubaldi, 2013). It has been stated that African countries are now opening up their data to attain transparency and accountability. However, although the data portals are being created, there is need for data provisioning in the open formats to populate the portals (Bello et al., 2016). This shows that in both the developed and developing nations, the knowledge of the principles of open government and open government data are known but the challenge is on implementing them.

Open government and open government data in Zimbabwe

In Zimbabwe, although the data is available, it is not accessible as pointed out by researchers who attended the

open data day commemorations at Bindura University of Science Education Library (Open Knowledge Foundation, 2019). Although there have been some baby steps toward the implementation of OECD open government data initiatives in Zimbabwe, the data is only available as summaries (The World Bank, 2019). The open data barometer (Web Foundation, 2018c) which is a global measure indicating how governments are publishing and using open data for accountability, innovation and social impact is one of the measurements that can be used to assess if Zimbabwe is moving along with the current trends in open government data. The open data barometer is also regarded as a research tool that is used to measure the prevalence and impact of open data initiatives in governments around the world (Web Foundation, 2018a).

Beneficiaries of open government data

Open government data is meant to benefit all the stakeholders that are involved which are the government, citizens, civil society and the wider economy including the private sector (Ubaldi, 2013; Gonzalez-Zapata and Heeks, 2015; Pereira et al., 2017; Open Data Govlab, 2019; Open Government Partnership, 2019; Zuiderwijk et al., 2019). The government would be able to make strategic decisions using the data and the data would provide ways of conducting business and allocating resources to be efficient and effective in delivering services. The data would assist in dealing with fraud and other corruption challenges being faced by the government and smarter and innovative services would be delivered to the public. This would also improve the way the government interacts with the users. Open government data help in offering services in a transparent way (OECD, 2022). Citizens benefit through public participation and social engagement in addressing the needs of the public. Therefore, open government data help citizens to increase their quality of life. Civil society organizations play a major role in working with vulnerable segments of the population, e.g., the Sunlight Foundation and the Open Forum Foundation in the United States of America, the Open Knowledge Foundation in Germany and the Open Rights Foundation in the United Kingdom. The private sector is one of the beneficiaries of open government data since they can use the datasets for commercial purposes (Open Government Partnership, 2019). This would assist in providing innovation and experimentation in service delivery. However, these can be regarded as competitors by the government and this can contribute to the reasons why the government institutions do not want to make their data open. However, from the general public and researchers' perspectives, it is difficult to derive these benefits since the data is not easily accessible (Open Knowledge Foundation, 2019).

Benefits of open government and open government data

Open government data is important in increasing public transparency where the public is able to understand what is being done by the government and how well it is performing (The World Bank, 2019). This helps in holding the government accountable for any wrongdoings or if it fails to achieve the results it would have set. It is also a way of showing the good things that the government would be doing since it is accountable to the general population in service delivery. Opening government data helps to generate insights on how to improve government performance through public participation and collaboration in creating value added services (The World Bank, 2019). This shows that individuals and governments would improve the decision making process whereby the public uses government data to have better decisions that improve the quality of life. As a result, the collaborative effort in decision making would ensure that there is trust between the service providers and those who are at the receiving end. Open government data is viewed as a source of economic growth, social innovation and new forms of entrepreneurships (Ubaldi, 2013). The goals of open government data in addition to higher transparency and public accountability is innovation, efficiency and flexibility in government (Yu and Robinson, 2012). Innovation calls for the introduction of new services that are beneficial to the nation leading to less resistance if changes are introduced. The Web Foundation (2018a) summarized the benefits of open data as improving how government resources are used, driving more transparency, accountability and participation, driving social impact by making the policy process more inclusive, and having positive economic impact by boosting economic growth. However, from the Zimbabwean perspective, the custodian of the data is the only stakeholder that is fully benefiting from such data since only summaries are available for public consumption.

Challenges against open data policy

The major challenges in regard to open government data include privacy, legal, financial and technological issues (Ubaldi, 2013). When dealing with data, it needs to be relevant, easily accessible, and reusable by everyone. As a result, there is need to protect the privacy of individuals and ensure that the data can be easily reused. The government should get feedback from data users to check if the data is useful, relevant and accessible to work toward ensuring openness, transparency, accountability, sharing, collaborating and public engagement (Ubaldi, 2013). The FAIR data principles also call for the findable, accessible, interoperable and reusable of data and the principles provide

guidance for data management and stewardship (Wilkinson et al., 2016; Liber, 2017).

Materials and methods

The study was qualitative in nature whereby data was collected from websites and documents mainly using the parameters of the open data barometer. Purposive sampling was used to select websites that document open data readiness and adoption so as to find out the stakeholders that should be involved in enacting and implementing a national open data policy. Web content analysis was used to visit government websites to confirm if the government data is now open. The open data barometer was also consulted to showcase where Zimbabwe is in terms of open data access and usage. Document analysis, looking at country reports and open data initiatives reports, was used to gather data on the infrastructure that is available and the opportunities that can be used to formulate a national policy on open data. Content analysis was used to analyse the results and they were presented thematically using the objectives of the study.

Results and discussion

Findings from the web content analysis reveal that there is no open data policy in place in Zimbabwe and it is not a member of the Open Government Partnerships (Murape, 2021). However, the government had endorsed a couple of open data frameworks such as the Zimbabwe Open Data Initiative (ZODI), a project of the Institute of Data Science Management, Data Management Society, and The Pan African Business Council that works with companies, civil society, and government to build an open and trustworthy data ecosystem (Zimbabwe Open Data Initiative, 2022). Small steps had also been registered through Zimbabwe National Statistics Agency (ZimStat, 2021). The major challenge is implementation and domestication of these conventions so that they can benefit people as stated by Wilkinson et al. (2016) and Liber (2017). The open data landscape in Zimbabwe showed that the data is available but not accessible since every ministry produced its own data and most of the data is available as summaries as stated by The World Bank (2019). However, most of the data is in print format and accessing such data online is not possible. Efforts are being made to digitize such datasets and of note is the Zimbabwe data portal on Open Data for Africa platform (African Development Bank, 2022), although most of the data is not current. Zimbabwe introduced the concept of data centers and the national data center was commissioned in February 2021 (Myles, 2021; Swinhoe, 2021; Towindo, 2021) and would be used to centralize and digitize government services. There is a Centre for High Performance Computing for big data analysis in Zimbabwe which is currently located at the University of Zimbabwe and is open to every researcher and can be fully utilized if open government data initiatives are fully adopted (ZCHPC, 2022). The findings revealed that there is an Academy of Science in Zimbabwe that can assist in opening government data.

The Ministry of Higher and Tertiary Education, Science and Technology Development launched Education 5.0 which is a new development thrust anchored on scientific innovations (Government of Zimbabwe, 2019) and the STEM education which concentrate on Science, Technology, Engineering and Mathematics. It stated that institutions of higher learning should deal with teaching, research, community services, innovation, and industrialization making use of the disruptive technologies such as Internet of Things (IoT), advanced robotics, and automation of knowledge. This led to the development of innovation hubs at six universities to provide platforms for innovators and researchers to develop their innovations with the aid of those with technical and entrepreneurial skills (Satumba, 2021). Universities with complete innovation hubs are National University of Science and Technology, Midlands State University, University of Zimbabwe, Harare Institute of Technology, Zimbabwe Defence University, and the Chinhoyi University of Technology (Nyikadzino, 2022). All these institutions are public universities and if the data is open, it would benefit a lot of researchers and the general populace as stated by OECD (2022).

A look at the open data barometer showed that there were 30 governments that have adopted the open data charter and have committed to G20 anti-corruption open data principles which are termed the leaders edition. The other 115 countries were just listed not as part of open data charter adopters in the year 2017 (Web Foundation, 2018c) and Zimbabwe is one of those 115 countries. This shows that there are signs that the Zimbabwean government is now moving toward open data initiatives but is not yet committed since it had not adopted the open data charter nor signed the G20 anti-corruption open data principles which are a globally agreed set of best practices for publishing, using and maximizing the potential of data (Web Foundation, 2018a). The barometer ranks governments on readiness for open data initiatives, implementation of open data programs, and impact that open data is having on business, politics, and civil society. There are three essential ingredients for good open data governance which are open by default, data infrastructure, and publishing with purpose. Open by default states that governments should build policies, skills and processes to enable a culture of data openness and acceptance of publishing open data. Data infrastructure deals with the building or improvement of technical infrastructure that supports openness in government and organizational transformation. Publishing with purpose is the consideration of the users of open data and what they use the data for to ensure that the data that is published is what is needed and the process is done in a way they can easily use the data (Web Foundation, 2018a). This

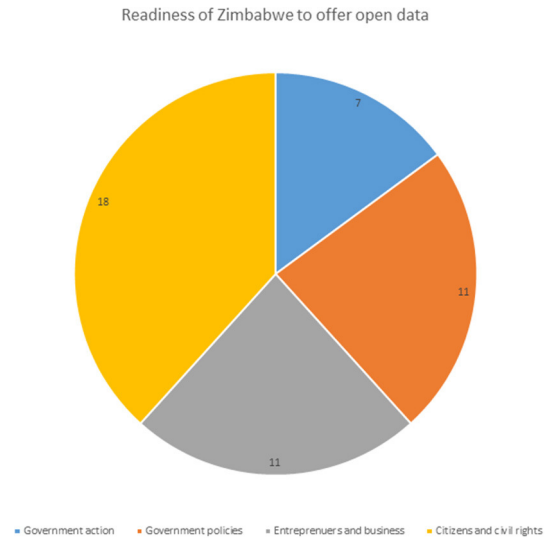


FIGURE 1
Readiness of Zimbabwe to offer open data–Data Source, Web Foundation (2018b).

buttresses what was stated by researchers during the open data day commemorations (Open Knowledge Foundation, 2019) who stated that they are not able to access the government data for research purposes. The readiness of the country to offer open government data is shown in Figure 1.

The summary of information about Zimbabwe on the open data barometer is shown in Figures 1, 2. The readiness status in Figure 1 showed that Zimbabwe scored 7 on government action, 11 on government policies, 11 on entrepreneurs and business and finally 18 on citizens and civil rights. This shows that all the sections of society should be involved in the process of moving toward open data as stated by Pereira et al. (2017), Open Data Govlab (2019), Open Government Partnership (2019), and Zuiderwijk et al. (2019).

Figure 2 showcases the datasets scored in various sectors toward the implementation of open data. A number of questions were asked which include does the data exist, is it available online from government in any form, is the dataset provided in machine-readable and reusable format, is the machine-readable and usable data available as a whole, is the dataset available free of charge, is the data openly licensed, is the dataset up to date, is the dataset being kept regularly updated, was it easy to find information about this dataset, and are data identifiers provided for key elements in the dataset? However, in terms of impact the country scored zero on political, social and economic impact of the available data.

The status of Zimbabwe on the open data barometer 4th edition (Web Foundation, 2018c) showed that Zimbabwe is on number 112 with a score of 2 out of 100 as shown in Figure 3. The score is an overall assessment of the prevalence of open data initiatives. On readiness, Zimbabwe scored 9 out of 100 and this is regarded as the readiness of states, citizens, and entrepreneurs to secure the benefits of open data. On implementation, the score is 4 out of 100 showing the extent to which accessible, timely, and open data is published by each country government on a selection of 15 key fields. On emerging impact, the score is 0 out of 100. This is the extent to which there is any evidence that open data release by the country government has had the impacts in a variety of different domains in the country. As a result, the findings showed that there is currently no evidence on the use and impact of open government data in Zimbabwe and therefore it is difficult to enjoy the benefits of open government data that were pointed out by Yu and Robinson (2012), Ubaldi (2013), Web Foundation (2018a), and The World Bank (2019).

The results indicated that open data must be made available and accessible in Zimbabwe as a matter of national policy to ensure transparency, innovation and efficiency as stated by Yu and Robinson (2012) and Ubaldi (2013). This calls for the need for advocacy and continuous awareness creation among the stakeholders so that a national open data policy can be crafted and enacted. This can be achieved by taking into consideration the open government data principles pointed out by Ubaldi (2013), Tauberer (2014), Laboutkova (2015), Gong (2016), Hodess (2019), and Tauberer (2022).

Conclusion and recommendations

In conclusion, it can be noted that a lot of government data is generated every day and there is need for a national open data policy that would clearly state what open data is and its benefits to the nation at large. It is feasible to have a national open data policy in Zimbabwe since there were some baby steps that were taken by ZimStat and the meteorological department to open up some data and this shows that the government is aware of the benefits of open data. The appearance of Zimbabwe on the open data barometer is a good indicator that there is open government and open government data in Zimbabwe though it is still minimal. The infrastructure that is needed to run the project is already available in Zimbabwe as evidenced by the commitment of various ministries in developing innovation hubs, and data centers which shows political will to make the data available. However, the major challenge that is faced is the accessibility of this data as well as the adoption of the open data charter so as to abide by the 10 principles of open government which are completeness, primacy, timeliness, ease of physical and electronic access, machine readability, non-discrimination, use of commonly owned standards, licensing, permanence and usage costs.

The author recommends the need to build capacity on open government data to ensure that researchers would benefit from the data that is collected using public funds, and to enhance technological and data skills among the researchers and librarians when dealing with data. The stakeholders should be knowledgeable about these principles to ensure that full

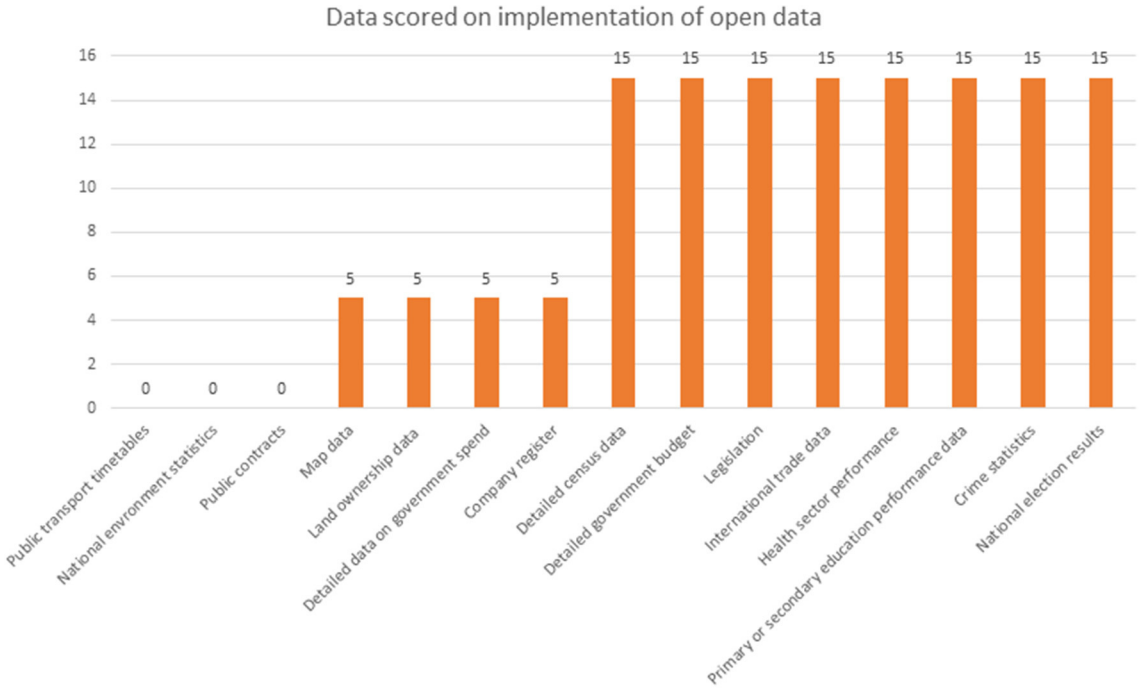


FIGURE 2
Status of Zimbabwe in implementation of open data–Data source, Web Foundation (2018b).

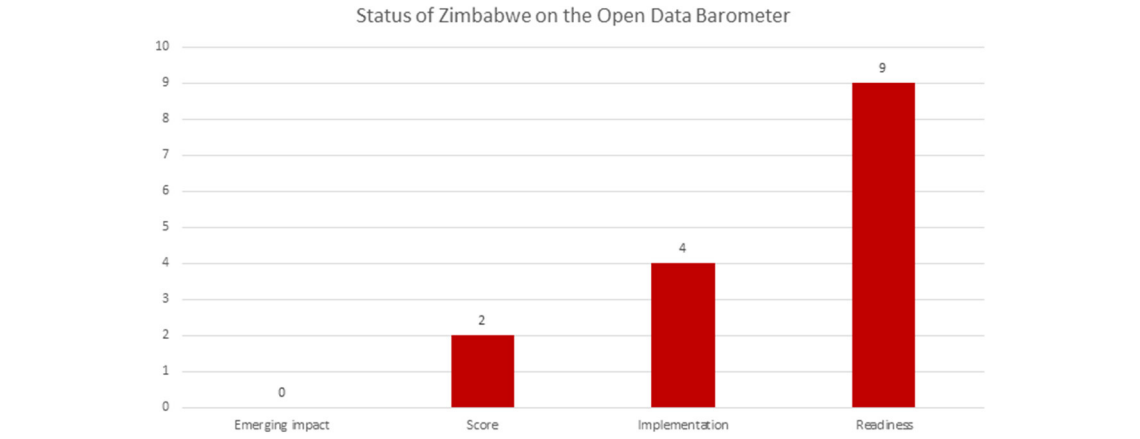


FIGURE 3
Status of Zimbabwe on the Open Data Barometer–Data Source, Web Foundation (2018c).

benefits are drawn from the open government data. Government data should be open by law and it helps to achieve national development and subsequently attain vision 2030 as stated in the sustainable development goals in Zimbabwe, and make research more valuable to the nation. There is need to ensure that there is advocacy and continuous awareness creation among all the stakeholders that are involved so that no one is left behind. The stakeholders include academia, government, corporate world and other professional bodies. This would ensure that equipment, infrastructure, human, and financial resources needed for open data are available. The infrastructure should also be maintained and upgraded to enhance service delivery. The government must prioritize open data governance so that open data becomes part and parcel of how they run their day to day business. There is need for the development of plans, guidelines and procedures of opening up data. The African Open Science Platform (AOSP) had been working on data policy issues with guidance from the Committee on Data for Science and Technology (CODATA) and Zimbabwe can benefit from this initiative to enact and implement a national open data policy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frma.2022.985999/full#supplementary-material

References

African Development Bank (2022). *Zimbabwe Data Portal*. Available online at; https://zimbabwe.opendataforafrica.org (accessed April 10, 2022).

Attard, J., Orlandi, F., Scerri, S., and Auer, S. (2015). A systematic review of open government data initiatives. *Govern. Inf. Q.* 32, 399–418. doi: 10.1016/j.giq.2015.07.006

Bello, O., Akinwade, V., Jolayemi, O., and Ibrahim, A. (2016). Open data portals in Africa: an analysis of open government data initiatives. *Afr. J. Library Arch. Inf. Sci.* 26, 97–106.

Chigwada, J., Chiparausha, B., and Kasiroori, J. (2017). Research data management in research institutions in Zimbabwe. *Data Sci. J.* 16, 31. doi: 10.5334/dsj-2017-031

Chiparausha, B., and Chigwada, J. P. (2019). "Accessibility of research data at academic institutions in Zimbabwe," in *Research Data Access and Management in Modern Libraries*, eds R. Bhardwaj and P. Banks (Hershey, PA: IGI Global), 81–89. doi: 10.4018/978-1-5225-8437-7.ch004

Gong, M. (2016). *Open Government Data Principles*. Six Thirty *Research*. Available online at: https://medium.com/sixthirty/ope-government-dataprinciples/9f06702c58b5 (accessed April 10, 2022).

Gonzalez-Zapata, F., and Heeks, R. (2015). The multiple meanings of open government data: understanding different stakeholders and their perspectives. *Govern. Inf. Q.* 3, 441–452. doi: 10.1016/j.giq.2015.09.001

Government of Zimbabwe (2019). *Doctrine Education 5.0*. Available online at: https://library.buse.ac.zw/docs/gvt-publications/higher-edu-plan.pdf (accessed January 10, 2020).

Hodess, R. (2019). *Open Government Must be More Than a Commitment on Paper*. Available online at: https://apolitical.co/solution_article/open-government-must-be-more-than-a-commitment-on-paper/ (accessed January 10, 2020).

Laboutkova, S. (2015). "Open government data- a lesson to be learned," in *Proceedings of the 12th International Conference Liberec Economic Forum, 2015, Vol. 1* (Liberec).

Liber (2017). *Implementing FAIR Data Principles: The Role of Libraries*. Available online at: https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf (accessed January 10, 2020).

Murape, T. (2021). *About the Open Government Partnership*. Available online at: https://www.zela.org/about-the-open-government-partnership (accessed March 10, 2022).

Myles, D. (2021). *Zimbabwe Establishes National Datacentre Amid Concern Over Data Privacy*. Extensia. Available online at: https://extensia-ltd.com/2021/02/26/zimbabwe-establishes-national-datacentre-amid-concern-over-data-privacy/ (accessed April 10, 2022).

Nyikadzino, G. (2022). *Zimbabwe: Innovation Hubs Spur Industrialisation*. The Herald. Available online at: https://www.herald.co.zw/innovation-hubs-spur-industrialisation (accessed March 10, 2022).

OECD (2022). *Open Government Data*. Available online at: https://www.oecd.org/gov/digital-government/open-government-data.htm (accessed April 10, 2022).

Open Data Govlab (2019). *Private Sector use of Open Government Data: Results From the Open Data 100 Mexico Initiative*. Available online at: https://www.oecd.org/mexico/open-government-data-review-of-mexico-9789264259270-en.htm (accessed January 10, 2020).

Open Data Watch (2020). *Zimbabwe*. Available online at: https://odin.opendatawatch.com/Report/countryProfileUpdated/ZWE?year=2020 (accessed July 26, 2022).

Open Government Partnership (2019). *Research on Impact and Effectiveness of Open Government*. Available online at: https://www.opengovpartnership.org/ogp-multi-donor-trust-fund/research-on-impact-and-effectiveness-of-open-government/ (accessed January 10, 2020).

Open Knowledge Foundation (2019). *Open data day in Taiwan and Zimbabwe*. Available online at: https://blog.okfn.org/2019/03/26/open-data-day-in-taiwan-and-zimbabwe/ (accessed January 10, 2020).

Pereira, G. V., Macadar, M. A., Luciano, E. M., and Testa, M. G. (2017). Delivering public value through open government data initiatives in a smart city context. *Inf. Syst. Front.* 19, 213–119. doi: 10.1007/s10796-016-9673-7

Satumba, C. (2021). *Zimbabwe Embarks on a Programme to Set Up Innovation Hubs at its Universities*. Available online at: https://www.aripo.org/success-stories/zimbabwe-embarks-on-a-programme-to-set-up-innovation-hubs-at-its-universities (accessed March 10, 2022).

Swinhoe, D. (2021). *National Data Centre in Zimbabwe Opens. Data Centre Dynamics*. Available online at: https://www.datacetredynamics.com/en/news/national-data-centre-Zimbabwe-opens (accessed April 10, 2022).

Tauberer, J. (2014). *Open Government Data: The Book*. Available online at: https://opengovdata.io/2014/8-principles/ (accessed January 10, 2020).

Tauberer, J. (2022). *The Annotated 8 Principles of Open Government Data*. Available online at: https://opengovdata.org/ (accessed April 10, 2022).

The World Bank (2019). *Data*. Available online at: https://data.worldbank.org/country/zimbabwe (accessed January 10, 2020).

Towindo, L. (2021). *Data Centre to Improve Service Delivery, Stem Corruption*. Available online at: https://www.sundaymail.co.zw/data-centre-to-improve-service-delivery-stem-corruption (accessed March 10, 2022).

Ubaldi, B. (2013). "Open government data: towards empirical analysis of open government data initiatives," in *OECD Working Papers on Public Governance* (Paris: OECD Publishing).

Web Foundation (2018a). *Open Data Barometer Leaders Edition: From Promise to Progress*. Washington DC: World Wide Web Foundation.

Web Foundation (2018b). *Country Detail*. Available online at: https://opendatabarometer.org/4thedition/detail-country/?_year=2016andindicator=ODBanddetail=ZWE (accessed April 10, 2022).

Web Foundation (2018c). *Open Data Barometer Country Sheets*. Available online at: https://opendatabarometer.org/country-sheets/ (accessed April 10, 2022).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., and Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3. doi: 10.1038/sdata.2016.18

Yu, H., and Robinson, D. G. (2012). The new ambiguity of "open government." *UCLA Law Rev. Discourse* 178–208. doi: 10.2139/ssrn.2012489

ZCHPC (2022). *Zimbabwe Centre for High Performing Computers*. Available online at: https://zchpc.ac.zw (accessed March 10, 2022).

Zimbabwe Open Data Initiative (2022). *Open Data for Development*. Available online at: https://zimbabweopendata.org (accessed April 10, 2022).

ZimStat (2021). *Zimbabwe National Statistics Agency*. Available online at: https://www.zimstat.co.zw (accessed April 10, 2022).

Zuiderwijk, A., Shinde, R., and Janssen, M. (2019). Investigating the attainment of open government data objectives: is there a mismatch between objectives and results? *Int. Rev. Administ. Sci.* 85, 645–672. doi: 10.1177/0020852317739115

frontiers | Frontiers in Research Metrics and Analytics

TYPE Review
PUBLISHED 03 November 2022
DOI 10.3389/frma.2022.950139

OPEN ACCESS

EDITED BY
Katie Wilson,
Curtin University, Australia

REVIEWED BY
Johann Van Wyk,
University of Pretoria, South Africa
Jc Burgelman,
Vrije University Brussel, Belgium

*CORRESPONDENCE
Elisha R. T. Chiware
chiwaree@cput.ac.za

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

RECEIVED 22 May 2022
ACCEPTED 13 October 2022
PUBLISHED 03 November 2022

CITATION
Chiware ERT and Skelly L (2022) Open Science in Africa: What policymakers should consider.
Front. Res. Metr. Anal. 7:950139.
doi: 10.3389/frma.2022.950139

Open Science in Africa: What policymakers should consider

Elisha R. T. Chiware[1*†] and Lara Skelly[2,3†]

[1]Cape Peninsula University of Technology Library, Cape Peninsula University of Technology, Cape Town, South Africa, [2]Loughborough University Library, Loughborough University, Loughborough, United Kingdom, [3]Stellenbosch Business School, Stellenbosch University, Stellenbosch, South Africa

As Open Science (OS) is being promoted as the best avenue to share and drive scientific discoveries at much lower costs and in transparent and credible ways, it is imperative that African governments and institutions take advantage of the momentum and build research infrastructures that are responsive to this movement. This paper aims to provide useful insight into the importance and implementation of OS policy frameworks. The paper uses a systematic review approach to review existing literature and analyse global OS policy development documents. The approach includes a review of existing OS policy frameworks that can guide similar work by African governments and institutions. This critical review also makes recommendations on key issues that Africa should consider in the process of OS policy development. These approaches can be widely used as further foundations for future developments in OS practices on the continent.

KEYWORDS

Africa, Open Science policy, Open Science, policy development, Open Access, Open Data

Introduction

The importance of promoting Open Science (OS) as the vision for the future of conducting science is shared by many, and it is gaining momentum across institutions, governments, and regions at a global level. In different areas–especially in Europe, the US, the UK, and Canada–governments have moved to create the necessary national OS policy frameworks to guide how institutions should respond to the call. The roadmaps toward the OS vision are being shaped by guiding principles such as Open Access; the adoption of Open Data and FAIR (findable, accessible, interoperable and reusable) data principles and citizenship science; the recognition, support and training of researchers; the participation of communities; the development of infrastructures, policies and regulations; and the need for broader stakeholder engagement, coordination and high-level government support (Boulton et al., 2020; Burgelman, 2021; Clark, 2021; Manco, 2021).

In 2021, the United Nations Educational, Scientific and Cultural Organization (UNESCO) provided a set of recommendations as "an international framework for OS policy and practice that recognizes disciplinary and regional differences in OS perspectives, takes into account academic freedom, gender-transformative approaches and the specific challenges of scientists and other Open Science actors in different countries and in particular in developing countries, and contributes to reducing the

digital, technological and knowledge divides existing between and within countries" (UNESCO, 2021). These recommendations are also seen as a support mechanism for a global response to fulfilling the Sustainable Development Goals (SDGs), especially among poorer countries in Africa, Latin America, and Asia. The principles of OS, which include the FAIR and open sharing of scientific research outputs, including data, are seen as an anchor to solving health, developmental, educational and social problems in a more coordinated way (Mwelwa et al., 2020; Abebe et al., 2021; UNESCO, 2021).

Concerns have been raised about the violation of some of the OS principles and its potential impact on the quality of research output during the COVID-19 pandemic–hence a call for "a wider adoption of OS practices in the hope that this work will encourage a broader endorsement of OS principles and serve as a reminder that science should always be a rigorous process, reliable and transparent, especially in the context of a pandemic where research findings are being translated into practice even more rapidly" (Besançon et al., 2021, p. 1). The Organization for Economic Co-operation and Development (OECD) also argued that "in global emergencies like the coronavirus (COVID-19) pandemic, OS policies can remove obstacles to the free flow of research data and ideas, and thus accelerate the pace of research critical to combating the disease" (OECD, 2020).

Drawing on the UNESCO recommendations and emerging research on the critical role of OS principles, there are many opportunities for African institutions and governments to shape their own roadmaps on OS through the development of research infrastructures and supporting, responsive and coordinated policy frameworks at national and institutional levels (Mwelwa et al., 2020). The OECD emphasizes that "to strengthen the contribution of OS to the COVID-19 response, for example, policymakers need to ensure adequate data governance models, interoperable standards, sustainable data-sharing agreements involving the public sector, private sector and civil society, provide incentives for researchers, build sustainable infrastructures, develop human and institutional capabilities and mechanisms for access to data across borders" (OECD, 2020). These principles apply to all other global developmental challenges hampered by climate change, energy provision, and lack of equity in education and other social services.

Reporting on the state of Open Science in research and innovation for development in sub- Saharan Africa, Boulton et al. (2020, p. 7) observed that "the basis of the OS revolution and its impacts" will leave Africa with no alternatives but to respond to its challenges. The value of an OS environment in Africa should be based on two fundamental premises: first, that data sharing and access to scientific data is affordable and easy, and second, that OS engages with society, business, policymakers, governments, communities and citizens as knowledge partners in ways that increase both effectiveness and socio-political legitimacy (Mwelwa et al., 2020).

Emerging continental, regional and national bodies and African programmes are working toward OS's development goals (Boulton et al., 2020; Chiware, 2020; Mwelwa et al., 2020; Abebe et al., 2021). One of the most ambitious projects is the African Open Science Platform, whose mission is to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for modern society. Its building blocks are federated hardware, communications and software infrastructure, including policies and enabling practices to support OS in the digital era; and a network of excellence in OS that supports scientists and other societal actors in accumulating and using modern data resources to maximize scientific, social and economic benefit (Smith and Veldsman, 2018).

Another important continental initiative toward the development of OS is LIBSENSE (2022), launched in 2016 to bring together the research and education networks (RENs) and academic library communities to strengthen OS in Africa. LIBSENSE provides an avenue through which different stakeholder communities can collaborate to define priority activities, share knowledge, and develop relevant services. LIBSENSE is led by the West and Central African Research and Education Network (WACREN) in collaboration with sister regional African RENs (ASREN and UbuntuNet Alliance). Other participating partners include several national RENs, libraries, library associations, universities and research communities in Africa, in conjunction with COAR, EIFL, University of Sheffield, National Institute of Informatics (Japan), GEANT, and OpenAIRE. Outcomes of the LIBSENSE initiative include metadata guidelines for repositories, plans for a regional repository hosting service, and national and institutional policy templates (COAR).

Leading international library and information services organizations are pivotal in enabling African institutions to engage in OS policy development. For example, Electronic Information for Libraries (EIFL, 2022), a not-for-profit organization, works with libraries to allow access to knowledge in developing and transition economy countries in Africa, Asia Pacific, Europe and Latin America. In Africa, EIFL has partnered with library consortia in countries like Ethiopia, Kenya, Uganda, and Zimbabwe and has launched projects to boost open access and OS policy development and to improve repositories and training. Botswana, Kenya, Madagascar, Mauritius, South Africa, and Uganda have progressed toward developing Open Data policies. Ethiopia is the first to have produced a national Open Access policy framework. Through collaborative dialogue with the European Union, South Africa has moved closer to finalizing a national OS framework.

Mwelwa and his fellow researchers (2020) have outlined some barriers, solutions and opportunities for OS in Africa. They have shown that the development of OS in Africa could be used to energize national science systems and enhance the roles they play in supporting the public and private sectors as

well as the general public. However, they pointed out some of the barriers to achieving openness in scientific research work, which include the lack of synergies among "African science systems that largely operate independently of each other, creating silos of incompatible policies, practices and data sets that are not mutually consistent or inter-operable" (Mwelwa et al., 2020, p. 1). Abebe et al. (2021) also argued that "the future of open data management and data sharing and their contribution to the advancement of science and technology in Africa will continue to increase, despite the slow pace caused by the lack of funding, redundant policy frameworks, and limited infrastructures." Abebe et al. (2021) explained that the African landscape is unique; the existing challenges and how they can be addressed will continue to be a big part of African participation in OS and open data global projects.

The purpose of this paper is to explore the significance of national and institutional policy frameworks in promoting OS and what policymakers should consider when developing these policy frameworks. The paper uses a systematic review approach to review existing literature and global OS policy development documents. The approach includes a review of international and national Open Science policy frameworks that can guide similar work by African governments and institutions.

Literature review

The development of national OS policy frameworks can also be guided by the principles that the policies should respond to and support national and institutional goals in order to advance science and knowledge production and sharing. OS policy development in Africa can be framed within key principles that include open access, open data, citizenship science, collaboration, and stakeholder and community engagements. OS policy development should be clearly understood in terms of these fundamental principles of institutions and governments to achieve the end goals of openness, integrity, and FAIR data sharing within the African and global research systems.

As mentioned in the introduction, UNESCO has released draft recommendations serving as "an international framework for OS policy and practice that recognizes disciplinary and regional differences in OS perspectives, takes into account academic freedom, gender-transformative approaches and the specific challenges of scientists and other OS actors in different countries and in particular in developing countries, and contributes to reducing the digital, technological and knowledge divides existing between and within countries" (UNESCO, 2021).

In Europe, Burgelman (2021) wrote about politics and OS and how the European Open Science Cloud (EOSC) has become a reality. The establishment of EOSC is said to be one of the key results that emerged from the policy intentions to foster OS in Europe through the European Open Science Strategy. The other components of this strategy include the Open Science Policy Platform, Open Access Publishing, and the EU Citizen Science Platform. The work to achieve this can also be attributed to the long and complex history of collaboration within the European Union.

In Canada, the government released the Roadmap for Open Science, a set of principles and recommendations to guide the country's federal scientific research. The guidelines and recommendations apply to research by federally employed researchers and research contracted by federal departments and agencies. The Roadmap was developed as part of the commitment to OS as outlined in Canada's 2018–2020 National Action Plan on Open Government (Government of Canada, 2020).

The approaches show governments' and continental bodies' commitment to OS through coordinated policy frameworks that provide institutions with clear guidelines on end goals. Similar continental collaborative OS approaches have evolved in Australia, New Zealand, and the United Kingdom. Africa has the African Union (AU) with a history of policy documents that are not followed through due to member states' lack of financial commitments. Some of the promising projects rely on donor funding. Tieku (2019) pointed out that the African Union has only been able to address the needs of the political elite. The AU, he asserted, has been less successful in connecting its activities and programmes to ordinary Africans. Tieku (2019) also pointed out that the AU has been less successful in providing common public goods and services and has failed to give a voice to the majority of young people and to promote intra-Africa trade. Other areas in which the AU has not fared well include good governance, the financial independence of the continent, and the struggle to address the expressed material needs and quotidian concerns of ordinary Africans (Tieku, 2019).

Boulton et al. (2020, p. vi) recommended that "science systems in Africa...must adapt their working practices" to the "'Open Science' movement," which has been made possible through the use of new digital technologies. This adaptation can be achieved through "the provision of IT infrastructure, policies, incentives, methods and standards for data sharing, policing of ethical standards, and systems and software needed by high-level analytic and AI procedures; such that no individual and few organizations or states in Africa could hope to provide them alone." Boulton et al. (2020, p. vi) also encouraged an approach that has proven effective at "institutional, disciplinary, national, or international levels in scaling up the effort to develop well-managed, integrated digital services and open, sharing practices through OS platforms or commons that serve a broad community through the more or less seamless provision of support and processes for highly creative interactions."

In Africa, Mwelwa et al. (2020) have identified one of the key barriers to OS in Africa and its institutions: the lack of policies, policy coherence, and alignment and harmonization to achieve one big goal of openness. OS practices are new

approaches to doing scientific research. In many ways, these new approaches interfere with established norms–hence the need to guide its uptake through policies and guidelines at institutional and national levels. The African research environment's participation in the global OS movement rests on solid policy frameworks. There is a need to review the progress and make recommendations on how this can be achieved.

In Europe, progress toward the European Open Science Cloud and related advanced research infrastructures has been made through collaboration, policies, engagement with researchers and communities, and the promotion of citizenship science. According to Carillo and Papagni (2014, p. 42), "the production of scientific knowledge is widely recognized as one of the key factors of the economic growth which has occurred in western countries since the Industrial Revolution" and has helped to advance OS to its current levels in those environments. Carillo and Papagni (2014) also regarded the institution of "Open Science" as a cause of scientific and economic inequalities among countries. In developing countries, the limits in knowledge production due to lack of investment in research, lack of incentives, brain drain and slow pace of digital infrastructure development have primarily accounted for the slow pace of OS advancement (Chiware, 2020). Therefore, a general understanding of the politics, progress and barriers regarding Africa's adoption of OS can assist in addressing some of the challenges and bolstering good practices (Boulton et al., 2020). Abebe et al. (2021, p. 9) emphasized this point by noting that "the unique African landscape, and especially the existing challenges and how they can be addressed, will continue to play a big part in African participation in OS and open data global projects." National and institutional policy frameworks, regulations or legislation on data sharing, access, and use are all necessary steps in enabling OS in African academic and research institutions.

Methods and tools

The systematic review followed the protocol described by Dempster (2003), which is specific for social science research. It brought in the following elements from the Preferred Reporting Items of Systematic Reviews and Meta-Analysis checklist (Page et al., 2021) to add to clarity.

Eligibility criteria

Results were not limited to any particular geographical area, date range or policy actor. The only criterion that was set was to include English language results only. These limitations were due to the language limitations of the researchers.

TABLE 1 Data sources used (searches performed on 16 February 2022).

Datasource	Results
Web of science	12
EbscoHost collections: Academic Search Premier; Africa-Wide Information; Business Source Premier; CINAHL; EconLit; ERIC; GreenFILE; Health Source: Nursing/Academic Edition; Library, Information Science & Technology Abstracts; MasterFILE Premier; MEDLINE; Newspaper Source	155
Scopus	93
Google scholar	127

Search strategy

The keywords used were "open research" OR "Open Science" appearing in the title field and "policy" OR "policies" appearing in the title or abstract fields. The exception for the latter was made in Google Scholar, which did not allow for searches in title *or* abstract fields, only title fields. Table 1 shows the data sources used and the number of results found.

Validity

Results discussing open research agendas, questions, directions or issues were excluded, as were any describing grassroots movements within disciplines. The included work had to address a particular policy actor(s). The vast majority of published scientific works mention policy as a tangential topic within conclusions. As policies were a key focus of this study, "policy" or "policies" had to appear at least five times in the body of the work. Where only the abstracts of the conference proceedings were available, they were excluded; full papers were included.

Data extraction

All results were exported to Mendeley as the preferred tool to house the documents. After that, the system's deduplication function was employed. Next, the titles were screened for relevance, followed by the abstracts. Where abstracts were missing, the first paragraph was screened. After the abstract screening, full texts were searched for five mentions of "policy" or "policies" and then screened for relevance. Figure 1 shows the flow diagram.

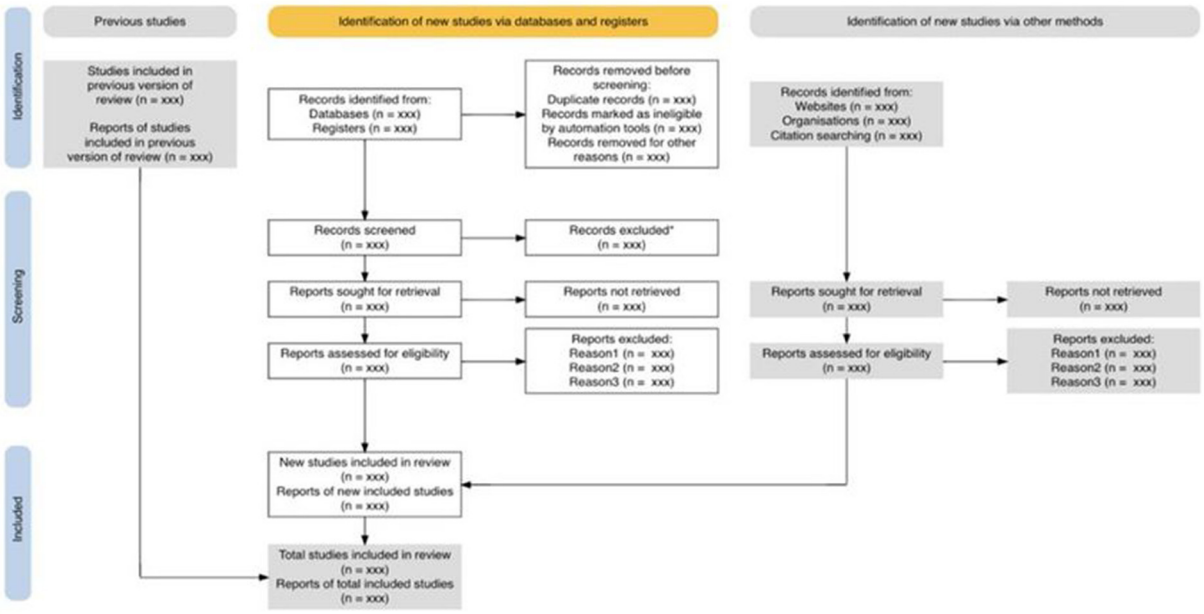


FIGURE 1
The PRISMA 2020 statement: an updated guideline for reporting systematic reviews (Adapted from Page et al., 2021).

TABLE 2 Number of articles per year.

Year	Number of articles
2015	2
2016	1
2017	5
2018	4
2019	3
2020	6
2021	7
2022	1

Synthesis methods

The included results were coded according to the scheme suggested for synthesizing the results of a systematic review of policy and practice by Snilstveit et al. (2012), who provided a three-part framework showing the importance of the policy (defining and framing the problem), examining policy examples (assessing potential policy options), and showing policy implementations (identifying policy implementation considerations for selected policy options). Parallel to the policy scheme coding, the policy actors were coded with an open coding method.

Results

Overview

The included papers are relatively new, with the oldest works coming from 2015. This spread reflects the relative newness of this topic in the published literature. Table 2 shows the distribution over time.

Despite the small time range covered by this literature review, one can see how interests in OS policies have moved among policy actors. Figure 2 is a semantic map showing that the earliest actors were authors primarily concerned with publications (papers) and that interest has shifted to an institutional level (universities) concerned with implementation. This shift reflects a certain degree of maturity in the policy landscape. That is, no longer are authors simply interested in the science; now, universities are interested in the practicalities.

The included papers come from a wide geographical spread, as shown in Table 3, with only the continents of South America and Australia not being represented.

Considering the currency of the topic of focus, it is not surprising that most inclusions are journal articles with a good representation of conference papers. Table 4 shows the varying publication formats of the included documents.

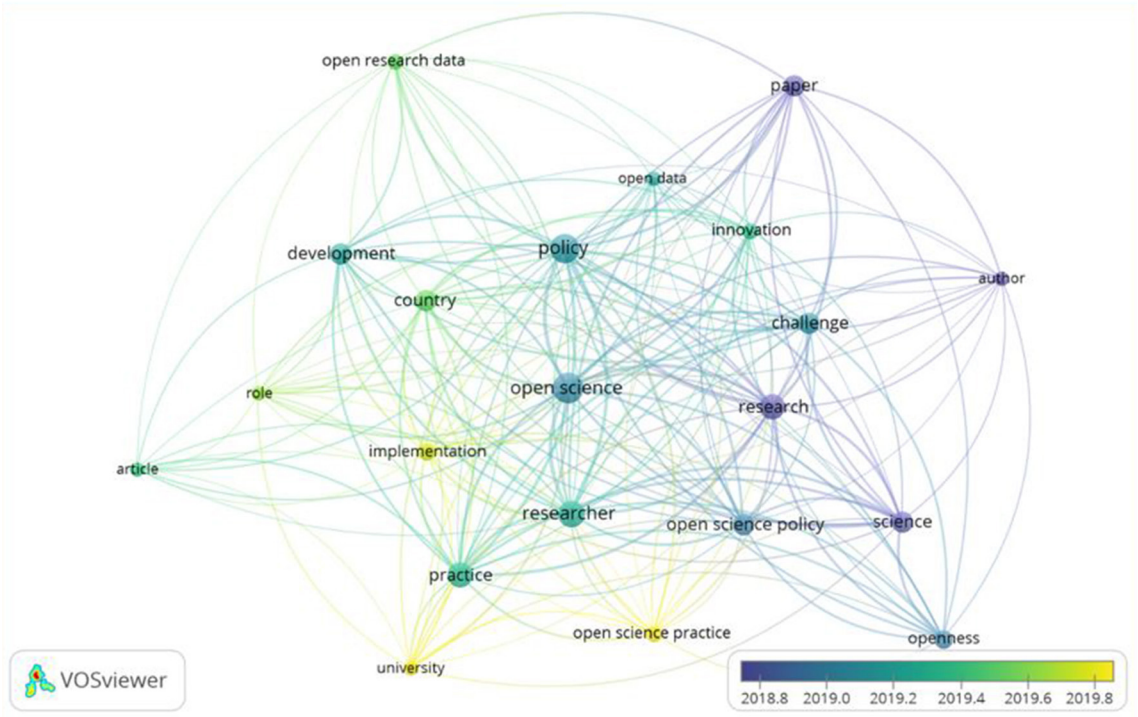


FIGURE 2
Semantic map of titles and abstracts of included articles, colors showing the average age of the keyword. VOSViewer settings: Abstract and fulltext fields; ignoring labels and statements; binary counting; minimum occurrences of terms: five; all 21 terms selected.

TABLE 3 Geographical spread of papers.

Geographic area	Number of articles
International	4
Europe	3
Canada	2
China	2
Finland	2
Africa	1
Albania	1
Botswana	1
Hong Kong	1
Malaysia	1
South Africa	1
United States of America	1
No geographical region	9

TABLE 4 Publication type of included papers.

Publication type	Number of papers
Conference paper	5
Journal article	16
Empirical journal article	7
Report	1

Analysis

Importance

This analysis begins with framing the importance of the OS policies. The literature is abundant with details of potential or realized benefits of various OS schemes. However, in this study, only four articles could be identified that focused on setting out the importance of the open research policies.

The first of these articles is a study by Albornoz et al. (2020), showing that policies are an expression of the policy actor's values and a codification of how such values are expressed. They conceived policies as "instruments that articulate paradigms that can sustain or relocate power and legitimacy" (p. 3). Policies cannot, therefore, ever be considered neutral, OS policies included. Casting policies in this light serves as a starting point for African policymakers to enter into discussions about OS policy development. What values are the OS policies embodying? To whom are they conferring power, and from

whom are they removing legitimacy? These questions could reveal possible objections or low uptake by target audiences.

Resources provide many with a solid power base, which is conferred by awarding research grants. Funders are relatively new players in the scene of OS policies. However, their importance has been growing, especially in low-resource environments where they play a significant role in funding research. The European Commission's Horizon 2020, published in 2014, mandated data management plans and open access publication (Burgelman et al., 2019). Funding was provided to share work earlier and share data. An important finding, which led the European Commission to support the principles of data that is findable, accessible, interoperable and reusable (FAIR), was the financial articulation of the opportunity cost of not having FAIR data: an estimated EUR10.2 billion for the European science system and EUR16 billion for the wider science system (Burgelman et al., 2019, p. 4). Policies that support open research have a clear financial benefit.

In thinking of who might lose power in a greater shift toward OS, journals that have enjoyed a position of exclusivity come to mind. A bold article published in the South African Journal of Industrial Psychology provides journal editors with several policy improvements that would enhance the journal's credibility and transparency (Efendic and Van Zyl, 2019). The authors proposed that journals that take the lead in OS policies have an opportunity to develop standards for a new scientific practice, which would, presumably, position those journals as continued preferred publishing avenues. Indeed, journals with OS policies are likely to be viewed more favorably by those in practice, where the openness assures readers outside of academia of the integrity of the science. Such research is more likely to be used for the benefit of society, a strong motivator for the importance of OS practices (Aguinis et al., 2020).

Implementation

Power, value, ethics

The expression of power and values that policies represent is a theme built upon by Lilja (2020) in her article looking at policy implementation. She raises issues from the perspective of the principal-agent theory. She puts forward that as policies shift power bases and as they express values that the implementing parties might not share, there is a risk of powerlessness and meaninglessness, which would create policy alienation. It is crucial, claimed Ali-Khan et al. (2017), that policies are born from the bottom-up, at least in part. Without the buy-in of those who will be implementing or be affected by the policies, there is a risk of policy alienation, resulting in an unsuccessful policy.

Policymakers should consider readiness for the OS policy, including "awareness, practices, and the perceived benefits" (Ahmed et al., 2019, p. 2). An earlier study from Canada, for example, showed that key partners must shape any successful policy. Still, before that can be done, there should be agreement on expectations, boundaries and engagement mechanisms (Ali-Khan et al., 2018). Armeni et al. (2021) showed the vital role that OS communities have in engaging role players to this end. Policymakers would be wise to partner with these communities to work toward their common goal.

However, not all stakeholders in the field of OS have a common goal. There is an evident tension between OS and the need to exploit findings for commercial gain. Industry partners have long relied upon researchers to provide sound conclusions on which to build products that meet consumer needs, but if those self-same researchers are pushed to share the findings openly, then the industry partners are left in a challenging position[1].

There is more tension than clarity in this area, according to Chataway et al. (2017), more questions than answers. Policymakers would be wise to pay heed to this as it has possible implications within research as well–for example, in the cases given by Levin and Leonelli (2017). In one case, they discuss whether a particular piece of software, which is both a tool and a product of research, should be made openly available. Staunch supporters of the open movement would not hesitate to agree that it should. However, the researchers are using the software as bait to collect other datasets, and make those datasets available to other researchers, which is a far greater boon to research. In such a case, a blanket policy for openness would not serve well. The second example makes this point more evident. Levin and Leonelli (2017) used the case of a researcher who used transgenic mice in their work. As was the case with the software, these mice are both a tool and a product of the research. Naturally, such mice cannot be made freely available. To do so would simply be unethical. The ethical debates of OS are well summarized by Beauvais et al. (2021), who encouraged policy actors to engage with them to ensure that "Open Science can achieve its full potential" (p. 5), a potential they continue "can be envisioned, metaphorically, as a marathon, not a sprint."

Implementation frameworks

Looking beyond the navigation of the problematic issues that OS brings regarding power, values, and ethics, several valuable frameworks can be found in the literature to assist policymakers in framing robust instruments. Pontika et al. (2015) provided a thorough taxonomy of the OS landscape (see Figure 3). Not all OS policies will or should address all these areas; policymakers in Africa can use this taxonomy as a mapping tool to chart out pathways to ever more openness.

Morais et al. (2021) provided a helpful list of emergent areas of OS–including open collaborative tools, open physical labs, and crowdsource practices–which can be used to expand Pontika et al.'s taxonomy (2021). The list of emerging areas

1 There have been movements that are not reported among the papers of this review and that reflect a growing openness in commercial spaces. These movements would stand in contrast to the findings here.

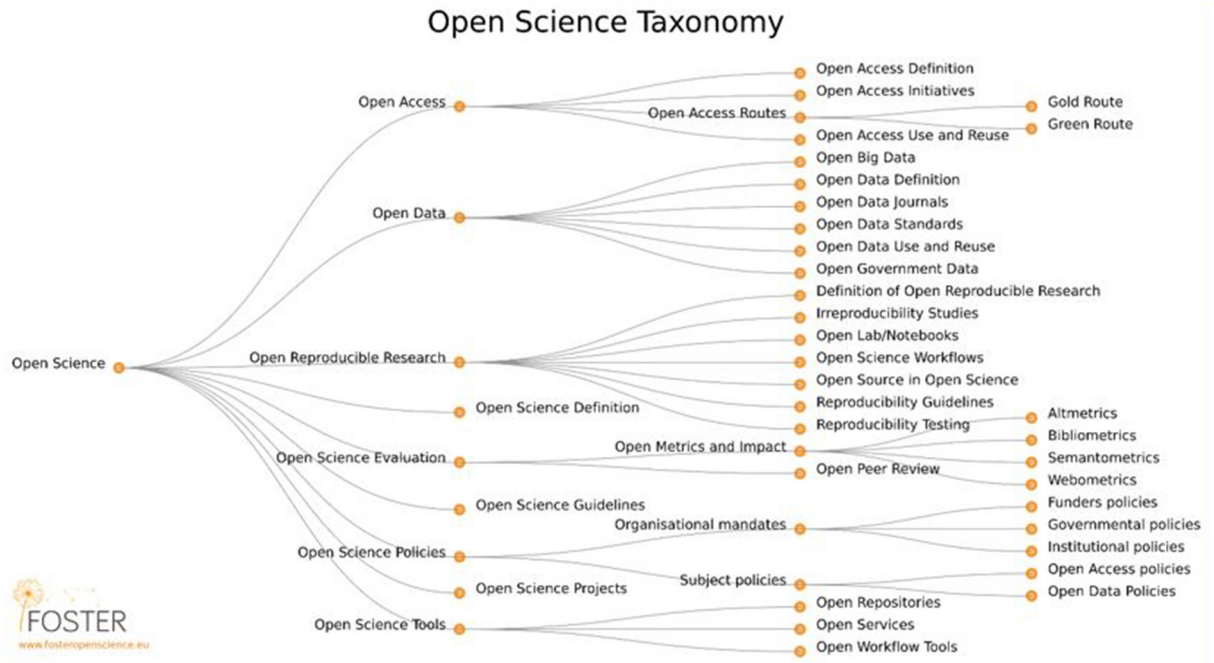


FIGURE 3
Open Science Taxonomy from Pontika et al. (2015). Figure available at https://doi.org/10.6084/m9.figshare.1508606.v3.

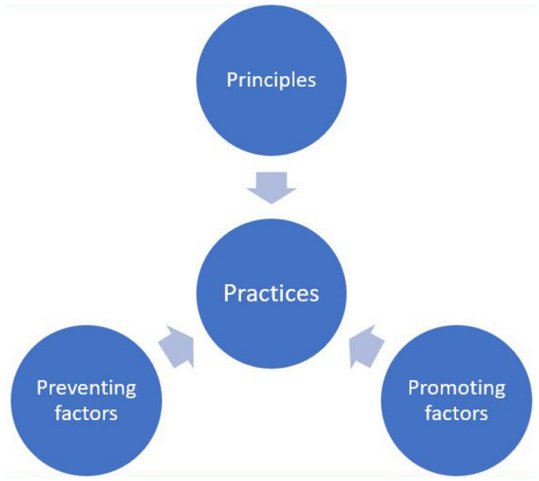


FIGURE 4
Conceptual model of governance of Open Science (adapted from Assist policymakers in framing robust instruments).

serves as a reminder to policymakers that the field of OS is by no means complete and that there will always be new areas and new issues to consider. Here, Vicente-Saez et al. (2020) provided a useful model for policymakers to bring new issues into their frame of thinking. Their model is adapted in Figure 4. It includes considerations around principles, promoting factors and preventing factors that all contribute to the practices within OS.

Examples

There is no shortage of examples of OS policies in the literature; those listed in Table 5 are a sample that emerged through the systematic review process. Policymakers who are disheartened in the challenging process can find solidarity in the cases outlined in these examples. For more examples of policies, without narratives, the Registry of Open Access Repository Mandates and Policies (ROARMAP) is a treasured resource.

Conclusion

The systematic review of literature highlighted what African policymakers should consider in terms of OS policy development in government and in academic and research institutions. It is clear from the review that what is important to policymakers in Africa is a consideration of the significance and value of OS and the accompanying policy frameworks. OS environments should be seen as more than technical problems and infrastructure development; they should also be seen as tools and mechanisms to solve broader societal problems. Levin and Leonelli (2017, p. 284) emphasized this point: "Openness is not only a technical problem to be solved but is also a social, cultural, and moral issue."

TABLE 5 Examples of Open Science policies.

Country/region	Detail	References
Africa	University open data policies	Chiware, 2020
Albania	National science policies	Hasani et al., 2021
Botswana	National open policies	Ntlotlang, 2019
Canada	Research institute	Poupon et al., 2017
China	National open research policies	Li et al., 2022
China	National open research data policies	Zhang et al., 2021
Europe	Behind-the-scenes, regional policy	Burgelman, 2021
Finland	University library policies	Saarti et al., 2020
Hong Kong	National Open Science policies	Sharif et al., 2018
International	Funder policies	Borchert and Proudman, 2018
International	Funder policies	Clobridge and Hinsdale, 2018
International	Overview	Kuchma, 2017
International	Journal Open Science policies	Nosek et al., 2015
United States	National research data policies	Joseph, 2016

Another critical point coming out of the analysis is the issues of OS policy readiness and, as Ahmed et al. (2019) pointed out, policymakers should consider the readiness for the OS policy, which would include awareness, practices, and the perceived benefits. Building onto this point are issues related to existing frameworks that should be considered in shaping the African OS policy environment. Existing frameworks, including tested taxonomies, are readily available and should be utilized. To strengthen these frameworks Beauvais et al. (2021, p. 5) advised that "technical considerations and responses to them must go hand in hand with ethical, legal and social ones." In addition, when considering the uniqueness of the African continent, Vicente-Saez et al. (2020) provided a useful model that policymakers could use to bring new issues into their frame of thinking. This model is centered around practice that should consider aspects of principles, promotion and presentation factors.

Future policymakers can use the findings of this review to engage with policy stakeholders in a manner that will hopefully allow them to enact their values meaningfully. It can be used to examine policy failure and plot a path to the future.

Author contributions

EC led the conceptualization, introduction, literature review, and conclusion. LS took the lead in methods, literature analysis, presentation of results, and discussion. All authors contributed to the article and approved the submitted version.

Funding

Stellenbosch Business School has partially funded this work.

Acknowledgments

The authors acknowledge Amanda Matthee for editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abebe, R., Aruleba, K., Birhane, A., Kingsley, S., Obaido, G., Remy, S., et al. (2021). "Narratives and counter narratives on data sharing in Africa," in *ACM FAccT Annual Conference* (Virtual Event Canada), 3–10.

Aguinis, H., Banks, G. C., Rogelberg, S. G., and Cascio, W. F. (2020). Actionable recommendations for narrowing the science-practice gap in open science. *Organ. Behav. Hum. Decis. Process.* 158, 27–35. doi: 10.1016/j.obhdp.2020.02.007

Ahmed, A. I. A., Ab Hamid, S. H., Gani, A., Khan, S., and Khan, M. K. (2019). Trust and reputation for internet of things: fundamentals, taxonomy, and open research challenges. *J. Network Comput. Appl.* 145. doi: 10.1016/j.jnca.2019.102409

Albornoz, D., Huang, M., Martin, I. M., Mateus, M., Touré, A. Y., and Chan, L. (2020). "Framing power: tracing key discourses in open science policies," in *22nd International Conference on Electronic Publishing - Connecting the Knowledge Commons: From projects to sustainable infrastructure*. OpenEdition Press. Available online at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085471179&partnerID=40&md5=4d9a26ae9080a4c38e819ac062070c9c (accessed October 24, 2022).

Ali-Khan, S. E., Harris, L. W., and Gold, E. R. (2017). Point of view: motivating participation in open science by examining researcher incentives. *ELIFE*. e29319: 1–12. doi: 10.7554/eLife.29319.005

Ali-Khan, S. E., Jean, A., and Gold, E. R. (2018). Identifying the challenges in implementing open science. *MNI Open Res.* 2, 5. doi: 10.12688/mniopenres.12805.1

Armeni, K., Brinkman, L., Carlsson, R., Eerland, A., Fijten, R., Fondberg, R., et al. (2021). Towards wide-scale adoption of open science practices: the role of open science communities. *Sci. Public Policy*. 48, 605–611. doi: 10.1093/scipol/scab039

Beauvais, M. J. S. S., Knoppers, B. M., and Illes, J. (2021). A marathon, not a sprint – neuroimaging, open science and ethics. *Neuroimage*. 236, 118041. doi: 10.1016/j.neuroimage.2021.118041

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., et al. (2021). Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med. Res. Methodol*. 21, 117. doi: 10.1186/s12874-021-01304-y

Borchert, M., and Proudman, V. (2018). "Someone has to pay: the global sustainability coalition for open science services (SCOSS)," in *IATUL Annual Conference Proceedings* (Oslo), 1–8.

Boulton, G., Loucoubar, C., Mwelwa, J., Wafula, M., Ozor, N., and Bolo, M. (2020). Digital revolution, open science and innovation for development in sub-Saharan Africa. A policy brief for the African science granting councils. Available online at: http://hdl.handle.net/10625/59529

Burgelman, J. C. (2021). Politics and open science: how the European open science cloud became reality (the untold story). *Data Intell.* 3, 1–15. doi: 10.1162/dint_a_00069

Burgelman, J. C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., et al. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Front Big Data*. 2, 43. doi: 10.3389/fdata.2019.00043

Carillo, M. R., and Papagni, E. (2014). Little science and big science: the institution of open science as a cause of scientific and economic inequalities among countries. *Econ. Model*. 43, 42–56. doi: 10.1016/j.econmod.2014.06.021

Chataway, J., Parks, S., and Smith, E. (2017). How will open science impact on university-industry collaboration? *Foresight STI Gov*. 11, 44–53. doi: 10.17323/2500-2597.2017.2.44.53

Chiware, E. R. T. (2020). Open research data in African academic and research libraries: a literature analysis. *Libr. Manag.* 4, 383–399. doi: 10.1108/LM-02-2020-0027

Clark, J. (2021). Open Science: a question of trust. *Data Intell.* 3, 64–70. doi: 10.1162/dint_a_00078

Clobridge, A., and Hinsdale, E. (2018). *Open Research Advocacy: Funding Organizations Step Up*. Available online at: https://www.infotoday.com/OnlineSearcher/Articles/The-Open-Road/Open-Research-Advocacy-Funding-Organizations-Step-Up-126172.shtml (accessed October 24, 2022).

Dempster, M. (2003). "Systematic review," in The A-Z of Social Research, eds R. L. Miller, and J. Brewer (London: SAGE). doi: 10.4135/9780857020024

Efendic, E., Van Zyl, L. E. (2019). On reproducibility and replicability: arguing for open science practices and methodological improvements. *SA. J. Ind. Psychol.* 45, 1–10. doi: 10.4102/sajip.v45i0.1607

EIFL (2022). *Electronic Information for Libraries*. Available online at: https://www.eifl.net/ (accessed October 24, 2022).

Government of Canada, Office of the Chief Science Advisor of Canada (2020). Roadmap for open science. Available online at: https://science.ised-isde.canada.ca/site/science/sites/default/files/attachments/2022/Roadmap-for-Open-Science.pdf

Hasani, S., Stefanova, E., Georgiev, A., Stefanov, K. (2021). "First steps towards open science in Albania," in *2021 International Conference Automatics and Informatics, ICAI* (Varna, Bulgaria), 235–238.

Joseph, H. (2016). The evolving US policy environment for open research data. *Inf. Serv. Use*. 36, 45–48. doi: 10.3233/ISU-160806

Kuchma, I. (2017). "Policy framework and roadmap for open access, open research data and open science," in *4th International Conference on Digital Presentation and Preservation of Cultural and Scientific Heritage, Bulgarian Academy of Sciences, Institute of Mathematics and Informatics*, eds R. Pavlov, and P. Stanchev (Sofia, Bulgaria: UNESCO), 333–352.

Levin, N., Leonelli, S. (2017). How does one "open" science? questions of value in biological research. *Sci. Technol. Human Values*. 42, 280–305. doi: 10.1177/0162243916672071

Li, C., Zhou, Y., Zheng, X., Zhang, Z., Jiang, L., Li, Z., et al. (2022). Tracing the footsteps of open research data in China. *Learn. Publ.* 35, 46–55. doi: 10.1002/leap.1439

LIBSENSE (2022). *The LIBSENSE Initiative*. Available online at:https://libsense.ren.africa/en/ (accessed October 24, 2022).

Lilja, E. (2020). Threat of policy alienation: exploring the implementation of open science policy in research practice. *Sci. Public Policy*. 47, 803–17. doi: 10.1093/scipol/scaa044

Manco, A. (2021). A landscape of open science policies research. *Advance*. 1–27. doi: 10.31124/advance.16553133

Morais, R., Saenen, B., Garbuglia, F., Berghmans, S., and Gaillard, V. (2021). From principles to practices: open science at Europe's universities. 2020-2021 EUA open science survey results. In: European University Association. Available online at: https://www.eua.eu/downloads/publications/2021%20os%20survey%20report.pdf

Mwelwa, J., Boulton, G., Wafula, J. M., Loucoubar, C. (2020). Developing open science in Africa: barriers, solutions and opportunities. *Data Sci. J.* 19, 31. doi: 10.5334/dsj-2020-031

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science* 348, 1422–1425. doi: 10.1126/science.aab2374

Ntlotlang, T. (2019). Embracing open science agenda through coordinated approach: the Botswana case. *Cadernos de Biblioteconomia, Arquivística e Documentação*. 1, 86–94. Available online at: https://www.bad.pt/publicacoes/index.php/cadernos/article/view/2015

OECD (2020). *Why Open Science is Critical to Combating COVID19. OECD Policy Responses to Coronavirus (COVID-19)*. Paris: OCDE.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 372, n71. doi: 10.1136/bmj.n71

Pontika, N., Knoth, P., Cancellieri, M., Pearce, S. (2015). "Fostering open science to research using a taxonomy and an elearning portalQ17," in *15th International Conference on Knowledge Technologies and Data-Driven Business, i-KNOW*. Graz Austria: ACM.

Poupon, V., Seyller, A., and Rouleau, G. A. (2017). The tanenbaum open science institute: leading a paradigm shift at the montreal neurological institute. *Neuron*. 95, 1002–1006. doi: 10.1016/j.neuron.2017.07.026

Saarti, J., Rosti, T., and Silvennoinen-Kuikka, H. (2020). Implementing open science policies into library processes: case study of the university of Eastern Finland library. *LIBER Q.* 30, 1. doi: 10.18352/lq.10336

Sharif, N., Ritter, W., Davidson, R. L., and Edmunds, S. C. (2018). An open science' state of the art' for Hong Kong: Making open research data available to support Hong Kong innovation policy. *J. Contemp. East. Asia*. 17, 200–221. https://doi.org/10.17477/jcea.2018.17.2.200

Smith, I., and Veldsman, S. (2018). "Data driving sustainability: the African Open Science Platform Project,"in *22nd International Conference on Electronic Library Publishing Conference*. Toronto.

Snilstveit, B., Oliver, S., and Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *J. Dev. Effect*. 4, 409–429. doi: 10.1080/19439342.2012.710641

Tieku, T. K. (2019). "The African Union: successes and failures,"in *Oxford Research Encyclopaedia of Politics*. Oxford: OUP.

UNESCO (2021). *Draft Recommendation on Open Science*. Paris, UNESCO.

Vicente-Saez, R., Gustafsson, R., and Van den Brande, L. (2020). The dawn of an open exploration era: emergent principles and practices of open science and innovation of university research teams in a digital world. *Technol. Forecast. Soc. Change*. 120037:1–10. doi: 10.1016/j.techfore.2020.120037

Zhang, L., Downs, R. R., Li, J., Wen, L., and Li, C. (2021). A review of open research data policies and practices in China. *Data Sci. J.* 20, 1–17. doi: 10.5334/dsj-2021-003

frontiers | Frontiers in Research Metrics and Analytics

TYPE Perspective
PUBLISHED 07 November 2022
DOI 10.3389/frma.2022.982435

OPEN ACCESS

EDITED BY
Hamid R. Jamali,
Charles Sturt University, Australia

REVIEWED BY
Anwar Vahed,
Council for Scientific and Industrial Research (CSIR), South Africa
Simon Wakeling,
Charles Sturt University, Australia

*CORRESPONDENCE
Tony Hey
tony.hey@stfc.ac.uk

SPECIALTY SECTION
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

RECEIVED 30 June 2022
ACCEPTED 17 October 2022
PUBLISHED 07 November 2022

CITATION
Hey T (2022) Open science and Big Data in South Africa.
Front. Res. Metr. Anal. 7:982435.
doi: 10.3389/frma.2022.982435

Open science and Big Data in South Africa

Tony Hey*

Scientific Computing Department, Rutherford Appleton Laboratory, Science and Technology Facilities Council/UK Research and Innovation (STFC/UKRI), Didcot, United Kingdom

With the Square Kilometer Array (SKA) project and the new Multi-Purpose Reactor (MPR) soon coming on-line, South Africa and other collaborating countries in Africa will need to make the management, analysis, publication, and curation of "Big Scientific Data" a priority. In addition, the recent draft Open Science policy from the South African Department of Science and Innovation (DSI) requires both Open Access to scholarly publications and research outputs, and an Open Data policy that facilitates equal opportunity of access to research data. The policy also endorses the deposit, discovery and dissemination of data and metadata in a manner consistent with the FAIR principles – making data Findable, Accessible, Interoperable and Re-usable (FAIR). The challenge to achieve Open Science in Africa starts with open access for research publications and the provision of persistent links to the supporting data. With the deluge of research data expected from the new experimental facilities in South Africa, the problem of how to make such data FAIR takes center stage. One promising approach to make such scientific datasets more "Findable" and "Interoperable" is to rely on the Dataset representation of the Schema.org vocabulary which has been endorsed by all the major search engines. The approach adds some semantic markup to Web pages and makes scientific datasets more "Findable" by search engines. This paper does not address all aspects of the Open Science agenda but instead is focused on the management and analysis challenges of the "Big Scientific Data" that will be produced by the SKA project. The paper summarizes the role of the SKA Regional Centers (SRCs) and then discusses the goal of ensuring reproducibility for the SKA data products. Experiments at the new MPR neutron source will also have to conform to the DSI's Open Science policy. The Open Science and FAIR data practices used at the ISIS Neutron source at the Rutherford Appleton Laboratory in the UK are then briefly described. The paper concludes with some remarks about the important role of interdisciplinary teams of research software engineers, data engineers and research librarians in research data management.

KEYWORDS

Open Science, SKA project, FAIR data, neutron data, research data management (RDM)

Big Scientific Data comes to South Africa

With the Square Kilometer Array (SKA) project (The SKA Project, 2022) and the new Multi-Purpose Reactor (MPR) to replace the existing Safari-1 neutron source (Necsa, 2022) soon coming on-line, South Africa and the other collaborating countries in Africa will have to cope with an increasing deluge of scientific data. In a talk given in January 2007, Turing award winner Jim Gray outlined the existing three paradigms of scientific discovery: first, empirical observations; second, theoretical explorations; and third, computational simulations. He then identified the emergence of a fourth paradigm: data-intensive scientific discovery (Hey et al., 2009):

> *"The techniques and technologies for data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration."*

Gray's fourth paradigm builds on the first three paradigms of observation, theory, and computation and requires that research scientists develop new skills in data management and data analysis. The breakthrough in image classification and recognition with "Deep Learning" artificial neural networks in 2012 (Sejnowski, 2018) has already transformed much of the commercial world and is now beginning to have a major impact on scientific data analysis (Stevens et al., 2020). The management, analysis, publication, and curation of "Big Scientific Data" will soon be an important component of open science in South Africa.

This paper is focused on data aspects of the Open Science agenda and on the management and analysis of scientific data sets. Borgman's book on "Big Data, Little Data, No Data: Scholarship in the Networked World" provides an excellent introduction to scientific data policy and practice, as well as discussing some case studies in data scholarship (Borgman, 2015). For the sciences, Borgman uses the example of astronomy as a "Big Data" research field and the example of sensor-networked science as a "little data" research field. However, the extreme data rates and volume of data from the SKA project will be at a totally different scale than any previous astronomy project. The SKA project therefore has the potential to be truly transformative for science and technology in Africa. South Africa will be the location for one of the SKA project's two Science Data Processors (SDP) and for an African SKA Regional Center (SRC). The output of the SDP will be distributed to a global network of SRCs which will produce science-ready data products and provide users with the necessary software tools for analyzing the data.

After a description of the data challenges and opportunities that will be posed to African scientists by SKA-Mid, the South African component of the SKA, the implications of the draft South African National Open Science policy (Pienaar, 2022) are discussed. For research funders globally, there is now an increasing focus on Scientific Data Management plans and the linking of the full text open access papers to the relevant supporting research data. Some evidence for better compliance by researchers with this component of Open Science is provided by the improving situation in the US and the UK.

The draft South African Open Science policy requires not only open access to the full text of the research paper but also access to the digital data necessary to validate the research findings described in the paper, as well as the availability of the software that was used to analyse the data. In addition, the draft policy also specifically requires that the research data should be "FAIR" compliant (FAIR, 2022). The FAIR initiative emphasizes the importance of having machine actionable metadata for interoperability rather than just community-agreed file formats. One approach to implementing FAIR data that has been adopted by the biosciences community is to extend the standard, industry-supported Schema.org vocabulary with specific types relevant to scientific datasets.

Another source of significant scientific datasets in South Africa – though not on the scale of SKA – will be from the beamlines to be built at the new MPR reactor facility. This will be a replacement for the Safari-1 reactor which is one of the top four medical radioisotope producers in the world as well as supporting a wide range of research and applications using their Neutron Diffraction Facility. The MPR project is now in the design phase with construction planned to start in 2025. A new Neutron Beam Line Center is being planned and a new software stack comprising the instrument control system, user interface and data reduction and analysis components will need to be built (Marais, 2022). In addition, the data management processes will need to support the generation of FAIR data. A short section describing the research data management practices and the progress toward generating FAIR datasets at the ISIS Neutron source at the UK's Rutherford Appleton Laboratory is included to illustrate the important components of such a FAIR data pipeline.

The paper ends with some remarks about the need for interdisciplinary teams in research data management and the important roles of research software engineers, data scientists and research librarians.

The SKA project in Africa

The Square Kilometer Array (SKA) project is an international effort to build the world's largest radio telescope, eventually covering over a square kilometer of collecting area. The data generated by the SKA has the potential to answer many open questions in modern astrophysics, ranging from mapping the early cosmic history of the universe to understanding how

galaxies form and evolve (Scaife, in press). The most suitable telescope locations for the SKA were determined to be in remote Western Australia, around 800 km north of Perth, and in the Karoo region in the Northern Cape of South Africa, where the national government has created a radio-quiet astronomy reserve.

The first of the two SKA telescopes will operate at low radio frequencies where radio signals have wavelengths of several meters. In this first phase of the SKA, the SKA1-LOW telescope comprising 130,000 dipole antennas will be sited in Western Australia. The second of the SKA telescopes will operate in the mid-frequency radio band where radio signals have wavelengths from around one meter to tens of centimeters. The desert regions of South Africa provide the perfect radio quiet location for this mid-frequency array, SKA1-MID (SKA South Africa, 2022). The technology for the SKA1-MID instrument will use familiar-looking radio dishes to receive incoming radio signals and, when complete, will comprise 197 individual radio dishes separated by distances of up to 200 km. The second phase of the SKA project (SKA2) will extend the mid-frequency dish array into the eight other SKA African partner countries – Botswana, Ghana, Kenya, Madagascar, Mauritius, Mozambique, Namibia, and Zambia.

In his talk in 2007 in which Jim Gray identified the emergence of a fourth paradigm of data-intensive scientific exploration and discovery, he went on to say (Hey et al., 2009):

> *"People now do not actually look through telescopes. Instead, they are 'looking' through large-scale, complex instruments which relay data to datacenters, and only then do they look at the information on computers. The world of science has changed, and there is no question about this. The new model is for the data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline."*

For the first generation of large-scale experimental facilities, it was possible to conceive of building a single facility that was able to provide end-to-end coverage of the data processing, storage and archiving needs of its users. However, with the new generation of large-scale projects, such as the Large Hadron Collider (LHC) at CERN and the global SKA Observatory project, this approach is no longer feasible. For example, the Worldwide LHC Computing Grid (WLCG) is an international collaborative project that consists of a grid-based computer network infrastructure incorporating over 170 computing centers in 42 countries (Worldwide LHC Computing Grid, 2022). It was designed by CERN to handle the many Petabytes of data produced by Large Hadron Collider (LHC) experiments. In similar fashion, the global SKA Observatory (SKAO) will require building a network of SKA Regional Centers (SRCs) to receive data from the two SKA Science Data Processors (SDPs).

When the two SKA telescopes become fully operational in the mid to late 2020s, the output data products are estimated to amount to approximately 300 Petabytes per telescope per year. Such "Extreme Scientific Data" will require the creation of a novel Science Data Processor (SDP) at each telescope that will be a schedulable part of the telescope (Chrysostomou, 2019). The goal is to reduce the raw data volume at the SDP before delivery to users. However, as for the CERN LHC data, there will also be a need to create a global network of SKA Regional Centers (SRCs) with one located in South Africa. Such a network is needed because the data volumes are so large that direct delivery to a distributed global community of end users is unfeasible. Moreover, the SKA data, as delivered to the SRCs from the telescope *via* the SDP, needs further processing to be in a state suitable for scientific analysis and publication.

Discussions about the precise roles of the SRCs are still ongoing but the SKAO has always had a very strong commitment toward implementing Open Science (Garrido et al., 2021):

> *"The SKAO and the SRC network are working to enable best practices that make data and other digital research objects (e.g., algorithms, tools, workflows, protocols, or services) 'findable, accessible, interoperable, and reusable' (FAIR). In particular, the SRC Coordination Group defined different requirements related to open science, highlighting the requirement of 'open access,' which relates to the need for public links to SKA science data products, and the 'reproducibility: provenance and workflow preservation' requirement, meaning that the SRCs must be capable of saving the provenance and workflow associated with the data products generated at each SRC."*

Remarkably, the SKAO is believed to be the first large-scale facility to include reproducibility as one of the scientific metrics of its success[1].

The South African government sees the SKA project as a catalyst for bringing new technology and skills to the African continent (Ratcliffe, 2022):

> *"Aside from the benefits to African science, Big Data capabilities could be our biggest spin-off from the SKA project. The innovations, skills development and commercial potential emerging as a result of the project are huge. The potential is not just academic – we develop the taxpayer-funded intellectual property to a point where it's ready to become commercialized and benefit the economy. We will increasingly be an incubator of science and technology innovation."*

1 https://www.skatelescope.org/wp-content/uploads/2021/03/22380_SKA_Est-Delivery-Plan_DIGITAL_v3.pdf p. 62–63 (accessed July 2022).

These new skills will involve exploitation of appropriate AI and Deep Learning technologies in both the SKA data pipeline and the analysis of the SKA data products (SKA South Africa, 2022).

Open access and Open Science

The South African Draft National Open Science Policy (Pienaar, 2022) defines Open Access as:

> *"a set of principles and a range of practices through which research outputs are distributed online, free of cost or other access barriers".*

Open Science is defined as:

> *"research and development that is collaborative, transparent and reproducible and whose outputs are publicly available".*

The policy will be applicable to all publicly funded research outputs and will require access to infrastructure at an institutional and national scale that supports the deposit, discovery and dissemination of data and metadata. One of the key guidelines for the successful implementation of the Open Science policy is the adoption of the FAIR principles for research data management and stewardship (FAIR, 2022). The challenge of making the data FAIR will be discussed in terms of SKA and neutron data later in this article.

It is worth comparing progress toward research data management in both the US and UK. In 2013, John Holdren, then director of the US Office of Science and Technology Policy (OSTP), issued a memorandum requiring all major Federal Funding Agencies develop plans to make available the direct results of federally funded scientific research for the public, industry, and the scientific community (Holdren, 2013). Such results include peer-reviewed publications and digital data. The memorandum defined digital data as:

> *"the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens."*

This was one of the earliest attempts to define what "digital data" should be included for research publications. This did not include all the raw observational data taken by experimentalists but only the subset that was relevant to the research publication.

The major US research Funding Agencies have now set up open access repositories and this represents a significant advance toward open access for US research publications. However, the OSTP has recently issued a memorandum[2] recommending that federal agencies:

> *"Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release."*

Since the US published 17% of peer-reviewed Scientific and Engineering articles in journals and conferences in 2018 (NSF, 2020), this OSTP directive is likely to have a major impact on both publishers and researchers globally.

In the UK, the UK Research and Innovation funding agency has also recently issued its official open access policy for peer-reviewed research articles as well as for monographs, book chapters, and edited collections (UKRI, 2022a). However, in both the US and the UK, the policy on data is less clear. In the UK, each of the seven research councils has its own data management policy although all research proposals are now required to include a Data Management Plan (DCC, 2022). UKRI does now require a "Data access statement" with research publications giving information as to how the supporting data for the reported research may be accessed (UKRI, 2022b). At a global level, currently 5.8% of all articles in Scopus have a link to a dataset (Scopus, 2020).

At the UK's National Laboratories, figures from their research publications repository ePubs[3] and the Web of Science Data Citation Index indicate a clear upward trend in the number of articles authored or co-authored by staff that now include Data Access Statements (see Table 1). The figures were obtained by downloading article data from ePubs which have Digital Object Identifiers (DOIs), and then running the DOIs through Web of Science.

TABLE 1 Research articles in the ePubs repository indexed in the Web of Science that have DOIs (20th April 2022).

Year of publication	Number of records returned in WoS	Articles with related data (WoS)
2019	1,328	42
2020	1,375	64
2021	1,234	92

2 https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf (accessed October 2022).

3 The ePubs repository: https://epubs.stfc.ac.uk/ (accessed June 2022).

FAIR data: Schema.org, Bioschemas, and W3C DCAT

The FAIR Data Principles are intended as a guide to making data Findable, Accessible, Interoperable and Reusable (FAIR, 2022). The *machine actionability* of the metadata associated with the data is an important aspect of FAIR data, referring to the ability of machines – and not just humans – to understand and manipulate the data. Until fairly recently, search engines had to scrape the Web using keywords and make a best guess about the actual page content. Even after applying natural language processing techniques to a page, there could still be significant ambiguity as to its content. To overcome this, the search engine companies collaborated in the development of the Schema.org vocabulary, a standard vocabulary of generic terms[4]. This allows web developers to provide a high-level overview of the content of the page by embedding machine processable markup within the page source. For example, for a web page about the movie Casablanca, the markup can specify that it was a Movie type (https://schema.org/Movie) and not a city in North Africa. The markup could also provide additional properties such as the title of the movie using (https://schema.org/name) and the names of the actors (https://schema.org/actor). This embedded markup is then used to create the knowledge graphs of the search engines. While the Schema.org vocabulary is not a World Wide Web Consortium (W3C) Recommendation, there is a W3C community group that oversees the development of the Schema.org[5].

In this paper, the focus is on describing the improvements in the Findability of scientific datasets through the embedding of machine interpretable markup within Web pages as well as improving dataset Interoperability from the use of a common vocabulary of terms. Note that the Schema.org markup only provides a high-level description of what the dataset contains but not all the scientific details required to fully interpret the data. However, the key advance of such web vocabularies is that the markup can be accessed using standard Web protocols and the data can be made available in the JSON-LD format. Users can therefore access data without needing to understand site specific APIs. It is important to stress that since the entry barrier for the data provider is very low, the long-tail of scientific datasets and not just major data providers in a particular field can easily be made accessible.

Although the main focus of Schema.org is to support general Web search, Google offers a dedicated search portal, Google Dataset Search, for collections of data[6]. The Dataset Search Portal allows the user to use a standard keyword search interface to search for collections of data. The content of the Google Dataset Search Portal is obtained from individual web pages using the Schema.org Dataset type. Details of the datasets found can then be shown as well as links to related research articles and the data download location. The corpus of the Google Dataset Search Portal has grown from 500 thousand records in 2016 to 28 million in 2020 (Benjelloun et al., 2020). The portal now contains datasets from a wide range of fields with the Social Sciences (about 26%) and Geosciences (about 19%) dominating.

The Findability of datasets can thus be enhanced by embedding Dataset markup within the homepage of the dataset[7]. When this markup is crawled by Google, the dataset will be added to their internal knowledge graph and become discoverable through their dedicated search portal for datasets as well as for their main search results. However, this search is based on keyword summaries of the dataset and, for example, can only find datasets about diseases in general and not for any specific disease. The Bioschemas initiative is an attempt by the life sciences community to extend the Schema.org vocabulary with specific types relevant to life sciences to enable a slightly deeper inspection of the contents of the resources (Gray et al., in press)[8]. This provides a representation of the key characteristics of the resource and enables an initial level of **I**nteroperability. Nevertheless, to obtain the full set of features of the data, the dataset needs to be retrieved from its original source in the detailed representation format in which it is published. Thus, the Bioschemas proposals do not replace any of the many existing domain ontologies.

The development of the Bioschemas types has followed the philosophy of the Schema.org vocabulary. Instead of trying to accurately capture all the underlying biology, the developed types only aim to capture the characteristics that are most widely used when searching for a concept. In addition, the Bioschemas initiative has developed community agreed usage profiles over the Schema.org types. These profiles identify the core set of properties (typically about 10) to describe a resource of a specific type from the sometimes 100s of available properties for the type. Web page developers simply follow the profile rather than needing to pick and choose which properties to use. The profiles also increase the consistency of the markup available to consuming applications meaning that the data is more viable for Reuse.

Since its inception in 2015, Bioschemas has developed 23 types for describing life sciences concepts, of which 6 have now been included into Schema.org. The community has also defined 37 profiles over these and existing Schema.org types with the goal of making them more accessible to life sciences resource providers. By limiting the number of properties, the process of developing markup for a site is simplified and allows developers to focus more on modeling their own data (Gray et al., in press).

4 https://schema.org/ (accessed June 2022).

5 https://www.w3c.org/community/schemaorg (accessed June 2022).

6 https://datasetsearch.research.google.com/ (accessed April 2022).

7 https://schema.org/Dataset (accessed June 2022).

8 https://bioschemas.org (accessed June 2022).

In summary, a key benefit of building on the Schema.org vocabulary is that it is a globally agreed model for representing data. The Schema.org approach, while not addressing important domain specific details, makes the data usable beyond the immediate community of interest, and due to the low deployment effort needed, this approach can be usefully applied to the long-tail of small datasets. Data marked up using Schema.org can be consumed both by community specific registries and wider cross-domain registries, thus dramatically increasing the reach of the data.

Google Dataset Search also supports datasets described using the Data Catalog (DCAT) vocabulary, which has been a World Wide Web Consortium (W3C) recommendation since 2014 (DCAT, 2022). Thus, any webpages describing datasets using the DCAT vocabulary are also crawled by Google and the datasets are displayed *via* their dedicated data search engine.

DCAT's initial version (DCAT v1.0 from 2014) targeted governmental data repositories but later versions have extended the DCAT vocabulary to include terminology that is more relevant for research data. For instance, DCAT v2.0 (a W3C Recommendation since 2020) now covers all the required terms for data citation, guidelines on identifiers, licensing and access rights, dataset quality information, properties to describe temporal and spatial resolutions of datasets. In addition, DCAT has been made more generic by supporting the cataloging and description of any Web Resource, and in particular, the description and cataloging of DataServices. The latest version, currently undergoing the recommendation process, includes treatment of versioning, dataset series and multiple other improvements.

DCAT is being used extensively in many data portals around the world (e.g., Europeana, Zenodo, governmental data catalogs), and the vocabulary itself has been extended *via* Application Profiles to add restrictions and terminologies required in specific domains (e.g., for statistical and geographical data).

Data management at the ISIS neutron source

Large-scale scientific facilities such as synchrotrons, neutron sources and lasers produce massive amounts of data, which are continuing to grow as their technology improves. Efficiently managing their data throughout its lifecycle is a fundamental activity to enable the science they produce. These management activities range from defining and maintaining a data policy, enabling data discovery and access, up to data archiving and preservation. The new Neutron Beam Line Center at the planned MPR facility is planning the development of a whole new software stack for neutron science (Marais, 2022). With the South Africa draft policy on Open Science this must now require the creation of FAIR datasets. The present Safari-1 team are collaborating with the ISIS Neutron Source at the Rutherford Appleton Laboratory (ISIS, 2022) and with other European neutron sources in the EU BrightnESS[2] project to bring together the neutron ecosystem for sustainable science[9]. It therefore seems useful to briefly describe how teams in ISIS and in the Scientific Computing Department at the laboratory develop, maintain and run the data management services for the ISIS data catalog (SCD, 2022).

The data management processes implement the ISIS open data policy (ISIS Data Policy, 2022), which is reviewed yearly and updated to reflect any changes in the practices and/or wider open data policy constraints. The data associated with experiments is given a Digital Object Identifier (DOI), which enables data citation[10]. Datasets landing pages are made available *via* a user-friendly interface. The DataGateway interface[11] provides access to the ISIS data catalog and enables browsing, searching and retrieving embargoed data to its owners and all of the open data to anyone in the world.

The backend system is composed of modular components of the ICAT ecosystem (ICAT Project, 2022). ICAT is an open-source collaborative project across multiple scientific facilities that need this type of data management. The ICAT project revolves around a metadata catalog component, a data retrieval module, a user-friendly interface and other components enabling searches, DOI creation, and so forth. Data may be stored and archived on disks or on tape, depending on the volume and access patterns. In the case of ISIS, the data is kept on disks. Other facilities that are part of the ICAT collaboration are the Diamond Light Source, the European Synchrotron Radiation Facility (ESRF), the ILL Neutron source, the Helmholtz-Zentrum Berlin für Materialien und Energie (HZB), the ALBA Synchrotron Light Source, and the Scientific Computing Department at RAL. These facilities share code, best practices and experiences *via* the ICAT collaboration, as well as collaborating *via* other European projects such as the ExPANDS (2022) and PaNOSC (2022) projects.

In addition to including the Datacite required metadata, the ISIS dataset landing pages have been marked up using the Schema.org mark-up vocabulary, according to Google's guidelines on structured data. The open datasets from the ISIS Neutron source are therefore available using Google's Dataset Search tool. Furthermore, in the ExPANDS project, the ICAT collaboration is planning extensions to the Schema.org vocabulary to better support FAIR data, in a similar fashion to the extensions proposed by the Bioschemas community described above. For photon and neutron datasets, these proposed changes involve adopting the PaNET ontology (NCBO

9 https://brightness.esss.se/ (accessed October 2022).

10 https://www.isis.stfc.ac.uk/Pages/Digital-Object-Identifiers-(DOIs)-for-ISIS-Data.aspx (accessed June 2022).

11 https://data.isis.stfc.ac.uk/ (accessed June 2022).

BioPortal, 2022). This provides a taxonomy and thesaurus of photon and neutron (PaN) experimental techniques, based mainly on accelerator-based light sources and neutron facilities. The ontology defines specific techniques in terms of more general technique classes and provides synonyms and references. The goal of using this ontology is to enhance the FAIRness of photon and neutron data catalog services.

The proposed Neutron Beam Line Center at the MPR will make use of similar technologies to those described above to ensure that experimental data generated at the MPR is compliant with the South Africa's draft Open Science policy.

Thoughts for the future

To support the SKA Science Data Processor and an SKA Regional Center in Africa will require the assembly of an interdisciplinary team consisting of research librarians, data scientists and research software engineers who are collectively skilled in complementary aspects of research data management. While the need for data scientists and software engineers is self-evident it is worth discussing the changing role of librarians in a time when almost all content is born digital.

The LIBER Consortium of research libraries (LIBER Europe, 2022) sees a role for libraries in four key areas of research infrastructure:

- Shared services and Cloud services.
- Semantic interoperability, open and linked data.
- Data stewardship.
- Disciplinary partnerships.

By developing such skills in research data management, libraries can continue to play a central role in supporting first-class research not only at universities but also at national and international facilities. However, this move will require librarians who are qualified to act as domain-specific data stewards to work collaboratively with research scientists, as well as with software engineers and data engineers, to support the emerging research data infrastructure. Research librarians can then play a central role in supporting scientists in coping with the requirements of FAIR data with actionable metadata and semantics.

In terms of the required skills in research software engineering in Africa, there are a number of activities already underway including:

- Research Software & Systems Engineers of Africa[12].
- Research Software Engineering (RSE) Group at Stellenbosch University, South Africa[13].

These communities are also liaising with the UK Society of Research Software Engineering[14] and the UK Software Sustainability Institute (Software Sustainability Institute, 2022). The latter has contributed to the provision of training activities including the software, data and library carpentries[15].

The ambitious objective of the European Open Science Cloud (EOSC) (European Commission, 2020) is to provide researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes. In addition, EOSC has launched the FAIR4S project[16] to help organizations identify the capabilities and skills required to implement the FAIR principles. EOSC ultimately aims to develop a "Web of FAIR Data and services" for science in Europe upon which a wide range of value-added services can be built.

An initiative to create a Global Open Science Cloud is being promoted by the CODATA organization (Global Open Science Cloud, 2022). CODATA is also supporting the development of a strategy and vision for an African Open Science Platform (AOSP) (African Open Science Platform, 2022):

> *"The Platform's mission is to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society".*

The SKA-Mid project will provide a unique opportunity to make the AOSP vision a reality.

Data availability statement

No new data were generated or analysed during this study.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This work was partially supported by the Facilities Funding from Science and Technology Facilities Council (STFC) of UKRI, and Wave 1 of the UKRI Strategic Priorities Fund under the EPSRC grant EP/T001569/1, particularly the 'AI for Science' theme within that grant, by the Alan Turing Institute.

12 https://rsse.africa/ (accessed June 2022).

13 https://rse-at-sun.github.io/RSE-at-SUN/ (accessed June 2022).

14 https://society-rse.org/ (accessed June 2022).

15 https://carpentries.org/ (accessed June 2022).

16 https://www.eoscpilot.eu/sites/default/files/fair4s_eoscpilot_skills_framework.pdf (accessed June 2022).

Acknowledgments

TH would like to acknowledge the invaluable assistance he has received from Alejandra Gonzalez Beltran, especially with section 5 on the ISIS data management pipeline and with her input on the DCAT and schema.org vocabularies. She also contributed to discussions around the required roles for FAIR and open data management, including software engineers and data stewards, and data stewards specialised by domain. TH also wishes to thank Anna Scaife, Rachel Ainsworth and Lourdes Verdes-Montenegro from the SKA project; Carole Goble and Alasdair Gray from Bioschemas; Deon Marais from NECSA; and Elizabeth Newbold, Tracy Colborne and Catherine Jones from the Rutherford Appleton Laboratory, all for helpful input. He also thanks Kim Martin of Stellenbosch University for information about the RSE activities in South Africa.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

African Open Science Platform (2022). CODATA, The Committee on Data for Science and Technology. Available online at: https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/african-open-science/ (accessed July, 2022)

Benjelloun, O., Chen, S., and Noy, N. (2020). "Google dataset search by the numbers," in *International Semantic Web Conference,* (Springer, Cham), 667–682. doi: 10.1007/978-3-030-62466-8_41

Borgman, C. L. (2015). Big Data, Little Data, No Data: Scholarship in the Networked World. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9963.001.0001

Chrysostomou, A. (2019). SDP and SRCs. Available online at: https://astronomers.skatelescope.org/wp-content/uploads/2019/05/SDP-and-SRCs.pdf (accessed June, 2022).

DCAT (2022). Data Catalog Vocabulary – Version 3 (w3.org). Available online at: https://www.w3.org/TR/vocab-dcat-3/#abstract (accessed June, 2022).

DCC (2022). Data Management Plans. Available online at: https://dcc.ac.uk/resources/data-management-plans (accessed June, 2022).

European Commission (2020). European Open Science Cloud (EOSC). Available online at: https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en (accessed June, 2022).

ExPANDS (2022). The ExPANDS project:. Available online at: https://expands.eu/ (accessed June, 2022).

FAIR (2022). Principles - GO FAIR (go-fair.org). Available online at: https://www.go-fair.org/fair-principles/ (accessed September, 2022).

Garrido, J., Darriba, L., Sánchez-Expósito, S., Parra-Royón, M., Moldón, J., Mendoza, M. Á., et al. (2021). Toward a Spanish SKA Regional Centre fully engaged with open science. *J. Astron. Telesc. Instr. Syst.* 8, 011004. doi: 10.1117/1.JATIS.8.1.011004

Global Open Science Cloud (2022). CODATA, The Committee on Data for Science and Technology. Available online at: https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/ (accessed July, 2022)

Gray, A., Castro, L., Juty, N., Goble, C. (in press) "Schema.org for Scientific Data," in "*AI for Science*" eds A. Choudhary, G. Fox, and T. Hey (Singapore: World Scientific Press).

Hey, T., Tansley, S., and Tolle, K. M. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research.

Holdren, J. (2013). Microsoft Word – Public Access Memo 022113 FINAL. Available online at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (accessed June, 2022).

ICAT Project (2022). The ICAT Project. Available online at: https://icatproject.org/ (accessed June, 2022).

ISIS (2022). Neutron and Muon Source. Available online at: https://www.isis.stfc.ac.uk/Pages/home.aspx (accessed June, 2022).

ISIS Data Policy. (2022). Available online at: https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx (accessed June 3, 2022)

LIBER Europe (2022). LIBER Consortium of research libraries: Home. Available online at: https://libereurope.eu/ (accessed June, 2022).

Marais, D. (2022). Presentation at NOBUGS 2022 (19-22 September 2022): Current state of data analysis at the Necsa Neutron Diffaction Facility · Indico. Available online at: https://indico.psi.ch/event/12738/contributions/38886/ (accessed September, 2022).

NCBO BioPortal (2022). PaN Experimental technique - Summary (bioontology.org). Available online at: https://bioportal.bioontology.org/ontologies/PANET (accessed June, 2022).

Necsa (2022). Cabinet approval of multipurpose reactor ensures South Africa retains its Nucear technology global footprint. Available online at: https://www.necsa.co.za/cabinet-approval-of-multipurpose-reactor-ensures-south-africa-retains-its-nucear-technology-global-footprint/ (accessed June, 2022).

NSF (2020). Publications Output: U.S. Trends and International Comparisons | NSF - National Science Foundation. Available online at: https://ncses.nsf.gov/pubs/nsb20206/executive-summary (accessed October, 2022)

PaNOSC (2022). The PaNOSC project: The Photon and Neutron Open Science Cloud (PaNOSC). Available online at: https://www.panosc.eu/ (accessed June, 2022).

Pienaar, H. (2022). *Draft National Open Science Policy*. Department of Science and Innovation, Republic of South Africa.

Ratcliffe, S. (2022). SKA a game changer for African tech. Available online at: https://businesstech.co.za/news/business-opinion/45930/ska-a-game-changer-for-african-tech/ (accessed October, 2022)

Scaife, A. (in press). Chapter in "AI for Science', edited by Alok Choudhary, Geoffrey Fox, and Tony Hey, to be published by World Scientific Press, Singapore (2023).

SCD (2022). About Us. Available online at: https://www.scd.stfc.ac.uk/Pages/About-Us.aspx (accessed June, 2022).

Scopus (2020). Scholix analysis (OpenAire Scholexplorer) matched with All articles.

Sejnowski, T. J. (2018). *The Deep Learning Revolution*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11474.001.0001

SKA South Africa (2022). About SKA South Africa. Available online at: https://southafrica.skatelescope.org/ (accessed June, 2022).

Software Sustainability Institute (2022). The Software Sustainability Institute. Available online at: https://www.software.ac.uk/ (accessed June, 2022).

Stevens, R., Taylor, V., Nichols, J., Maccabe, A. B., Yelick, K., and Brown, D. (2020). *AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science (No. ANL-20/17)*. Argonne, IL: Argonne National Lab.(ANL). Available online at: https://www.anl.gov/ai-for-science-report (accessed June, 2022). doi: 10.2172/1604756

The SKA Project (2022). Available online at: https://www.skatelescope.org/the-ska-project/ (accessed June, 2022).

UKRI (2022a). UKRI announces new Open Access Policy. Available online at: https://www.ukri.org/news/ukri-announces-new-open-access-policy/ (accessed June, 2022).

UKRI (2022b). Making your research data open. Available online at: https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/ (accessed June, 2022).

Worldwide LHC Computing Grid (2022). Available online at: https://en.wikipedia.org/wiki/Worldwide_LHC_Computing_Grid (accessed June, 2022).

frontiers | Frontiers in Research Metrics and Analytics

TYPE Review
PUBLISHED 02 December 2022
DOI 10.3389/frma.2022.975109

OPEN ACCESS

EDITED BY
Katie Wilson,
Curtin University, Australia

REVIEWED BY
Rania Mohamed Hassan Baleela,
University of Khartoum, Sudan
Dale Peters,
University of Cape Town, South Africa
E. Tobias Krause,
Friedrich-Loeffler-Institute, Germany

*CORRESPONDENCE
Michael S. Pepper
michael.pepper@up.ac.za

SPECIALTY SECTION
This article was submitted to Scholarly Communication, a section of the journal Frontiers in Research Metrics and Analytics

RECEIVED 21 June 2022
ACCEPTED 15 November 2022
PUBLISHED 02 December 2022

CITATION
Strydom A, Mellet J, Van Rensburg J, Viljoen I, Athanasiadis A and Pepper MS (2022) Open access and its potential impact on public health – A South African perspective.
Front. Res. Metr. Anal. 7:975109.
doi: 10.3389/frma.2022.975109

Open access and its potential impact on public health – A South African perspective

Adéle Strydom, Juanita Mellet, Jeanne Van Rensburg, Ignatius Viljoen, Anastasios Athanasiadis and Michael S. Pepper*

SAMRC Extramural Unit for Stem Cell Research and Therapy, Department of Immunology, Faculty of Health Sciences, Institute for Cellular and Molecular Medicine, University of Pretoria, Pretoria, South Africa

Traditionally, access to research information has been restricted through journal subscriptions. This means that research entities and individuals who were unable to afford subscription costs did not have access to journal articles. There has however been a progressive shift toward electronic access to journal publications and subsequently growth in the number of journals available globally. In the context of electronic journals, both open access and restricted access options exist. While the latter option is comparable to traditional, subscription-based paper journals, open access journal publications follow an "open science" publishing model allowing scholarly communications and outputs to be publicly available online at no cost to the reader. However, for readers to enjoy open access, publication costs are shifted elsewhere, typically onto academic institutions and authors. SARS-CoV-2, and the resulting COVID-19 pandemic have highlighted the benefits of open science through accelerated research and unprecedented levels of collaboration and data sharing. South Africa is one of the leading open access countries on the African continent. This paper focuses on open access in the South African higher education research context with an emphasis on our Institution and our own experiences. It also addresses the financial implications of open access and provides possible solutions for reducing the cost of publication for researchers and their institutions. Privacy in open access and the role of the Protection of Personal Information Act (POPIA) in medical research and secondary use of data in South Africa will also be discussed.

KEYWORDS

open science, open access, publication costs, privacy, POPIA, secondary use of data

Introduction

Traditional subscription-based publishing models where individuals or institutions pay subscription fees in order to gain access to scientific material have been the modus operandi for many years. More recently, "Open Science," a global movement that aims to make the conduct and dissemination of research methods and results accessible to all, has been gaining traction. This is done in order to promote transparency and collaboration to the benefit of the global community (Besançon et al., 2021), and is applied through Open

Science practices that include open access, open source, open data, open methodology, and open peer-review. Open access strives to remove the financial and legal restrictions that can prevent individuals from accessing research publications and outputs (Prlic and Procter, 2012; Tennant et al., 2016; Muñoz-Tamayo et al., 2022; UNESCO, 2022a). Open access can also include early distribution of manuscripts in the form of preprints, in other words, draft articles that have not yet been peer-reviewed or published in scientific journals. Open source, open data and open methodology ensure that all research data and the various tools used to acquire and analyze the data are shared in an unrestricted manner, thereby promoting and facilitating the rapid replication of studies, increasing re-use of data, and assisting in the peer-review process. Open source, open data and open methodology should provide data publicly, and without cost and access restrictions (York, 2022). Open peer-review allows for the public sharing of peer-review reports and author responses in a transparent manner. This practice maintains a high quality of peer-review and reduces the risk of hidden conflicts of interest (Szekely et al., 2014).

The increase in online journals in the 1990s initiated the open science movement with the purpose of supporting transparency and collaboration in research and scientific communication (Huston et al., 2019). In public health, open science provided benefits such as opportunities for scientific collaboration and partnerships, increased research and analytical capacity, early detection of health and environmental threats, monitoring of real-time response, informed policy decisions, more capacity for public participation, transparency and better accountability. The COVID-19 pandemic highlighted the benefits of open science but also exposed the challenges related to the accuracy and validity of scientific information (Besançon et al., 2021). During the pandemic, researchers relied less on traditional systems of publishing and embraced open access platforms and preprint repositories to disseminate their COVID-19-related research results as quickly as possible. A preprint is a version of a scientific manuscript posted on a public server prior to formal peer review. Even though preprints may contain errors and potentially increase the risk of disseminating misinformation, they provide an open and transparent publication mechanism, thereby accelerating communication between scientists.

Although COVID-19 exposed the need for open science, open access to scientific knowledge is still a dilemma for many scientists, especially in resource-constrained countries. Scientific studies can consequently become locked behind subscription paywalls thereby blocking those lacking appropriate journal subscriptions or financial resources from obtaining access to research material (International Science Council, 2022). Except for diamond publishing where authors do not pay for open access publishing, the typical gold open access model affords access to publications but transfers the responsibility of payment from individuals to academics/authors and their institutions. While it is recognized that open access to peer-reviewed publications is critical for scientific advancement and affords readers unrestricted access to information, payment is still required to cover costs relating to editing, typesetting, printing, binding, marketing, distribution and archiving (The Conversation Africa, 2022b).

Despite inadequate funding and limited research capacity, African scientists have made valuable but limited contributions to COVID-19 research. Two studies concluded that only 3% of all COVID-19-related articles (not including preprints) were authored by African scientists and just 4.3% contained information specific to Africa and/or African countries (Kana et al., 2021; Naidoo et al., 2021). South Africa, Egypt and Nigeria collectively provided 65% of the COVID-19 articles produced by Africans. According to Naidoo et al. (2021), one in five African COVID-19 papers had no African authors, and approximately 66% of authors on papers with research originating from African populations were non-African nationals. Both these studies highlight the need to boost the research production by African scientists and the need to support the publication of their research findings.

South Africa is considered a pioneer in Africa regarding open access policies and the structures that enforce them (UNESCO, 2017). As of February 2022, South Africa had indexed more than 100 South African-based open access journals in the DOAJ (2022) (Directory of Open Access Journals). Open data and the sharing of health data for research should nevertheless be subjected to legal and ethical procedures (Staunton et al., 2021), especially for secondary use of such data. This has changed notably in South Africa since the implementation of the Protection of Personal Information Act (POPIA) No. 4 of 2013.

This paper will discuss the opportunities and challenges associated with open access to research in the South African higher education context. It will also address the financial implications of open access to academic institutions, amongst other role players, and provide possible solutions to reduce publication costs for researchers and their associated institutions. Since South Africa is classified as an upper middle-income country (UMIC; The World Bank, 2022) and considering the limited research resources available, this article will also address the publishing policies and journal selection processes that need to be considered when publishing open access. Privacy in open access and the role of the POPIA in medical research and secondary use of data, will also be discussed.

Open access – Opportunities and challenges

There are multiple journal ranking systems that serve as proxies for the quality of a given journal. The most utilized is the

impact factor (IF) which is the ratio of citations to the number of citable publications published by a given journal. It is calculated based on a two-year period where the number of annual citations is divided by the number of publications in the previous 2 years (Sharma et al., 2014).

Each article undergoes a rigorous peer-review process. Peer-review is a process in which experts in the field under consideration serve as a quality control checkpoint where scientific processes and claims are verified, rendering approved articles scientifically reliable and valid. The higher the quality of the journal, the more arduous the reviewing process (making it difficult to publish in highly revered journals). Credibility is often reflected by the number of citations a publication has received, which in turn, increases the journal's IF.

Dissemination of research information can be accelerated through preprints and expedited peer-review and publishing processes, as was observed during the COVID-19 pandemic.

Journals indexed in major databases such as Medline, Scopus and PubMed are considered to be of acceptable scientific quality. A journal's publishing history and scope are likewise considered important and are reflected in the journal's impact. These measures affect the relative importance of journals in a field and also reflect the journal's reach. When deviating from these well-established publication practices, publications may end up in predatory journals. "Predatory" journals and publishers make false claims and provide misleading information to manipulate authors into publishing therein. Several predatory journals have author-pays practices and should be avoided, since publishing in such journals may harm the reputation of the authors and their affiliated institutions, and provides a highly questionable way around the all-important peer-review process.

This section serves to explore themes associated with open access publications, preprints and predatory journals.

Open access publications

Medical journals such as the New England Journal of Medicine, the British Medical Journal and the Journal of the American Medical Association are published in both paper and electronic formats. These journals rely on association membership fees and journal subscriptions to cover publication costs. With the growth of digital media, publication cost and revenue models have changed, and new funding models have had to be developed. Publishing open access allows readers to have immediate access without the need for subscriptions through institutions. Various types of open access models exist, and includes gold, green, bronze, hybrid, and diamond. The majority of open access journals, with the exception of the latter, incur a cost to the authors and/or funders for peer-review and publication, referred to as article processing charges (APCs). Gold open access is when authors and funders pay for published articles to allow immediate access without any restrictions. With green open access, a self-archived version of the manuscript is made available through an open access repository or website (Piwowar et al., 2018). In the case of an embargo period (~6–12 months), readers are required to pay a fee to access these articles during that period. In 2018, the majority of open access articles were published as "bronze". Articles published under this open access model are free to read on publisher websites but do not have a formal license for reuse (Piwowar et al., 2018). Hybrid journals charge APCs in addition to subscription costs that allow readers to access the full contents of the journal (Piwowar et al., 2018). At the extremes of the spectrum, we have pure subscription journals and pure "pay for publication" journals. In-between, some subscription journals offer free "green open access" and paid "gold open access" options. A wide range of hybrid journals offer subscriptions, site licenses and pay-for-use to readers, with different paid open access options to authors.

In the South African context, publishing research in a journal accredited by the Department of Higher Education and Training (DHET) has benefits for both the researcher responsible for the article as well as the institution they belong to. Increased access and readership will likely result in improved citation that might lead to wider visibility for authors and their institutions (Shuai et al., 2012). The number of articles that researchers publish in reputable journals is taken into consideration by academic and private institutions when performance-related decisions are made. Additionally, in the greater scheme of science and information sharing, it is through publication that information circulates through to others in the field, expanding the pool of knowledge and ultimately advancing scientific progress. Additional benefit is derived by South African academic institutions when publishing research in accredited journals, as these publications receive a subsidy from the South African government through the DHET (Woodiwiss, 2012). South Africa's experience with open access publishing is not new and the issues about high publishing costs have been discussed by other authors (Czerniewicz and Goodier, 2014; Hartman and Wu, 2018; Bawa, 2020).

Currently, journal articles have a system of reviewer recruitment which usually does not benefit the reviewer. However, depending on the publishing authority, reviewers may benefit through the wavering of APCs for articles that they wish to publish in the future. In some instances, the reviewers' contributions may be uploaded onto a commercial website named Publons (2022). Publons is a free platform that ensures recognition for peer-review and editorial contributions. Publons generates a review record that can be used in CVs, job and funding applications, and performance evaluations.

Lastly, publishing institutions benefit through the sale of published articles, and monthly/yearly subscriptions from academic institutions from which the author and/or reviewer benefit as a consequence of exposure rather than financial gain.

Preprints

More than 80,000 COVID-19 and SARS-CoV-2-related preprints and peer-reviewed articles have been published since the emergence of this virus in December 2019 (Besançon et al., 2021). In 2020, preprints accounted for 17–30% of all COVID-19 research papers (Else, 2020). In South Africa, the first COVID-19 case was reported in March 2020 and the Network for Genomic Surveillance in South Africa (NGS-SA) was created to investigate the dynamics of the SARS-CoV-2 epidemic (Giandhari et al., 2021). Consequently, preprints on South African SARS-CoV-2 variants and the immune response have been posted since 2020 (Callaway, 2021; Giandhari et al., 2021). Overall, preprints have accelerated the pace of research, and public-health policies have been directly informed by them. The use of preprints may thus be beneficial, but their unfettered use also raises concerns (Horby, 2022). Although preprints enable quicker data sharing during a crisis and allow scientists to improve their work with informal feedback, it also opens the door to use of predatory journals (Watson, 2022) and the use of social media to disseminate preprint findings (Koerber, 2021). Unfortunately, rapid preprints potentially increase the risk of fraudulent, deceptive or poor quality research (Horby, 2022). This may result in premature and misguided claims and increased confusion, especially when the distinction between preprints and standard peer-reviewed articles is poorly communicated or misunderstood by the media and the public. Publications in predatory journals that are not peer-reviewed may endanger public health by publishing inaccurate and unvalidated findings (Watson, 2022). This may contribute to unreliable meta-analyses, or flawed research data and findings.

Preprint publications contributed important knowledge on COVID-19, despite possibly having methodological weaknesses that could limit the interpretation of the results and provide misleading or false claims that could greatly impact public health (Besançon et al., 2021). Despite the fact that some studies were retracted, their claims still contributed to the body of knowledge. As a case in point, when the retraction rate of COVID-19-related articles was compared to publications in related research fields, the authors concluded that the rate was approximately four retractions for every 10,000 papers. This may have been the result of researchers rushing to submit manuscripts for publication and the expedited peer-review and publication process of COVID-19 articles by some journals (Yeo-Teh and Tang, 2021). This poses a direct threat to public health and leads to wastage of scientific resources and public confusion (Besançon et al., 2021). While academic communities rapidly disputed false claims, public perceptions are influenced by the dissemination of preprint information in mainstream media articles (Brierley, 2021; Fraser et al., 2021). The World Health Organization (WHO) has been raising awareness about an "infodemic" as social media has amplified and exacerbated misinformation and uncertainty (Vraga et al., 2020). Promoting news and science literacy allows people to determine whether information about COVID-19, or any other disease, is accurate, and empowers them to take active control of their social media feeds and protect themselves and others.

The quality of research becomes questionable when peer review is absent. This has the potential to drive negative perceptions, particularly within clinical research, and may therefore impact the sector as a whole, and not just at an individual level (Kwon, 2020). Consequently, preprint servers, such as BioRxiv (2022) and MedRxiv (2022), now have enhanced screening procedures in place. Both bioRxiv and medRxiv screen articles in a two-tiered approach which firstly requires that in-house staff examine the manuscript before seeking expert opinion concerning scientific merit and validity. This is done in order to ensure that scientifically sound, original research is being placed in the public domain. Health professionals and principal investigators are primarily used to review submitted preprints in medRxiv and bioRxiv. While the former (in-house) screening requires more time to complete, the latter is typically finished within 2 days. Since papers in medRxiv may be more relevant to health, they are scrutinized more closely and therefore take longer to evaluate. Rather than determining research quality, vetting is primarily used to identify potentially harmful articles, including those that do not provide evidence-based conclusions and/or make statements viewed as contradictory without just cause (Kwon, 2020). Recently though, this vetting process has been extended to exclude computational modeling papers considered to be "speculative" in nature.

Therefore, to ensure good quality research through open science initiatives, this process must go hand in hand with, for example, full data sharing and the publication of study protocols approved by International Regulatory Boards or Research Ethics Committees prior to the initiation of clinical trials (Watson, 2022). Similarly, use in policy decisions and modeling must be undertaken with caution and be transparent, while preprint servers may need to put additional screening measures and procedures into place to block the distribution of poor-quality manuscripts. Disclaimers relating to the preprint status of articles could also be used to combat publishing of preprints by predatory journals.

Predatory journals

In March 2022, the InterAcademy Partnership (IAP) published a report entitled "Combatting Predatory Academic Journals and Conferences" (Inter Academy Partnership, 2022). One of the main aims of the report was to find ways to prevent and reduce the number of predatory journals and conferences. According to the report, there are over 15 500 predatory journals around the world with widespread predatory practices. Predatory practices such as phishing, being misleading

by using false identities or re-publishing papers already in legitimate journals without permission will continue to be fueled by the digitization of academic publishing, gold and author-pays open access publishing models, and research evaluation criteria that emphasize quantity over quality. Authors have succumbed to predatory journals who seek papers *via* email and social media, promising to publish open access articles rapidly and with minimum review, frequently for a charge (Vervoort et al., 2020). The "publish-or-perish" mindset in academia, combined with obstacles to publish encountered by researchers from low to middle income countries (LMICs) has resulted in a rapid increase in the number of predatory journals which are easy to publish in (Vervoort et al., 2020). This affects the visibility of credible scientific research, stifles scientific progress, and jeopardizes the reputations of authors and the institutions they represent.

According to a previous study, South African academics published 728 articles in only five predatory journals over the three and a half years covered by the study (De Jager et al., 2017). This may have been due to local researchers collectively legitimizing these predatory journals. In this study, the level of predatory publications was found to have financial implications. Using an estimated subsidy of ZAR100 000 per article as an example, it was estimated that a total of ZAR70 million might possibly have been paid for publications in journals that did not meet strict academic research quality criteria. A combination of the South African subsidy system, the involvement of South African academics on editorial panels and reviewer lists, promotional material, and a substantial amount of South African written articles in such publications, had to some extent legitimized these journals in the South African system.

Using open access policies, funders can encourage funded researchers to publish in credible journals that adhere to established open access principles (Shamseer et al., 2021). Open access is not understood by most researchers beyond making research free to read. Journals with deceptive or nefarious publishing operations might have gained from or taken advantage of the inexperience of some authors. Predatory journals do not always include licensing information for articles or provide information on reusing published research. Scientists who publish in predatory journals are likely to be violating their funders' open access policies (Shamseer et al., 2021).

APCs and publisher profits

Article processing charges

In contrast to subscription-based publication models whereby publication costs are absorbed by the reader or their institution, the cost of open access models to academics is substantial. As an example, Nature and Lancet respectively charge academics €9 500 and US$5 000 per open access paper (The Conversation Africa, 2022b). Both Nature and Lancet call this APCs while other journals refer to a "publication fee". In Africa, the publication costs to researchers are hugely burdensome as a consequence of currency exchange rates. To demonstrate this practically, at an approximated exchange rate of ZAR17,00 to €1 and ZAR16,00 to US$1, the APCs in South Africa would be the equivalent of paying ZAR161 500 and ZAR80 000, or 16.2 and 8.0% of a one million Rand budget, respectively. This is equivalent to a €9 500 cost on a €60 000 budget, or US$5 000 cost to a budget of US$62 500. Several LMICs have weaker currencies when compared to South Africa. This further demonstrates the potential cost implications for researchers in LMICs and the financial pressure that many LMIC universities and associated researchers face with regard to APCs.

The authors of this paper are all members of the Institute for Cellular and Molecular Medicine (ICMM) of the Department of Immunology in the Faculty of Health Science at the University of Pretoria, South Africa. The ICMM is a transdisciplinary, translational, highly collaborative entity that aims to understand and manage specific contributors to the disease burden in South Africa and Africa. Active research projects cover a wide range of disciplines and entities, with a particular focus on molecular, cell and computational biology, and the ethical, legal, and social implications of research in cell and gene therapy. The ICMM comprises senior researchers, post-doctoral scientists, and postgraduate students of medical, scientific, ethical, data sciences and legal disciplines. Research papers are prepared and submitted for publication to a wide spectrum of academic journals.

African research groups are uniquely positioned and able to more accurately perform, collaborate with, sustain and describe research endeavors in Africa (Kana et al., 2021). However, as alluded to previously, this comes with notable costs to researchers. To demonstrate this practically, APCs paid during the course of 2021 for 10 publications associated with authors representing the ICMM totaled nearly ZAR330 000. This was distributed across APCs that were charged in Swiss Francs, Euros, and US Dollars. As shown in Table 1, the exchange rates increased the relative Rand (ZAR) cost of the APCs anywhere between 8.6 and 17.7 times that of the foreign currency equivalent. The average APC per published article was ZAR32 803,94, with the total costs representing nearly 33% of a ZAR 1 million budget. While not reflected in Table 1, institutional contributions totaling R86 075,00 were received during the 2021 period for six of the 10 articles with contributions still pending for the remaining four publications. While institutional contributions may cover up to 50% of the APCs, they are not guaranteed and can also take several months to reflect in research accounts.

While reimbursements and financial "rewards" for publications are provided at some academic institutions, the financial resources to do so must come from somewhere. The University of Pretoria, as with most other academic institutions

TABLE 1 A summary of the APCs paid for 10 published articles during 2021.

	CHF ($N = 4$)	EUR ($N = 1$)	USD ($N = 5$)
Min cost per article	CHF 1 440.00	N/A	USD 1 867.50
Max cost per article	CHF 1 800.00	N/A	USD 2 950.00
Mean cost	CHF 1 710.00	N/A	USD 2 243.50
Total cost	CHF 6 840.00	EUR 3 582.25	USD 11 217.50
Min exchange rate (ZAR:1 FC)	ZAR 8.59	ZAR 16.12	ZAR 14.51
Max exchange rate (ZAR:1 FC)	ZAR 17.74	N/A	ZAR 15.54
Mean exchange rate (ZAR:1 FC)	ZAR 14.79	N/A	ZAR 15.10
Min cost in ZAR	ZAR 15 469.90	N/A	ZAR 28 279.16
Max cost in ZAR	ZAR 31 923.49	N/A	ZAR 45 853.03
Mean cost in ZAR	ZAR 25 132.55	N/A	ZAR 33 955.48
Total cost in ZAR	ZAR 100 530.18	ZAR 57 731.88	ZAR 169 777.38

ZAR, South African Rand; FC, foreign currency; CHF, Swiss Franc; EUR, Euro; USD, US Dollar.

in South Africa, receives a subsidy per publication in an accredited journal from the South African government on an annual basis through the DHET. This subsidy is allocated to the Institution, Faculties, Schools, Departments and ultimately to academics according to various payment structures determined by the institution itself. Funding to academics who raise the funding, do the research and publish their manuscripts is on an *ad hoc* basis and may only be sufficient to cover the cost of one or two manuscripts. These payment structures vary across academic institutions and do not appear to be consistently observed. They are also restricted according to the availability of funds. Internally, the University of Pretoria open access fund provides partial support for APCs for articles published in accredited open access journals when no alternative funding is available. This support is based on a set of criteria provided in guidelines for applications to the fund. These criteria and the proportion of the APCs to be refunded per article may be reviewed and revised. The fund is supported by a reserve set aside from the annual resource allocation provided to the Department of Library Services. The fund does not support APCs for hybrid (open choice) journals and excludes monographs, book chapters and publications. Since support is dependent on the availability of funds, payment is not guaranteed. As such, the university urges researchers to request a waiver of APCs from target journals and to request APC support from the library services before submitting articles to journals.

While we have some appreciation of APCs relative to South African institutions, the true cost of open access for researchers based in LMICs is largely unknown (The Conversation Africa, 2022b). An urgent discussion is therefore needed on the financial implications of open access publishing for academic and research institutions in Africa. The rising cost of journal subscriptions is leading to an increasing number of questions concerning the academic publishing establishment (Van Noorden, 2013), and while some journals consider waiving the costs for 47 historically disadvantaged academic institutions in LMICs, research institutions from the remaining 58 nations are expected to cover all or reduced publication costs (The Conversation Africa, 2022b). This potentially creates a conundrum for researchers when applying for research grants to cover the costs of open access publications. Since grant funding is limited, researchers must carefully consider what they wish to publish and how they wish to publish prior to commencing the research; there is usually little room to adjust this plan once grant funding is approved. Using the costs incurred by the ICMM for 10 open access articles published in 2021 as an example, it is clearly not feasible to request support for open access publications through grant applications valued at less than R100 000 when the average cost of a single article in a medical sciences discipline would likely cost 30–50% of the total value of such a grant.

Publisher profits

While the philosophy behind open access requires that authors retain their copyright, in practice researchers and their institutions are required to assign copyright to the publisher in addition to paying APCs to publishers who generate profits through this process. Peer-review is typically also done without compensation. This has created an "asymmetric business model" (International Science Council, 2022). What contributes to the high cost of open access publishing? Commercial, profit-driven publishing houses sell journal subscriptions and site licenses or charge pay-per-use fees. They also apply embargoes that may range from 6 months to potentially indefinite periods of time where they own the copyright. All of these journals claim to be "open access". While "green open access" options are seldom provided, journals require payment of APCs to give

readers "gold open access". This has made academic publishing a lucrative business. As a case in point, it was reported that the revenue generated by the scientific publishing industry in 2011 totaled US$9.4 billion. This included the roughly US$5 000 generated per article of the nearly 1.8 million English-language articles published at that time (Van Noorden, 2013). The total revenue generated by the academic publishing industry increased to more than US$19 billion in 2017 (Hagve, 2020).

Over 50% of the publication industry is dominated by five publishers. These include Elsevier, Black and Wiley, Taylor and Francis, Springer Nature, and SAGE. Approximately 16% of the market is held by Elsevier, which publishes more than 3 000 academic journals. Unlike companies such as Microsoft, Google, and Coca Cola, Elsevier boasts a profit margin approaching 40% and the curve indicates an upward trend (Hagve, 2020). Between 1991 and 2013, profit margins for all of Reed Elsevier's enterprises ranged between an estimated 14 and 27% (Larivière et al., 2015), which aligns with the average estimated 20–30% profit margins currently noted for other publishers. However, when focusing on their "Scientific, Technical and Medical division," profit margins across the 1991–2013 period reportedly ranged between an estimated 30 and 42% while operating profits at the same time-period ranged between approximately 9 (1991) and 43% (2013). Operating profits were noted to have a strong upward trend between 1991 and 2013 (Larivière et al., 2015).

To place this into context from a South African perspective, at an approximate exchange rate of ZAR20.00 to £1.00, Elsevier's recently reported revenues of £2.64 billion and net profit of £1.922 billion (72.8%) equates to a staggering revenue of ZAR52.8 billion, and net profit of ZAR38.44 billion. This represents roughly 95% of the ZAR40.4 billion and 15% of the ZAR259 billion budget respectively allocated toward provincial hospital services and the entire public health sector for the 2022 financial period by the South African Treasury [National Treasury (RSA), 2022]. Representing the South African public hospital industry, the provincial hospital services sector provides hospital care to approximately 80% of the South African population (South African Government, 2022). Given that the South African population comprises just over 60 million people (Statistics South Africa, 2022), this implies that the provincial hospital services sector must serve just over 48 million people with a budget that is a mere 5% more than recently reported net profits reported by Elsevier.

As a consequence of high profit margins, there has been a proliferation of publishing houses and journals and the capacity for academic journals to turn the situation of production costs on its head. For example, a traditional newspaper, whose profit tends to be 10–15%, incurs expenses for wages for its journalists, editors, and graphic artists, as well as expenses for research, printing, and distribution, all paid through sales and advertising. In the case of academic journals, production cost are paid for by research funds, researcher salaries, and the costs involved in undertaking research. However, academic editors receive symbolic pay, as quality control and fact-checking are done through peer-review, which is voluntary. Most of the access is digital, and therefore the only real cost incurred by the publisher is for graphic design of the article (Hagve, 2020).

For publishing houses, open access has provided a new way to generate a profit. However, it comes at a high cost to authors, with the price of publication often ranging from US$1 500–US$3 000 in a fully open access journal, and up to US$6 000 in traditional subscription journals. Although open access fees are transparent, revenue may also be generated through other means (Van Noorden, 2013). These include revenue generated through membership or subscription fees and subsidies that may be received, notably by smaller publishers. Subscription-based journals may additionally derive their revenue from cross-subsidies, by offering advertising opportunities, and charging fees for reprinting of articles. One reason for the lower costs of purely open access publishers is that they are providing a digital product from a business model that is less established than the traditional, subscription- and paper-based journals. Unlike their more established counterparts, this creates flexibility regarding presentation of their articles and subsequent reduction in production costs.

Ultimately, the success of publishing houses is dependent on how well their products sell. This is in turn dependent on the quality of their products. The quality of academic journals is measured through an IF, with a high IF being important for financial success. As indicated previously, the journal IF is based on the number of citations a journal's articles receives in a given period of time, with frequent citations increasing the journal's perceived importance and value (Hagve, 2020). The 'exclusivity' of journals, as measured by their rejection rate, has also provided grounds for increased publication fees. The rejection process may prompt authors to consider alternative journals. Since journals perceived as being of greater value will naturally receive a greater number of submissions, publishers have argued that this is essential for researchers whose task is to sort through millions of published articles each year to determine which are worth reading.

University libraries and their relation to APCs

The South African National Library and Information Consortium (SANLiC) is a non-profit organization that facilitates the process involved in obtaining licensing agreements for electronically accessed information. Members of SANLiC notably include public higher education and research institutions (SANLiC, 2021). Furthermore, SANLiC is committed to promoting open access for South African research outputs by increasing access to scholarly information, reducing the cost of library subscriptions, and looking for alternative

forums for the distribution of South African scholarships. SANLiC has been successful in lowering the cost of subscription access for member libraries resulting in the expansion of their collections. In 2019, an overall 87.4% cost avoidance on subscription list prices was negotiated. Unfortunately, according to the organization, this is not enough to address the unsustainability of the entire pay-to-read model. The main annual expenditure on scholarly literature by South African higher education and research institutions is allocated to pay-to-read subscriptions and approximately 80% of these subscriptions are based on deals negotiated by SANLiC. In 2020, SANLiC spent US$27 130 138 (82% of their journal expenditure) on Big Five journal packages (Elsevier's Science Direct, Wiley, Springer Nature, Taylor and Francis, and SAGE). Only 52% of South African research output, i.e., research with South African corresponding authors, is published in journals covered by these packages. SANLiC did a data analysis in 2020 on research and review articles published between 2014 and 2019, and of the 62 549 publications assessed, approximately 33% were open access, while the remainder were behind paywalls. Recently, SANLiC signed a number of Read and Publish Agreements on behalf of South African higher education institutions. The list of current negotiated agreements is now available, thus facilitating publication in gold open access journals.

The total amount that South Africa spends on journal subscriptions is unknown. This is largely due to the fact that university libraries, research departments, and research institutions each have their own budgets for this. Additionally, there is a lack of transparency regarding publisher fees owing to nondisclosure agreements signed by research institutions (Mail and Guardian, 2022; The Conversation Africa, 2022a). It is important for academics to publish their work, not only to advance their careers, but also to increase research citations and visibility. This is a long-running issue between researchers and publishers, as journals make their profit from research while restricting access to it (Mail and Guardian, 2022). According to Elsevier, the embargo period in green open access journals is justifiable on the basis that it is (1) not uncommon practice to have embargo periods of 12–14 months for publications, and (2) that the publisher requires revenue from subscriptions to compensate for the publishing costs (Mail and Guardian, 2022). While implementation of embargo periods is not new, Elsevier's updated regulations regarding embargoes resulted in a petition being launched against it. Signatories of this petition not only include SANLiC and SANLiC affiliates, but also non-SANLiC members (Mail and Guardian, 2022). The signing of this petition was largely driven by the principles governed by open science, notably in the form of open access.

According to a 2015 White Paper published by the Max Planck Digital Library, it has been suggested that scientific subscription journals should alter their business models to adopt open access business models instead (The Conversation Africa, 2022b). It has also been suggested that such changes should be reflected in how countries challenge the publisher costs through amendments to their legal and financial structures (Schimmer et al., 2015). This may not be easy to implement in countries that like South Africa have little published data on fees charged by publishers, how much is actually spent on various publishing fees, or what discounts, waivers and/or rebates are granted by publishers. Nevertheless, through a 2018 survey to which 15 of the 26 South African public university libraries provided a response, it was found that more than ZAR1 billion (US$68 020 593) was paid toward fees for e-resources, book budgets, and copyright licenses (The Conversation Africa, 2022a). As a result of the increasingly unfavorable foreign exchange rate, it has been speculated that this amount may increase by about 5% per annum. Additionally, 14 of the 15 institutions pay roughly ZAR31 million (US$2 106 307.37) for copyright licenses on prescribed works. While limited, the noted expenditure for research and teaching purposes should be a major concern, especially when considering that an estimated 80% of literature purchased for use in academic libraries is produced by international publishers. Furthermore, a great portion of locally produced research is made visible through the publication of work using international publishers (The Conversation Africa, 2022b). More financial data is however required before the combined efforts aimed at impacting these costs can be experienced by researchers and associated institutions in LMICs.

In order to fully benefit from the principles that govern open access, university libraries in Africa have actively promoted the open access movement in a variety of ways. This has not only been seen within the academic research sector through the establishment and maintenance of institutional repositories (IRs) but has also included the cataloging of journals that facilitate and promote open access publications by University library services (Research Gate, 2022). Assuming that they are properly maintained, IRs therefore have the capacity to increase the visibility of research activities and outputs achieved by academic institutions. While it is noted that this information can change rapidly owing to daily revision of IR data, according to OpenDOAR, eight African countries currently have IRs. These include South Africa, Kenya, Nigeria, Algeria, Tanzania, Zimbabwe, Sudan and Ghana. IRs total 48, 44, 31, 20, 14, 11, 12, and 6 per country, respectively (OpenDOAR, 2022). The University of Pretoria hosts both a Research Data Repository (Figshare, 2022) and an Institutional Repository (UP Space, 2022). Both are operated by the Department of Library Services.

The standard means of accessing journals at present is through academic library subscriptions and private purchases. Due to paywalls that continue to frustrate access to journals and articles, multiple alternative options have been developed to provide free access. Sci-Hub is a controversial website that has emerged as a consequence. Sci-Hub (2022) provides mass public access to research papers located behind paywalls thus enabling the free sharing of information. However, this platform is illegal as it allows copyright infringement. A non-controversial version of Sci-Hub, named Unpaywall (2022),

is a legal, commonly utilized browser extension of Google Chrome and Firefox that provides access to a repository of freely available scientific articles. When encountering an academic article online, a pop-up will appear providing the option to download the article for free if the article is available on Unpaywall.

Institution-specific policies on open access and publishing

In October 2021, the University of Pretoria approved a Policy on Open Access to Research Papers and Creative Outputs Authored by University of Pretoria Researchers. While the policy itself is not publicly available, the purpose of the policy can be summarized as follows: (1) "to support several Open Access initiatives, including the Berlin Declaration"; (2) "to ensure that research conducted at the University is conducted to the highest possible standard, is made freely available to increase local and global visibility, and facilitates greater research impact and benefit to all stakeholders"; (3) "to honor ethical research and business standards, contractual obligations, legal restrictions, archiving requirements of research funders, and publishers' copyright regulations"; (4) "to support global initiatives to influence the current copyright practices of publishers and authors thereby expanding the rights of its authors and researchers"; and (5) "to provide directives for the archiving and dissemination of academic journal articles, conference papers and creative outputs authored or co-authored by University of Pretoria researchers which have been or will be published".

The aim of this policy is to ensure that all published University of Pretoria research and creative outputs are available for use within the University and are freely accessible to any other student, researcher, or member of the public with a non-commercial requirement for access to the information. The policy applies to all postgraduate students, research staff, employees, visiting researchers, and postdoctoral fellows engaged in publishing and/or disseminating research outputs under the auspices of the University, even when they co-author with researchers from other institutions. Other sections in the policy include a policy statement, definitions, associated documents, roles and responsibilities of authors, Deans of Faculties, and the Department of Library Services, and describe where it is not applicable as well as the consequences of non-compliance. The policy is reviewed every 3 years.

In July 2014, the University approved their Policy on Open Access Publishing Processing Charges (UP Policies, 2022). The purpose of this policy is to facilitate open access publishing of research by students and staff at the University. It provides the principles for support of open access publishing by researchers and the criteria for funding of APCs through an Open Access Fund. Support is provided for articles to be published in peer-reviewed, international open access journals. A list of eligible open access journals, together with their IFs, is provided *via* restricted access by the Department of Library Services. This policy is also reviewed every 3 years. The current policy was reviewed in 2018 and remains unchanged.

In keeping with the objectives of the indicated policies, the Department of Library Services, as part of SANLiC signed transformative (read-and-publish) agreements in March 2022 with the following publishers: Wiley, Emerald, and Association for Computing Machinery (ACM; UP news, 2022). The main benefit is that publications in journals from these publishers are not subjected to APCs. The Department of Library Services has further expanded its services to support researchers to publish open access articles for free in hybrid journals from these publishers and has compiled a list of accredited journals which are part of these agreements. The list will constantly be updated as negotiations with publishers on transformative agreements are ongoing. In the meantime, authors can start submitting their manuscripts to Wiley and Emerald. Subscription to the Wiley hybrid open access journals provides read access and enables eligible corresponding University of Pretoria authors to publish articles at no extra charge. However, publishing in fully open access journals with Wiley is not free and may require authors to pay publishing fees. Regarding Emerald hybrid journals, research can be published through prepaid open access publishing vouchers if journals are eligible. All ACM open access journals publish articles for free.

Privacy, POPIA, and research ethics

Privacy and the impact of POPIA on open data sharing

Governments have recognized the value of open science and open access, particularly as they pertain to biological samples and their associated data. South Africa is no exception, with the South African Department of Science and Innovation recognizing its pertinence to genomic research and the Fourth Industrial Revolution (Staunton et al., 2019). For example, the sharing of genomic data has several benefits. These include ensuring the optimal use of resources such as facilitating studies that require larger sample sizes to ensure that they are statistically more powered, thereby facilitating reproducible research, creating new research opportunities from pre-existing data sets, and promoting research innovation. Notably, open data initiatives have often been considered important to ensure that a replication crisis does not occur, even if sharing raw data may be difficult due to compliance with data protection regulations in medical research. Despite this difficulty, they are often seen as an essential component of the peer-review system. Except where sharing of data is prohibited for privacy

reasons, the drive for data sharing has been optimized by the increasing number of policy documents developed to facilitate this process (Besançon et al., 2021). However, data sharing must be governed according to ethical standards that ensure no risk of harm to participants and promote public trust in such endeavors (Staunton et al., 2019). In South Africa, this governance framework was recently established by the gazetting and enforcement of the POPIA (no. 4 of 2013). Questions have consequently been raised by researchers about the impact of POPIA on informed consent, standards of anonymization, secondary use of data, and privacy in open access.

A strict interpretation of section 13(1) of POPIA suggests that it is only permissible to request specific consent from participants. Under this interpretation, unless consent for sharing of specific data and/or samples is obtained at the onset of the study, it is not possible for researchers to engage in data/sample sharing practices (Staunton et al., 2019). Practically, this is considered inefficient and potentially wasteful. As such, POPIA permits the secondary use of personal information within the context of research without the need to obtain further consent if "the research is necessary to prevent or mitigate a serious and imminent threat to public health or safety" or "if the personal information will only be used for research AND it will not be published in an identifiable form". As broad consent is a tool through which sample and data sharing has frequently been made possible, phrasing used within POPIA regarding informed consent has initiated a debate among researchers regarding the legality of broad consent (Staunton et al., 2019, 2020). The National Department of Health (DoH) Ethics in Health Research Guidelines defines broad consent as follows: "the donor permits use of the specimen for current research, for storage and possible future research purposes, even though the precise nature of future research may be unclear at present" (National Department of Health, 2022). In contrast to the strict interpretation of POPIA section 13(1), Thaldar and Townsend (2020) have posited that section 15 of POPIA makes provision for further research without obtaining new consent if the personal information collected previously or elsewhere is to be used for a specific purpose. Based on their interpretation, once specific consent has been obtained, researchers may continue to conduct their ongoing research under the provisions set forth in POPIA section 15 (Thaldar and Townsend, 2020). While there are enforceable conditions such as data security governance and participant risk of harm linked to this interpretation, from a practical perspective, this latter perspective essentially considers the receipt of specific consent as grounds for extended research privileges and some degree of broad consent.

Regardless of the argument made, until judicial case studies become available, the practical application of such clauses within POPIA remain debatable and open to interpretation. This is because POPIA-driven processing of personal information is principle-based, rather than sector-specific (Staunton et al., 2021). This has consequentially resulted in uncertainty regarding the appropriate application of POPIA in relation to health information for research purposes. In order to provide clarity in the healthcare/medical research sectors, the Academy of Science of South Africa (2022) and several of its stakeholders commissioned the development of a Code of Conduct (Staunton et al., 2020, 2021). This Code aims to compliment POPIA and to provide sector-specific guidance on its interpretation and application. In so-doing, the Code aims to clearly communicate the expectations placed upon researchers when working with health information or engaging in health-related research. The final draft of the Code of Conduct is currently being finalized (Academy of Science of South Africa, 2022).

Research ethics

In addition to fulfilling POPIA requirements, scientists are equally bound by research ethics. This is important given that researchers are increasingly applying open science principles and making anonymized data available for analysis *via* publicly accessible repositories (Besançon et al., 2021). Data that is ethically the most sensitive can sometimes be the most valuable, and the ability to utilize it depends on the ability to preserve the privacy of the research subjects (Dennis et al., 2019). Research ethics committees (RECs) or similar regulatory bodies are tasked in the same way as their legislative peers with ensuring that no harm comes to research participants as a result of data sharing. Researchers are therefore not only bound by considerations of legislation, but also by the interpretation of the legislation by RECs. This may create a scenario where a REC may not approve research activity to satisfy open science principles, including open data, open source, and sample sharing, even though provisions are made for this under the legislative framework. It is therefore to the benefit of researchers that they are cognizant of this fact and work in collaboration with those able to provide legal and ethical guidance during the construction of their research protocols. This is particularly important in LMICs where resources needed to repeat certain aspects of their studies are often lacking due to legal or ethical constraints.

Researchers should similarly be sensitive to the fact that once data is shared, it is very hard to take it back. Additionally, despite the obvious ease of identifying research participants using personal information such as names and addresses, it is possible to reverse-engineer an identity from a wide variety of anonymized sources (Narayanan and Shmatikov, 2010). To protect sensitive data from unauthorized use, computational analysis must be accompanied by strict access control mechanisms and non-technical measures such as informed consent. It is considered unethical to upload data that has not been anonymized; recruiting research participants would not be permitted if this were not done (Dennis et al., 2019). As such, open sharing of sensitive data may be deemed illegal and may hold dire consequences for those who partake

in such practices without adequate authorization (Dennis et al., 2019). Problems associated with data sharing are therefore critical to answer, especially for those engaged with qualitative research and associated data (Kirilova and Karcher, 2017).

Because sophisticated reverse-engineering identity techniques may allow for the re-identification of research participants, it is not possible to guarantee anonymity. While it has been suggested that data should never be shared, many are in favor of sharing and are actively engaged in developing policies and processes that would facilitate data sharing through ethically and legally sound means (Kirilova and Karcher, 2017). With regard to privacy, key issues such as the nature of the consent and who owns the data must be considered. This is important since most research institutions claim ownership of data collected by their researchers (Dennis et al., 2019). It is recognized however that research participants remain the owners of their information, even if collected for a research study, and therefore have the legal and ethical right to request that their data be amended and/or deleted without question or consequence. However, in practice, few participants execute their authority to do so largely due to logistical barriers. In addition, some researchers treat data like they own it and retain the data upon moving between institutions. In extreme and limited cases, while lacking the appropriate institutional approval to do so, researchers may publish the data on open platforms (Dennis et al., 2019). Some open data policies permit the secondary use of research data. Under these circumstances, data is used by researchers not involved in obtaining the initial consent for purposes and studies outside of the initial consent (Cummings et al., 2015). Research participants are not necessarily informed of this practice. Furthermore, research participants are not informed of the purpose for which their data will be reused under these open data policies. Such policies may have an impact on obtaining informed consent, especially if potential participants refuse to participate because of these open data policies, which may result in unreliable sample information and databases (Cummings et al., 2015). This may further result in legal consequences such as fines and criminal penalties, violations of ethical standards or data protection regulations that may result in irreparable damage to a provider's reputation (Wiesenauer et al., 2012).

It is thus imperative that the principles that govern ethical open science practices, including the sharing of samples and data, be observed for all data in order to experience the maximum benefit from such data and information while ensuring protection of the research participants (Martani et al., 2019). When it comes to secondary use of data, the risk-benefit considerations must be balanced so as to provide a useful resource to others while limiting the risk of exposure to participants. This is particularly important in health research as the secondary use of data increases the range of research projects that can be conducted, reduces not only the time spent on projects but also the operational and research costs, and increases the capacity of healthcare professionals to make evidence-based decisions for the continued improvement and delivery of good quality healthcare. Martani et al. (2019) have reported three categories of cutting-edge research initiatives within the healthcare sector. These include reusing data for: (1) "genomics and environmental health research;" (2) "clinical research in order to more rapidly identify and potentially recruit research participants;" and (3) "retrospective comparison of data from patients that have received conventional or alternative treatments, respectively."

Open science: Other barriers and misconceptions

The cost of accessing subscription journals, as previously discussed, is one of the most prominent barriers to dissemination of research findings in LMICs and may prevent research from being accessible to scientists and the public alike (Newton, 2020; Kwon, 2022). This challenge has however been exacerbated by the shift to online and open access publication models, given the difficulties that LMICs may experience with access to the Internet. When Internet access is possible, it still remains expensive and sometimes unstable within LMIC settings. Another barrier to open access for LMICs is the exorbitant and often prohibitively high APCs/publication fees which researchers or their associated institutions are unable or unwilling to pay. Many of these fees are more than the annual subscription to the journal, and often exceed the monthly salary of a researcher. Some researchers from LMICs are exempt from fees, but this is often reserved for the countries with the lowest gross domestic products and weakest economies. Many open access publishers impose a delay thereby decreasing the immediacy of the research or an embargo period, for example, 6 months for "green" open access. As indicated by Mwelwa et al. (2020), other barriers to open access and open science in Africa can broadly be summarized as deficiencies or limitations with regard to governmental or political, regulatory, institutional, financial, and researcher-centric structures. These categories include those challenges created by a lack of resources, such as access to research databases and journals, human capital and information and communication technology infrastructure, as well as the distrust or concerns that researchers may have regarding the ownership of published findings and any subsequent product developments. This latter point also concerns academic institutions and funding agencies, notably those located in Africa, who are adversely affected by the costs of APCs for reasons previously discussed. These challenges are perpetuated at government level owing to "a lack of political commitment in governments" and "a lack of national and institutional policies to provide a legal and regulatory framework for open science" (Mwelwa et al., 2020). In South Africa this problem is being addressed through a National Open Science Policy currently in draft form (Research Professional News, 2022).

Perception of research quality

The initial perception was that open access publication and research would be of lower quality. As open access has expanded, misinformation and concerns have decreased and most researchers now have positive attitudes to open access publishing (Nobes and Harris, 2019). The quality of open access publishing would only be compromised if journals did not follow a rigorous peer-review process. Authors should choose reputable journals for open access publishing. Predatory and fraudulent journals do not provide the same quality publications as reputable journals and should therefore be avoided.

It is well-known that articles published in subscription-based journals are initially only visible to people at institutions which have a license for these specific journals and are thus less visible than those published in open access journals. Despite this well-known fact and as measured through citations, researchers do not necessarily prioritize publishing in open access journals (Perianes-Rodr-Guez, 2019).

To increase the visibility and credibility of research findings, researchers may make research data available in open access repositories (Misgar et al., 2020). The Registry of Research Data Repositories provides an overview of repositories available for research data across all academic disciplines and is funded through the German Research Foundation (Registry of Research Data Repositories, 2022). In 2019, the registry indexed 2264 repositories with a metadata description (Misgar et al., 2020). Digital Object Identifiers (DOIs) are used to access and cite registry records. Open access research data repositories have also been developed by Brazil, Russian Federation, India, China, and South Africa, an association of five countries also referred to as the BRICS (Misgar et al., 2020). The highest number of repositories is found in China (81), followed by India (51), Russian Federation (23), Brazil (18), and South Africa (15). English is the common language among all BRICS countries.

Getting practical: Real-world impact of APCs

The previous sections have provided background on open science, and in particular open access. The COVID-19 pandemic has revealed that open access is inequitable, especially as far as LMICs and UMICs are concerned. The opportunities of preprints and challenges with predatory journals and research quality are briefly discussed. However, the core of this paper exposes the challenges that authors/researchers experience with high APCs and the resultant profits that publishers make. The role of higher education and research institutions in providing resources for authors/researchers are explained with the emphasis on UMICs, especially South Africa. Privacy and research ethics in medical research becomes challenging in open data sharing environments and need to be strictly regulated. The following is an overview of how researchers at the ICMM at the University of Pretoria manages the high APCs and their choice of journals with limited research funds.

From a financial perspective, the first question asked concerning a manuscript is whether or not there are APCs. Briefly, if no APCs are charged, assuming that the journal meets the various quality standards set by the research industry and/or institute (Figure 1), authors will proceed to submit manuscripts for publication. If APCs are charged, several additional questions are asked prior to submission of the manuscript. These include whether there are fee waivers, fee discounts or other funding opportunities available for publications. Where a suitable journal cannot be selected owing to cost or lack of funding, lower APCs, or no APCs may be considered instead. Should no suitable journal be found for the manuscript, submission of the manuscript may be delayed until a financially suitable publication option becomes available.

From an academic perspective and as illustrated in Figure 1, manuscript submission is determined by journal suitability, accreditation and quality (as measured through the IF). While subject specificity is evaluated through journal titles and research focus areas (Figures 1, 2), at the University of Pretoria, journals are considered to be accredited if they appear on lists generated by Clarivate Analytics Web of Science, IBSS, the South African DHET, Norwegian, SciELO SA, Scopus, and the DOAJ. This typically results in a list of several potential journals that are further evaluated according to their IFs. Traditionally, journals with higher IFs are believed to publish research of higher quality and are therefore more likely to gain greater research exposure and readership. As reported by Alberts (2013), there are unfortunately evaluation structures that consider journals with IFs less than 5.0 to be "of zero value". While this is largely dependent on the field of research and data used to determine such metrics, journals with IFs of at least 5.0 are meant to represent the top 10–20% of all journals within the medical sciences and may therefore be seen by some as an arbitrary benchmark of "good quality" (Alberts, 2013). Clarivate's Journal Citation Reports are often used for this purpose (Clarivate, 2022). While this standard is not strictly adhered to within the ICMM, journals with higher IFs will nevertheless be favored over lower IF journals for manuscript submission (Figure 1). Assuming that the academic and financial components have been adequately met, the manuscript may be submitted for review.

Practically applied, the University of Pretoria's Library Services website provides a list of 1,581 journal titles associated with fee waivers for open access (Figure 2). Considering that the ICMM conducts inter- and cross-disciplinary research, by observing the journal title alone, 129 (8.2%) and 385 (24.4%) journals would be perceived as being potentially relevant or relevant for manuscript submission, respectively. At the ICMM, journals focused on stem cell research, obesity, diabetes, cancer, cystic fibrosis, hypoxic ischemic encephalopathy, human

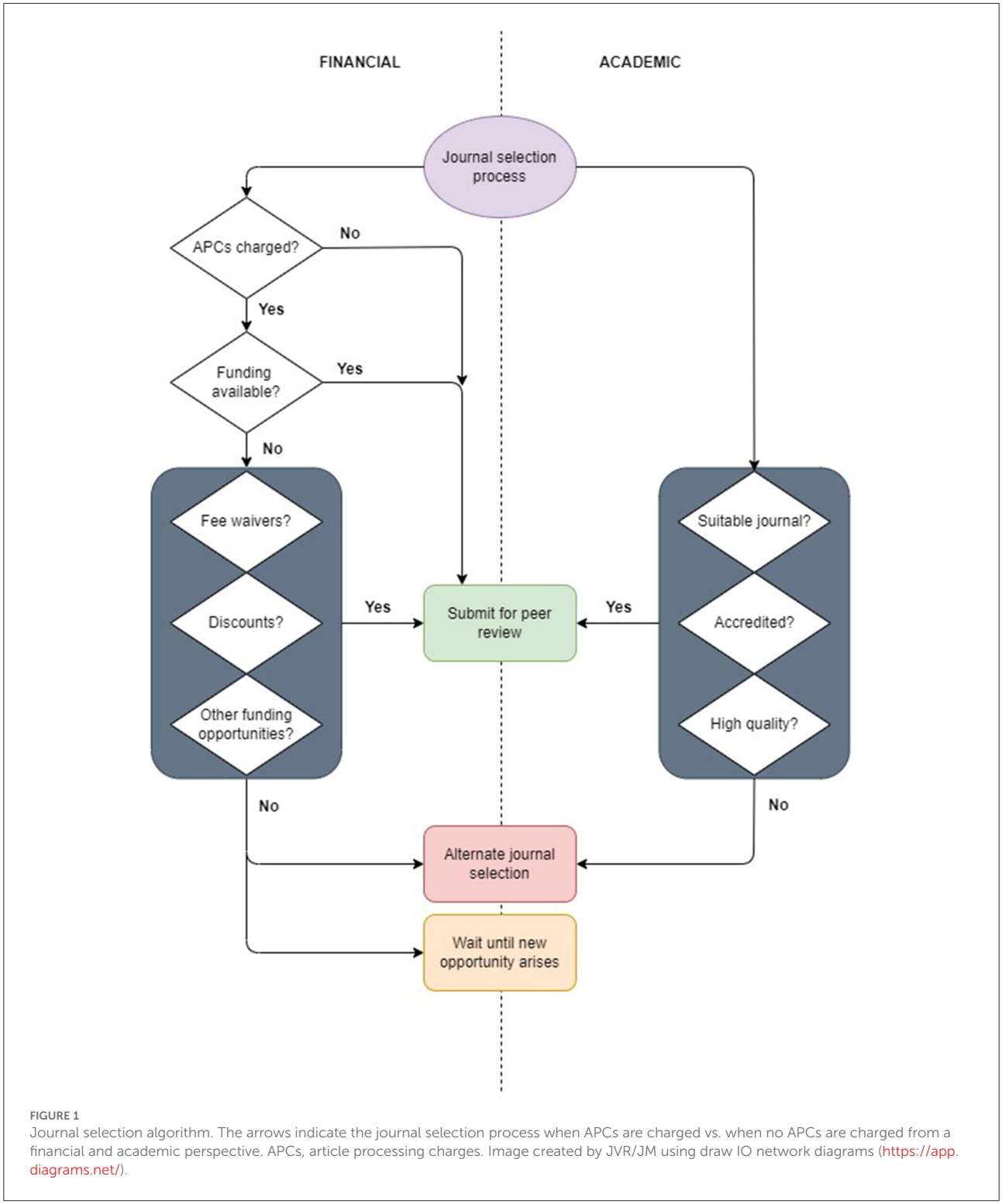


FIGURE 1
Journal selection algorithm. The arrows indicate the journal selection process when APCs are charged vs. when no APCs are charged from a financial and academic perspective. APCs, article processing charges. Image created by JVR/JM using draw IO network diagrams (https://app.diagrams.net/).

immunodeficiency virus, human leukocyte antigen studies, and genetic susceptibility to disease would be considered from a cellular and molecular perspective. Where research has been conducted in cross-disciplinary fields, relevant law, computational biology, and engineering journals may also be considered. As such, when considering these potential journal titles based on research content, only 121 journals (7.7%) would be further considered for publication. Of these, impact factors were only available for 97 journal titles (6.1%), with a total of 24 journals (1.5%) having an IF of at least 5.0 (Figure 2, Table 2).

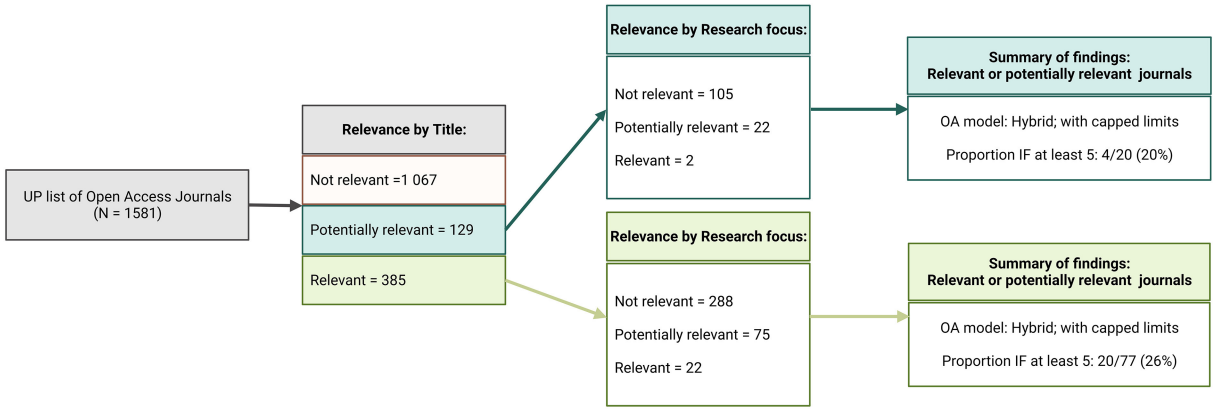


FIGURE 2
Open access journals with fee waivers considered to be relevant or potentially relevant to research conducted at the ICMM. UP, University of Pretoria; OA, open access; IF, impact factor. Image created by JVR/AA using BioRender (https://Biorender.com/).

While the median and mean IFs were respectively found to be 3.617 and 4.409, the maximum and minimum IFs were noted as 25.113 and 0.910, respectively. Each of these titles are associated with a hybrid open access model with "first come, first served" capped limits to fee waivers.

While the finer breakdown of this exercise can be seen in both Figure 2 and Table 2, it must however be noted that this form of evaluation is subjective in nature, so there may be some differences between what one researcher notes as potentially being of value compared to another. This may similarly be extended to what is understood by "predatory journal." In the context of this exercise, a predatory journal represents a journal that is (1) not accredited by the University of Pretoria standards previously indicated; (2) requests payment for manuscript review; (3) does not offer peer-review or offers sub-standard peer-review; (4) has a very low impact factor; and (5) promises rapid manuscript review, acceptance and publication. Regardless of these points, following the examination of the open access journals for which fee waivers were noted according to the list available through the University of Pretoria's Library Services website, it is important to note that capped limits for hybrid open access journals exist for all journals that would potentially or definitely be considered for manuscript submission by researchers at the ICMM. While it is not publicly known whether the capped limits are per faculty, per academic institution, or per country, or whether the limits are set to one, ten, 100 or 1,000 or more publications, that they are hybrid journals in nature *per se* is not of concern. What is of concern is that the University of Pretoria (and other academic institutions in South Africa) does not contribute to APCs for hybrid journals. Recently signed agreements between select publishers and research institutions like the University of Pretoria and others in South Africa have created an opportunity for complete fee waiving on select hybrid journals (UP news, 2022). This is of significant value to South African researchers operating within participating institutions (such as the University of Pretoria) and has the potential to have far-reaching benefits to individuals in public and private entities.

Solutions to the challenges

Approximately a third of global research articles are now being published as open access and there is a strong drive to further increase this number (STM Global, 2021; Delta Think, 2022). Currently, the peer-review process has little benefit for the reviewer. Some journals offer incentives and rewards to reviewers such as subscription access for a limited period of time (The Conversation Africa, 2022b). This is not ideal for universities as it only benefits the individual reviewer. Instead, publishers may consider a voucher approach where vouchers are provided to reviewers' institutions. In LMICs this may contribute toward journal subscription costs or APCs and may also encourage academics to be involved in the review process. The high APCs associated with open access publishing remain a challenge for researchers in LMICs (Kwon, 2022). However, there are some options to consider when APCs are required and funding is limited; these include: (1) asking the publisher whether the journal would consider waiving or reducing APCs for researchers in LMICs; (2) enquiring whether the representative institution maintains an agreement with certain publishers that will allow publishing for free or at a discounted rate; and (3) enquiring from funders about the availability of funds for publications related to awarded grants. Publishing in societal journals is another potential solution as profits from these journals are re-invested into supporting a wide range of research activities. Researchers should also include publishing costs in grant applications. This is already encouraged by some funders such as South Africa's

TABLE 2 The trends in impact factor relative to journals marked as potentially relevant or relevant to researchers based at the ICMM.

Impact factor details	Potentially relevant journals	Relevant journal	All journals
N (IF < 5.0)	60	13	73
N (IF ≥ 5.0)	19	5	24
N (IF not reflected)	18	6	24
Minimum	0.91	1.615	0.910
1st IQR	2.452	2.087	2.447
Median	3.780	3.587	3.617
Mean	4.419	4.365	4.409
3rd IQR	4.972	4.971	4.982
Maximum	25.113	11.598	25.113

IQR, Interquartile range; IF, impact factor.

National Research Foundation. Other proposed solutions include increasing government subsidy for universities to aid in covering full APCs. Current government subsidies received by academic institutions to cover publication costs are divided amongst the Institution, Faculties, Schools and Departments, with a very limited amount of funding trickling down to researchers on an *ad hoc* basis. The full subsidy or a portion of this should in all fairness be returned to researchers to cover the costs of future publications. Additional agreements should be put in place between universities and publishing houses to assist researchers in LMICs to publish high quality research in reputable journals.

In relation to cost and in order to address some of the current pressures, Plan S and several other initiatives have been established to increase open access. Plan S is an initiative that was implemented by Science Europe, a group of state funded scientists and researchers from 12 national European funding agencies referred to as "cOAlition S" (Plan, 2022). In 2021, as part of Plan S, a group of international funders (including UK Research and Innovation, the Bill and Melinda Gates Foundation, the Wellcome Trust, and the World Health Organization) launched a major reform of the way funded research is published (Mering, 2020; Shamseer et al., 2021). Recipients of grants funded by agencies affiliated to Plan S are required to publish research in open access journals or platforms and make publications available *via* open access repositories (Plan, 2022). In order to facilitate this, funders will pay APCs (up to a limit still to be determined) to "gold" open access journals. Furthermore, Plan S supports the retention of publication copyrights by authors and their affiliated institutions, as well as the use of open licenses (Mering, 2020). Although the hybrid journal publishing model is not supported, transformative journal agreements are provided as an option to gradually increase the amount of journal open access content. The aim is for all journals to be open access by the end of 2024. Among the few non-European agencies, the South African Medical Research Council has also joined Plan S and may serve as an early indication of how Plan S will operate in LMICs (The Scientist, 2022). Some requirements of Plan S already encourage more equitable publishing practices. Open access journals or platforms publishing results generated using funds from Plan S signatories should provide APC relief either through waivers or discounts for researchers from LMICs (The Scientist, 2022). Some researchers have suggested that Plan S should support "diamond" open access journals, allowing free reading and free publishing. These journals are often supported by scholarly societies, receive funding from higher education institutions, and the editorial boards consist of volunteer editors (The Scientist, 2022). According to Robert Kiley, the head of strategy of cOAlition S and the head of open research development at the Wellcome Trust, publishers are required to share their pricing and service data with cOAlition S, starting in July 2022. Incentives will be provided for publishers to do so, such as continued funding by cOAlition S to cover open access publication costs. In the future, researchers should be able to choose to pay only for essential publication services and should be exempt from covering marketing and other non-essential costs (The Scientist, 2022). There are various other open access projects under the SHERPA (Securing a Hybrid Environment for Research Preservation and Access) organization (SHERPA, 2022), UNESCO's Open Science project (UNESCO, 2022b) and Research4Life (R4L) (Research4Life, 2022), formerly the WHO Hinari Program.

According to the journal Science, authors of research papers will be allowed to share an almost final version of their articles in a public repository of their choice without paying any fees from January 2023 (Else, 2022). The policy will apply to all five subscription journals in the Science family. Currently, most authors can share their accepted articles only in an institutional repository or on a personal website. Authors have an embargo period of 6 months after publication before they can add their papers to other repositories, such as PubMed. However, there are exceptions for some authors supported by funders from cOAlition S.

Concluding remarks

Limited access by researchers in LMICs and UMICs to the latest research information affects their ability to provide quality work of current relevance, and this barrier must be lifted. Open access serves to bridge this gap and will provide equity in the research space. The National Research Foundation's mandate is to contribute to national development by:

- "Supporting, promoting and advancing research and human capacity development, through funding and the provision of the necessary research infrastructure, in order to facilitate the creation of knowledge, innovation and development in all fields of science and technology, including humanities, social sciences and indigenous knowledge";
- "Developing, supporting and maintaining national research facilities";
- "Supporting and promoting public awareness of, and engagement with, science; and"
- "Promoting the development and maintenance of the national science system and support of Government priorities".

In the present technology driven era, it is critical to strengthen open access initiatives and to transform journals, platforms and repositories to make research freely available online. This will enable sharing of knowledge and enhance global communication, while improving research potential and visibility of institutions and researchers. The increase in predatory journals and misconceptions regarding open access are challenges that need to be overcome in order for open access to achieve its full potential. In order for this endeavor to be successful in LMICs and UMICs (such as South Africa), governments and funding agencies need to adopt and improve open access initiatives in support of African research. The number of open access journals that offer fee waivers for LMICs should be increased to broaden journal options for publication. Until all countries and scientific communities have equal access to all research available, the impact of open access will remain limited and inequitable.

Author contributions

AS and MSP completed the final preparation and editing of the manuscript. JVR, JM, and AA created the tables and figures. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the South African Medical Research Council Extramural Unit for Stem Cell Research and Therapy and the University of Pretoria through the Institute for Cellular and Molecular Medicine. AS, JM, JVR, and MSP receive funding from the Bill and Melinda Gates Foundation (INV-022216).

Acknowledgments

The authors thank the South African Medical Research Council and the University of Pretoria for supporting this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Academy of Science of South Africa (2022). *POPIA: A Code of Conduct for Research*. Available online at: https://www.assaf.org.za/popia-a-code-of-conduct-for-research/ (accessed May 4, 2022)

Alberts, B. (2013). Impact factor distortions. *Science*. 340, 787. doi: 10.1126/science.1240319

Bawa, A. C. (2020). South Africa's journey towards open access publishing. *Biochemist*. 42, 30–33. doi: 10.1042/BIO20200029

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., et al. (2021). Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med. Res. Methodol.* 21, 117. doi: 10.1186/s12874-021-01304-y

BioRxiv (2022). *Preprint server for biology*. Available online at: https://www.biorxiv.org/ (accessed March 28, 2022).

Brierley, L. (2021). Lessons from the influx of preprints during the early COVID-19 pandemic. *Lancet. Planet. Health*. 5, e115–e117. doi: 10.1016/S2542-5196(21)00011-5

Callaway, E. (2021). Rare COVID reactions might hold key to variant-proof vaccines. *Nat. Inte. Weekly J. Sci.* 592, 20–21. doi: 10.1038/d41586-021-00722-8

Clarivate (2022). *Journal Citation Reports*. Available online at: https://jcr.clarivate.com/jcr/home (accessed October 19, 2022).

Cummings, J. A., Zagrodney, J. M., Day, T. E., and Metcalfe, J. Z. (2015). Impact of open data policies on consent to participate in human subjects research: discrepancies between participant action and reported concerns. *PLoS ONE*. 10, e0125208. doi: 10.1371/journal.pone.0125208

Czerniewicz, L., and Goodier, S. (2014). Open access in South Africa: a case study and reflections. *South Afr. J. Sci.* 110, 1–9. doi: 10.1590/sajs.2014/20140111

De Jager, P., De Kock, F., and Van Der Spuy, P. (2017). Do not feed the predators. *South Afr. J. Bus. Manage*. 48, 35–45. doi: 10.4102/sajbm.v48i3.34

Delta Think (2022). *Open Access Market Sizing Update*. Available online at: https://deltathink.com/news-views-open-access-market-sizing-update-2021/#:$\sim$:text=The%20open%20access%20market%20is,17.6%25%20in%20OA%20market%20value (accessed October 13, 2022).

Dennis, S., Garrett, P., Yim, H., Hamm, J., Osth, A. F., Sreekumar, V., et al. (2019). Privacy versus open science. Behav. Res. *Methods*. 51, 1839–1848. doi: 10.3758/s13428-019-01259-5

DOAJ (2022). *Directory of Open Access Journals*. Available online at: https://doaj.org/ (accessed March 1, 2022).

Else, H. (2020). How a torrent of COVID science changed research publishing — in seven charts. *Nat.: Int. Weekly J. Sci.* 588, 553. doi: 10.1038/d41586-020-03564-y

Else, H. (2022). Science's no-fee public-access policy will take effect in 2023. *Nature*. doi: 10.1038/d41586-022-03128-2

Figshare (2022). *UP Research Data Repository*. Available online at: https://researchdata.up.ac.za/ (accessed April 7, 2022).

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., et al. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biol.* 19, e3000959. doi: 10.1371/journal.pbio.3000959

Giandhari, J., Pillay, S., Wilkinson, E., Tegally, H., Sinayskiy, I., Schuld, M., et al. (2021). Early transmission of SARS-CoV-2 in South Africa: an epidemiological and phylogenetic report. *Int. J. Infect. Dis*. 103, 234–241. doi: 10.1016/j.ijid.2020.11.128

Hagve, M. (2020). The money behind academic publishing. *Tidsskr. Nor. Laegeforen*. 140, 19. doi: 10.4045/tidsskr.20.0118

Hartman, M. S., and Wu, S. (2018). Open access in South Africa: a coherent strategy is needed. *South Afr. J. Sci.* 114, 1–2. doi: 10.17159/sajs.2018/a0259C

Horby, P. (2022). Why preprints are good for patients. *Nat. Med*. 28, 1109. doi: 10.1038/s41591-022-01812-4

Huston, P., Edge, V. L., and Bernier, E. (2019). Reaping the benefits of Open Data in public health. *Can. Commun. Dis. Rep*. 45, 252–256. doi: 10.14745/ccdr.v45i10a01

Inter Academy Partnership (2022). *Predatory Publishing*. Available online at: https://www.interacademies.org/project/predatorypublishing (accessed March 17, 2022).

International Science Council (2022). *COVID-19 and Access to Scientific Knowledge*. Available online at: https://council.science/current/blog/covid-19-and-access-to-scientific-knowledge/ (accessed February 24, 2022).

Kana, M. A., Laporte, R., and Jaye, A. (2021). Africa's contribution to the science of the COVID-19/SARS-CoV-2 pandemic. *BMJ Global Health*. 6, e004059. doi: 10.1136/bmjgh-2020-004059

Kirilova, D., and Karcher, S. (2017). Rethinking data sharing and human participant protection in social science research: applications from the qualitative realm. *Data Sci. J*. 16, 43. doi: 10.5334/dsj-2017-043

Koerber, A. (2021). Is it fake news or is it open science? *J.Bus.Tech.Commun*. 35, 22–27. doi: 10.1177/1050651920958506

Kwon, D. (2020). How swamped preprint servers are blocking bad coronavirus research. *Nat. Int. Weekly J. Sci*. 581, 130–131. doi: 10.1038/d41586-020-01394-6

Kwon, D. (2022). *Open-access publishing fees deter researchers in the global south. Nature* doi: 10.1038/d41586-022-00342-w

Larivière, V., Haustein, S., and Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE*. 10, e0127502. doi: 10.1371/journal.pone.0127502

Mail and Guardian (2022). *Open Access to Academic Work*. Available online at: https://mg.co.za/article/2015-06-25-door-slammed-on-open-access-to-academic-work/ (accessed April 6, 2022).

Martani, A., Geneviève, L. D., Pauli-Magnus, C., Mclennan, S., and Elger, B. S. (2019). Regulating the secondary use of data for research: arguments against genetic exceptionalism. *Front. Gen*. 101254. doi: 10.3389/fgene.2019.01254

MedRxiv (2022). *Preprint Server for Health Sciences*. Available online at: https://www.medrxiv.org/ (accessed March 28, 2022).

Mering, M. (2020). Open access mandates and policies: the basics. *Serials Rev*. 46, 157–159. doi: 10.1080/00987913.2020.1760707

Misgar, S. M., Bhat, A., and Wani, Z. A. (2020). A study of open access research data repositories developed by BRICS countries. *Digital Lib. Persp*. 38, 45–54. doi: 10.1108/DLP-02-2020-0012

Muñoz-Tamayo, R., Nielsen, B. L., Gagaoua, M., Gondret, F., Krause, E. T., Morgavi, D. P., et al. (2022). Seven steps to enhance open science practices in animal science. *PNAS Nexus*. 1, 1–6. doi: 10.1093/pnasnexus/pgac106

Mwelwa, J., Boulton, G., Wafula, J. M., and Loucoubar, C. (2020). Developing open science in Africa: barriers, solutions and opportunities. *Data Sci. J*. 19, 31. doi: 10.5334/dsj-2020-031

Naidoo, A. V., Hodkinson, P., Lai King, L., and Wallis, L. A. (2021). African authorship on African papers during the COVID-19 pandemic. *BMJ Global Health*. 6, e004612. doi: 10.1136/bmjgh-2020-004612

Narayanan, A., and Shmatikov, V. (2010). Myths and fallacies of "personally identifiable information". *Commun. ACM*. 53, 24–26. doi: 10.1145/1743546.1743558

National Department of Health (2022). *Ethics in Health Research*. Available online at: https://www.health.gov.za/nhrec-guidelines/ (accessed May 3, 2022).

National Treasury (RSA) (2022). *Budget Review, 2022*.Available online at: http://www.treasury.gov.za/documents/national%20budget/2022/review/FullBR.pdf (accessed October 19, 2022).

Newton, C. R. (2020). Research and open access from low- and middle-income countries. *Dev. Med. Child Neurol*. 62, 537. doi: 10.1111/dmcn.14513

Nobes, A., and Harris, S. (2019). Open access in low- and middle-income countries: attitudes and experiences of researchers. *Emerald Open Res*. 17, 11–24. doi: 10.35241/emeraldopenres.13325.1

OpenDOAR (2022). *Directory of Open Access Repositories*. Available online at: https://v2.sherpa.ac.uk/opendoar/ (accessed April 7, 2022).

Perianes-Rodr-Guez, A, and Olmeda-GMez, C. (2019). Effects of journal choice on the visibility of scientific publications: a comparison between subscription-based and full open access models. *Scientometrics* 121, 1737–1752. doi: 10.1007/s11192-019-03265-y

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., et al. (2018). The state of OA: a large-scale analysis of the prevalence and impact of open access articles. *PeerJ*. 6, e4375. doi: 10.7717/peerj.4375

Plan, S. (2022). *Making Full and Immediate Open Access a Reality*. Available online at: https://www.coalition-s.org/ (accessed March 29, 2022)

Prlic, A., and Procter, J. B. (2012). Ten simple rules for the open development of scientific software. *PLoS Comput. Biol*. 8, e1002802. doi: 10.1371/journal.pcbi.1002802

Publons (2022). *Track More of Your Research Impact*. Available online at: https://publons.com/about/home/ (accessed May 23, 2022).

Registry of Research Data Repositories (2022). *Index of Research Data Repositories*. Available online at: https://www.re3data.org/ (accessed April 5, 2022).

Research Gate (2022). *The Influence of the Open Access Movement on Faculty Research Output and University Libraries in Africa*. Available online at: https://www.researchgate.net/publication/353886046 (accessed April 6, 2022).

Research Professional News (2022). *South Africa's Open Science Policy*. Available online at: https://www.researchprofessionalnews.com/rr-news-africa-south-2022-2-south-africa-s-draft-open-science-policy-promises-shake-up/ (accessed July 13, 2022).

Research4Life (2022). Online Access for LMICs to Academic and profEssional Content. Available online at: https://www.research4life.org/ (accessed May 11, 2022).

SANLiC (2021). *South African National Library and Information Consortium*. Available online at: https://sanlic.org.za/wp-content/uploads/2021/05/SANLiC-Statement-on-Open-Access-November-2020.pdf (accessed April 6, 2022).

Schimmer, R., Geschuhn, K., and Vogler, A. (2015). Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. *Sci. Open Res*. 1–7. doi: 10.14293/S2199-1006.1.SOR-EDU.AJRG23.v1

Sci-Hub (2022). *Free and Unrestricted Access to All Scientific Knowledge*. Available online at: https://sci-hub.se/ (accessed May 23, 2022).

Shamseer, L., Cobey, K. D., Page, M. J., Brehaut, J. C., Grimshaw, J. M., Straus, S. E., et al. (2021). Top health research funders' guidance on selecting journals for funded research. *F1000Research* 10, 100. doi: 10.12688/f1000research.27745.2

Sharma, M., Sarin, A., Gupta, P., Sachdeva, S., and Desai, A. V. (2014). Journal impact factor: its use, significance and limitations. *World J. Nucl. Med*. 13, 146. doi: 10.4103/1450-1147.139151

SHERPA (2022). *Securing a Hybrid Environment for Research Preservation and Access*. Available online at: https://v2.sherpa.ac.uk/ (accessed May 11, 2022).

Shuai, X., Pepe, A., and Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: article downloads, twitter mentions, and citations. *PLoS ONE*. 7, e47523. doi: 10.1371/journal.pone.0047523

South African Government (2022). *DoH Policy on Quality in Health Care for South Africa*. Available online at: https://www.gov.za/documents/policy-quality-health-care-south-africa (accessed October 19, 2022).

Statistics South Africa (2022). *Statistical Release*. Available online at: https://www.statssa.gov.za/publications/P0302/P03022019.pdf (accessed October 19, 2022).

Staunton, C., Adams, R., Botes, M., Dove, E. S., Horn, L., Labuschaigne, M., et al. (2020). Privacy rights of human research participants in South Africa must be taken seriously. *S. Afr. Med. J.* 110, 175–176. doi: 10.7196/SAMJ.2020.v110i3.14450

Staunton, C., Adams, R., Botes, M., Vries, D. E., Labuschaigne, P. J., Loots, M., et al. (2021). Enabling the use of health data for research : developing a POPIA code of conduct for research in South Africa. *South Afr. J. Bioethics Law*. 14, 33–36. doi: 10.7196/SAJBL.2021.v14i1.740

Staunton, C., Adams, R., Botes, M., Dove, E. S., Horn, L., Labuschaigne, M., et al. (2019). Safeguarding the future of genomic research in South Africa: broad consent and the protection of personal information act no. 4 of 2013. *S. Afr. Med. J.* 109, 468–470. doi: 10.7196/SAMJ.2019.v109i7.14148

STM Global (2021). *Economics and Market Size*. Available online at: https://www.stm-assoc.org/wp-content/uploads/2021_10_19_STM_Global_Brief_2021_Economics_and_Market_Size-1.pdf (accessed October 13, 2022).

Szekely, T., Kruger, O., and Krause, E. T. (2014). Errors in science: the role of reviewers. *Trends Ecol. Evol.* 29, 371–373. doi: 10.1016/j.tree.2014.05.001

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., Hartgerink, C. H. J., et al. (2016). The academic, economic and societal impacts of open access: an evidence-based review. *F1000Research* 5, 632. doi: 10.12688/f1000research.8460.3

Thaldar, D. W., and Townsend, B. (2020). Genomic research and privacy: a response to staunton et al. *S. Afr. Med. J.* 110, 172–174. doi: 10.7196/SAMJ.2020.v110i3.14431

The Conversation Africa (2022a). *Open Access in Poorer Countries*. Available online at: https://theconversation.com/how-the-open-access-model-hurts-academics-in-poorer-countries-113856 (accessed February 28, 2022).

The Conversation Africa (2022b). *Cost of Accessing Academic Research*. Available online at: https://theconversation.com/the-cost-of-accessing-academic-research-is-way-too-high-this-must-change-105583 (accessed April 6, 2022).

The Scientist (2022). *Impact of Plan S*. Available online at: https://www.the-scientist.com/news-opinion/as-plan-s-takes-effect-some-anticipate-inequitable-outcomes-69058 (accessed April 13, 2022).

The World Bank (2022). *The World by Income and Region*. Available online at: https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html (accessed August 2, 2022).

UNESCO (2017). *Access by Region*. Available online at: http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/access-by-region/africa/south-africa/ (accessed March 1, 2022).

UNESCO (2022a). *Open Access Publications*. Available online at: https://en.unesco.org/open-access/what-open-access (accessed May 16, 2022).

UNESCO (2022b). *Open Science Project*. Available online at: https://www.unesco.org/en/natural-sciences/open-science (accessed May 11, 2022).

Unpaywall (2022). *Open Database of Free Scholarly Articles*. Available online at: https://unpaywall.org/ (accessed May 23, 2022).

UP News (2022). *Research Published OA at No Cost*. Available online at: https://www.up.ac.za/news/post_3053626-up-researchers-can-now-publish-their-research-open-access-at-no-cost (accessed March 18, 2022).

UP Policies (2022). *OA Publishing Charges*. Available online at: https://www.up.ac.za/media/shared/538/ZP_Files/s4420_14-policy-on-open-access-publishing-processing-charges.zp106764.pdf (accessed February 15, 2022).

UP Space (2022). *UP Institutional Repository*. Available online at: https://repository.up.ac.za/ (accessed April 7, 2022).

Van Noorden, R. (2013). Open access: the true cost of science publishing. *Nature* 495, 426–429. doi: 10.1038/495426a

Vervoort, D., Ma, X., and Shrime, M. G. (2020). Money down the drain: predatory publishing in the COVID-19 era. *Can. J. Pub. Health*. 111, 665–666. doi: 10.17269/s41997-020-00411-5

Vraga, E. K., Tully, M., and Bode, L. (2020). Empowering users to respond to misinformation about COVID-19. *Media Commun*. 8, 475–479. doi: 10.17645/mac.v8i2.3200

Watson, C. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nat. Med*. 28, 2–5. doi: 10.1038/s41591-021-01654-6

Wiesenauer, M., Johner, C., and Rohrig, R. (2012). Secondary use of clinical data in healthcare providers – An overview on research, regulatory and ethical requirements. *Stud. Health Technol. Inform*. 180, 614–618. doi: 10.3233/978-1-61499-101-4-614

Woodiwiss, A. J. (2012). Publication subsidies: challenges and dilemmas facing South African researchers. *Cardiovasc. J. Afr*. 23, 421–427.

Yeo-Teh, N. S. L., and Tang, B. L. (2021). An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). *Account. Res*. 28, 47–53. doi: 10.1080/08989621.2020.1782203

York, U. (2022). *Open Access and Open Data*. Available online at: https://www.library.yorku.ca/web/open/overview/open-data/ (accessed May 16, 2022).

frontiers | Frontiers in Research Metrics and Analytics

TYPE Original Research
PUBLISHED 30 May 2023
DOI 10.3389/frma.2023.1070390

OPEN ACCESS

EDITED BY
Heila Pienaar,
University of Pretoria, South Africa

REVIEWED BY
Prashanth N. Suravajhala,
Amrita Vishwa Vidyapeetham University, India
Josiline Phiri Chigwada,
University of South Africa, South Africa

*CORRESPONDENCE
Caleb K. Kibet
✉ ckibet@icipe.org

RECEIVED 14 October 2022
ACCEPTED 10 May 2023
PUBLISHED 30 May 2023

CITATION
Karega P, Mwaura DK, Mwangi KW, Wanjiku M, Landi M and Kibet CK (2023) Building awareness and capacity of bioinformatics and open science skills in Kenya: a sensitize, train, hack, and collaborate model.
Front. Res. Metr. Anal. 8:1070390.
doi: 10.3389/frma.2023.1070390

Building awareness and capacity of bioinformatics and open science skills in Kenya: a sensitize, train, hack, and collaborate model

Pauline Karega[1,2], David K. Mwaura[3], Kennedy W. Mwangi[4], Margaret Wanjiku[5], Michael Landi[6,7] and Caleb K. Kibet[1]*

[1]International Center of Insect Physiology and Ecology, Nairobi, Kenya, [2]Department of Biochemistry, University of Nairobi, Nairobi, Kenya, [3]Institute of Primate Research (IPR), Nairobi, Kenya, [4]International Livestock Research Institute (ILRI), Nairobi, Kenya, [5]Department of Biology, San Diego State University, San Diego, CA, United States, [6]Department of Bioinformatics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [7]International Institute of Tropical Agriculture, Nairobi, Kenya

We have applied the sensitize-train-hack-community model to build awareness of and capacity in bioinformatics in Kenya. Open science is the practice of science openly and collaboratively, with tools, techniques, and data freely shared to facilitate reuse and collaboration. Open science is not a mandatory curriculum course in schools, whereas bioinformatics is relatively new in some African regions. Open science tools can significantly enhance bioinformatics, leading to increased reproducibility. However, open science and bioinformatics skills, especially blended, are still lacking among students and researchers in resource-constrained regions. We note the need to be aware of the power of open science among the bioinformatics community and a clear strategy to learn bioinformatics and open science skills for use in research. Using the OpenScienceKE framework—Sensitize, Train, Hack, Collaborate/Community—the BOSS (Bioinformatics and Open Science Skills) virtual events built awareness and empowered researchers with the skills and tools in open science and bioinformatics. Sensitization was achieved through a symposium, training through a workshop and train-the-trainer program, hack through mini-projects, community through conferences, and continuous meet-ups. In this paper, we discuss how we applied the framework during the BOSS events and highlight lessons learnt in planning and executing the events and their impact on the outcome of each phase. We evaluate the impact of the events through anonymous surveys. We show that sensitizing and empowering researchers with the skills works best when the participants apply the skills to real-world problems: project-based learning. Furthermore, we have demonstrated how to implement virtual events in resource-constrained settings by providing Internet and equipment support to participants, thus improving accessibility and diversity.

KEYWORDS

open science, bioinformatics, BHKi, OpenScienceKE, Africa

Introduction

Bioinformatics is the field that uses computational tools to capture, analyze, and interpret biological data (Bayat, 2002). Genome projects have increased manifold since the advent of cheaper next-generation sequencing technologies (Batley and Edwards, 2009), leading to an explosion of genomic sequences in public and private databases that need analysis and interpretation (Schneider et al., 2010). This explosion of data has also resulted

in new analysis techniques and bioinformatics solutions to process them. Therefore, a skilled bioinformatics workforce is constantly required (Braga et al., 2021). The workforce would consist of researchers with the skills or core competencies listed by the International Society of Computational Biology (ISCB) (Mulder et al., 2018). Bioinformatics training in Kenya and most parts of the world occurs primarily in graduate programs and, in some cases, at the undergraduate level (Sayres et al., 2018). However, the graduate programs are usually tailored for those who want to specialize in bioinformatics and may not be ideal for those who want to use bioinformatics as a tool (Aron et al., 2021; Ras et al., 2021). Therefore, short training for skill development programs is essential to equip students with the necessary bioinformatics skills to conduct research and data analysis. Short training courses tend to have customized specificities not covered by traditional courses (Braga et al., 2021) and also have leeway in their design, factoring in a specific audience in attendance, a luxury that traditional courses do not have.

Data availability for reuse by researchers, especially from resource-constrained settings, results from a strong move toward open data sharing, a practice known as open science.

Open science is an umbrella term comprising open access to publications, open research data, open-source software, open collaboration, peer review, notebooks, educational resources, monographs, citizen science, or research crowdfunding. Data is only reusable when FAIR—Findable, Accessible, Interoperable, Reusable. Therefore, data-generating researchers must be trained to make their data FAIR. Open science skills are essential in bioinformatics to facilitate open, reproducible, and collaborative research stages in the research life cycle (OECD, 2015). Open science training is usually through informal training by grassroots communities through workshops, short courses, and MOOCs on open science. However, it is necessary to combine bioinformatics and open science training, which are traditionally not taught together in formal curricula in Kenya and surrounding countries (Mwangi et al., 2021) to facilitate the effective adoption of open practices in bioinformatics research.

The COVID-19 pandemic demonstrated the need for open science adoption in biomedical research (Okafor et al., 2022). Practices such as open data, open access, open source, and open peer review enabled quick and timely responses from researchers globally who were experiencing restrictive movements to work together collaboratively (Besançon et al., 2021). The demand further demonstrates the need for awareness of open science, more so, open science by design, where there is the intent to conduct research transparently and openly at all stages of the research life cycle. We note that in computational analysis, the scholarly contribution is the data and the code that generated the results, with a strong move in the field toward collaborative and reproducible research. Therefore, this study sought to empower researchers with skills and tools in bioinformatics through a series of virtual events dubbed Bioinformatics and Open Science Skills (BOSS) using the sensitize-train-hack-collaborate model. The main objective of the series of events was to blend open science and bioinformatics and to train researchers to conduct open, reproducible, and collaborative research in bioinformatics. This paper describes how we applied the model virtually and highlights the successes and challenges of organizing such events in a resource-constrained region.

The BOSS events

The BOSS events had five phases: sensitize through a symposium, training through workshops, hacking through mini-projects, sustain through instructor training, and a conference to showcase work done during the events and network. The main objective of the series of events was to blend open science and bioinformatics and to train researchers to conduct open, reproducible, and collaborative research in bioinformatics. We aimed to reach an audience with beginner, intermediate, and advanced knowledge of bioinformatics and open science. We adopted the sensitize-train-hack-collaborate model previously developed to empower researchers with open science skills (Mwangi et al., 2021). We adjusted the model to include sustainability, an opportunity to network and share work through conferences, and redefining collaboration and community to encompass all aspects of the model (Figure 1). The model and steps incorporated Bloom's taxonomy of effective learning to ensure that the content equipped trainees with knowledge and skills (Bloom, 1969). All the events were virtual using Zoom video conferencing.

Sensitize: the open science FAIR symposium

Creating awareness and providing information on the need and benefits of practicing open science principles in bioinformatics formed the foundation step of the framework during its conception. Therefore, the FAIR Open Science Symposium aimed to create awareness of the need and the benefits of practicing open science in research (Table 1).

The symposium received 130 applications from various African countries such as Kenya, Uganda, Cameroon, Morocco, Nigeria, Tunisia, Botswana, Ghana, South Africa, Tanzania, and Zimbabwe (Figure 2). The event included participants outside Africa: the United States of America, India, Lithuania, France, Bangladesh, Iran, New Zealand, South Korea and Portugal. However, only 55 participants attended, with numbers varying throughout the week, with a minimum of 35. The active participants were mainly from Kenya, Uganda, Tanzania, the United States, Zimbabwe, South Africa, and South Korea.

The 5-day event had sessions that were in the form of talks, open panel discussions, and practical lessons. The sessions included expert presentations followed by panel discussions guided by questions from the audience. All sessions were accessible to all registered participants.

The open science session addressed the topics of the status of open science in Kenya and Africa, policies in higher education and research institutions, and the challenges faced, especially in the practice of open science. The speakers gave visual presentations on adopting open-access in Kenya and Africa, pointing out the growth of open access journals in the region.

The sessions on research data management aimed to sensitize the FAIR principles, introducing the participants to the definitions of each component of FAIR. The speakers covered scientific research data management, the importance of proper metadata handling, open data tools such as Dryad, persistent identifiers to

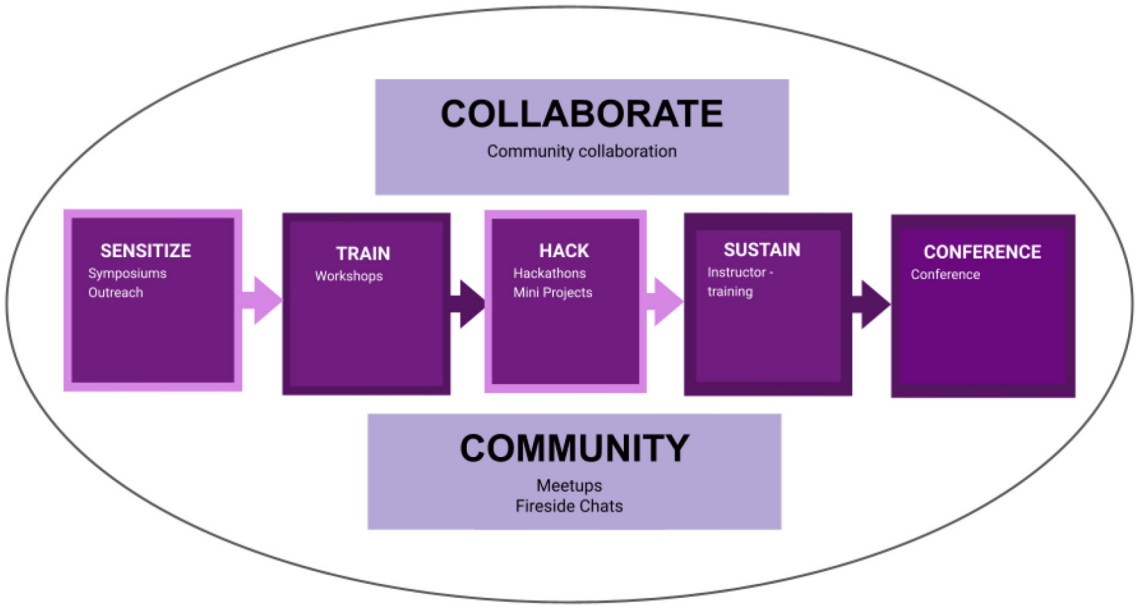


FIGURE 1
The model was adapted from the OpenScienceKE framework and used during the BOSS events. The sensitization phase involved the Open Science FAIR symposium; the training phase involved two parts; the BOSS workshop and Instructor training, the hack phase was done through mini-projects, which also encompassed the aspect of collaboration, and the conference covered the community aspect.

TABLE 1 The topics covered during the 5-day symposium.

Day	Theme
Day 1: 11 October 2021	Open Science
Day 2: 12 October 2021	Research data management
Day 3: 13 October 2021	Project Planning
Day 4: 14 October 2021	Reproducibility in research
Day 5: 15 October 2021	Contribution to open source projects: a practical approach

monitor the use of open data and the need for clear research data management policies. A panel discussion of the day addressed issues such as confidence in sharing data openly due to fear of scooping, a significant hindrance to open science adoption in the Global South. The challenges noted in openly sharing data include ownership issues and misinterpretation of the data.

The project planning session targeted MSc and Ph.D. students, and researchers. The central theme of the session was "open by design research to foster reproducibility and collaboration." These sessions motivated the benefits of reproducible research in promoting collaboration, avoiding misinformation, and ensuring continuity in research. The sessions also provided a foundation to open science through project planning. They highlighted the vast array of open science tools researchers could use in each step of the research life cycle. Amazon Cloud, Jupyter Notebooks, RStudio, Singularity and Docker Containers, Nextflow and Snakemake workflows, Dryad, and Figshare. The symposium also included a practical session to demonstrate how to contribute to open source projects using GitHub to give them an appreciation of the importance of open source in research.

Feedback from the participants after the sessions revealed that the symposium met their expectations, with the discussion on low-cost publication options for the Global South and the research findings and work done to map how open science is progressing in Kenya being a favorite. However, some noted the need for more practical sessions and discussions on open policy.

Train phase: bioinformatics workshops

With participants aware of the need for open science in collaborative and reproducible bioinformatics research, the second phase aimed to empower participants by equipping them with bioinformatics and open science tools introduced in the symposium. BOSS events implemented this phase through a workshop that taught introductory bioinformatics and open science skills. The pre-workshop survey revealed that most trainees were interested in learning new skills to apply to their current and future work (Figure 3).

Of 68 applications, 35 undergraduate and 25 master's, and 7 Ph.D. students. We tailored the workshop content to target audiences with beginner to intermediate skills (Table 2).

Training included two daily sessions, a morning lecture, and an afternoon practical. Participants' feedback from the Symposiums pointed to the need to prepare and share materials early and provide guidance on technical aspects of the program, such as the installation of tools. Before the training, the instructors uploaded the materials to the Canvas Learning Platform (https://www.canvas.net/). The platform allowed sharing of exercises, learning material, and a discussion platform with other course participants. We set aside a day before the main workshop program to assist

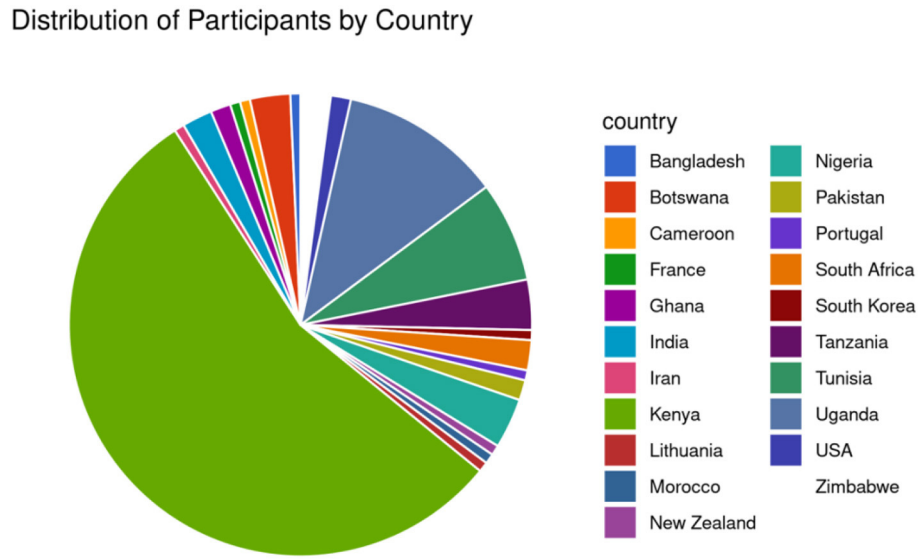


FIGURE 2
The proportion of participants registered for the Open Science FAIR symposium.

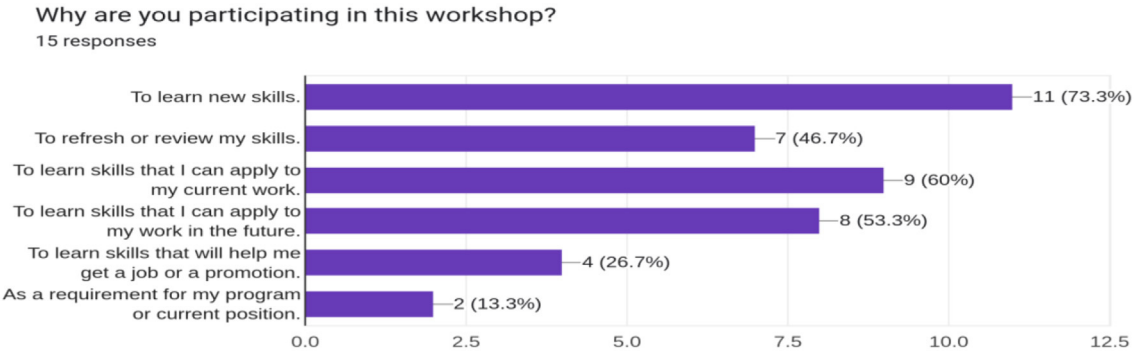


FIGURE 3
The response of the participants to their motivation to attend the bioinformatics workshops.

the participants in setting up workshop material such as GitBash, Ubuntu virtual machine, or a Windows Subsystem for Linux (WSL). We also dedicated an hour before the workshops to answer questions and address any difficulties the participants may have faced the previous day.

After the training, we administered a post-workshop survey to obtain feedback on the sessions. The participants greatly appreciated the programming and practical sessions in the afternoon and requested advanced training workshops. Participants also greatly appreciated training on platforms such as Galaxy since they were not aware of the platform's capabilities. Most participants agreed that the topics covered were relevant to them, the content was well-organized and easy to follow, and the distributed materials were helpful. The time allocated to the sessions was the only challenge for the participants. Although some participants had Internet connectivity problems, 93.3% completed the training. At the end of the training, most trainees were comfortable with the less technical modules, namely sequencing technologies, and familiarization of different data file formats, compared to the more technical ones, namely the Linux command line and quality control assessment (Figure 4).

Hack: mini-projects

This next phase brought together participants to apply skills acquired in the training phase to collaborative projects. We tasked registered participants for this phase to replicate methodologies and reproduce results obtained in selected papers of different research interests. Participants were assigned mini-projects listed in Table 3 to answer research questions using published data. We designed the mini-projects to demonstrate the need for open and reproducible research while imparting technical skills in genomic data analysis, collaboration, and teamwork. We selected published projects whose data met the FAIR principles and reproducible methods. Each of the five projects—plant genomics, viral research, metagenomics

TABLE 2 The training curriculum for the Bioinformatics and Open Science workshops.

Day	Morning session	Time	Afternoon session (1400-1630hrs)
Monday	Intro to sequencing technologies	9:00–10:00 am	• Introduction to High Performance Computing (HPC) • Assignments on data file formats and Introduction to Unix
	Data file formats	10:15–11:00 am	
	Introduction to unix	11:15–12:30 pm	
Tuesday	Advanced Linux, Awk, and Sed	9:00–11:00 pm	Assignments—Unix, Sed, and Awk
Wednesday	Quality control and assessment	9:00–10:00 am	Assignments—QC
	Practical session—QC	10:15–12:00 pm	
Thursday	Sequence alignment	9:00–11:15 am	Assignments—Sequence alignment and assembly
	Assembly	11:30–1:00 pm	
Friday	Introduction to Git/GitHub	9:00–10:00 am	Assignments—Introduction to the Galaxy—Genomics
	Introduction to the Galaxy	10:15–11:15 am	

Morning sessions covered theoretical concepts, while we dedicated the afternoon sessions to practical sessions and assignments.

studies, open science in Africa, and research data management handbook—had 5 participants. We assigned a mentor and expert from various institutions to each group.

We used GitHub to manage the mini-projects from applications, creating issues for each project and assigning tasks to the participants. Of 31 applicants, we selected 25 based on their experiences and motivation; the selected participants came from Kenya, Tanzania, Uganda, and Ghana. The groups selected a project leader, read their respective manuscripts, prepared a presentation of their approach, and virtually presented it to all participants. The groups met their mentors weekly. The participants then extracted data from the databases in the first week. In the following 2 weeks, groups reproduced the analysis and made a final presentation.

For plant, viral, and metagenomic projects, participants faced challenges around the absence of scripts, sketchy details to reproduce the steps, difficulty coming up with custom scripts for analysis given the timeline, and difficulty using the high-performance cluster. The mentioned challenges significantly affected the completion of mini-projects. The groups working on the Open Science in Africa project completed the project on time. Despite the minimal data science skills of the project participants, they managed to pick up from the already existing code and reproduce results for the status of open science in Africa and interpret it. The last project on the research data management did not start as it did not garner a sufficient response from the participants.

A final presentation was made virtually for each project after 3 weeks, where the participants communicated their progress, challenges, and experience of the mini-projects, with two of the groups presenting at the BOSS conference.

Sustain: train the trainer

The "Sustain" phase aimed to increase the capacity of trainers in the region through instructor training. We implemented the phase through instructor training by partnering with Carpentries (Teal et al., 2015), which specializes in training subject experts with experience in coding and life sciences to be instructors. The Carpentries instructor training usually involves two phases: instructor training and checkout to become certified. Twenty-five members were selected to join the Carpentries training. Fifteen participants successfully registered for the training sessions: Four did not participate (poor internet connection), Eight completed the training, and six were certified. They gained skills on how people learn, how to provide feedback, teach, and live coding skills, and finally taught in a demo session for their checkout.

BOSS conference

The Bioinformatics and Open Science Skills (BOSS) virtual conference was the final part of the series of events in the study. The 4-day event themed "Bioinformatics and Open Science Empowerment in an Era of Genomics" brought together researchers and students in bioinformatics, life science, and other fields. Participants joined to present their work, network, and learn from invited speakers on genomics, bioinformatics, open science, careers, and community building. To implement this phase, we formed a 9-member committee to organize the BOSS conference. Planning involved:

- Program design. The committee came up with the conference topics: genomics, open science, research data management, science communication, open science career opportunities, and community development. See the detailed program on the conference website: https://bosscon2022.bhki.org/ and Table 4.
- Speaker selection and communication. One month before the conference, the committee invited experts to give keynotes and talks based on the topics.
- Conference hosting. The committee deliberated over the conference platform needed to accommodate many participants, allow interaction among participants, be easily manageable, and have a low bandwidth requirement. They considered multiple options, including Zoom, BigBlueButton, Aiirmeet, Streamyard, and Eventee, each with strengths and weaknesses. They selected Zoom Events, which is familiar and met the set requirements.
- Conference advertising and registration. The committee used BHKi mailing lists and social networks, primarily Twitter,

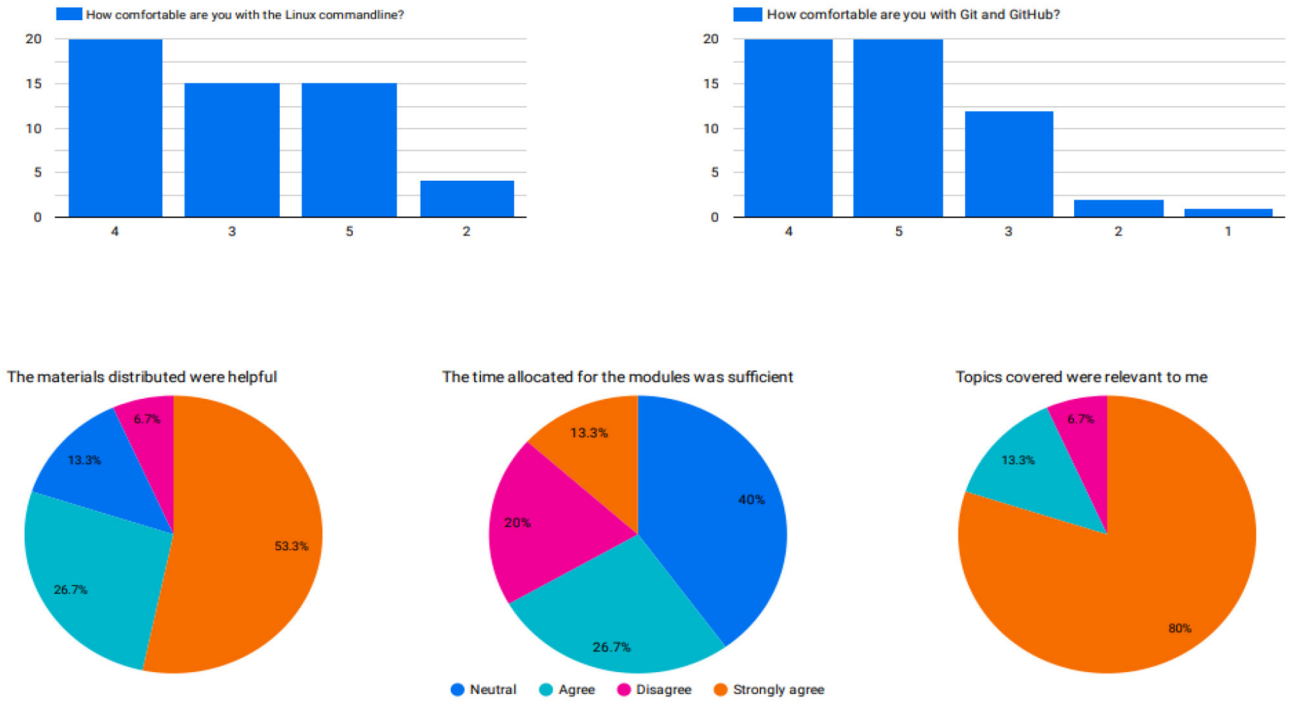


FIGURE 4
(Top) A pre-workshop survey response to assess participants' knowledge of key topics in bioinformatics. **(Bottom)** The post-workshop survey response indicates the satisfaction of participants on selected bioinformatics topics.

TABLE 3 Mini project topics, number of participants, and mentors.

Theme of mini-project	Title of the manuscript	Members	Mentors
Open science in Africa	Open Science in Kenya: Where are we? (Mwangi et al., 2021)	4	1
Metagenomics	Profiling of RNA viruses in biting butterflies (Ceratopogonidae) and related Diptera from Kenya using metagenomics and metabarcoding analysis (Langat et al., 2021).	6	1
Plant genomics	The draft genomes of five agriculturally important African orphan crops (Chang et al., 2019).	5	2
Viral mini-project	Phylogenetic analysis of the 2020 West Nile Virus (WNV) outbreak in Andalusia (Spain) (Casimiro-Soriguer et al., 2021).	7	2
Open science research data management handbook.	Conceptualized by the BOSS organizing team	3	2

to raise awareness of the conference and to call for abstract submissions from researchers and students who wanted to present their projects and register for the conference. Free registration through the Zoom Events platform was open for 3 weeks.

A total of 207 participants registered: over 60% were from Kenya, and the rest were from Nigeria, Ghana, the United States, South Africa, Cameroon, the United Kingdom, India, and Bangladesh (Figure 5B). The education level of the registrants is shown in Figure 5A.

We had 100 participants attending the conference, with the numbers fluctuating throughout, with a minimum of 50 at any given session. The audience was engaged during all sessions using icebreakers, questions, and polls. The lobby attendance feature of the Zoom Events platform proved advantageous, as participants did not need to log in to Zoom and could follow the proceedings from their browsers. The post-conference survey indicated that the conference met expectations, with 35 participants rating the conference 5, 40 at 4, and 3 at 3.

Collaborate and community

The African proverb "If you want to go fast, go alone; if you want to go far, go together" aptly captured the motivation behind the "collaborate and community" phase of our events. The impact of our work to sensitize and build capacity in bioinformatics and open science leveraged the power of collaborations. Collaboration and community are the key pillars of the model we used. This phase involves the engagement of the organizations and individuals to put together resources available to the different parties to achieve

TABLE 4 A detailed 4-day BOSS conference program.

Day 1: open science	Day 2: one health, plant genomics, and reproducibility
Opening remarks (Introduction to the conference)	Keynote: pathogen genomics
Introduction to BHKi and OpenScienceKE	QandA One Health
Icebreaker 1: participants' perspectives of OpenScienceKE and RDM	Plant genomics keynote
Keynote 1: research data management	Q&A plant genomics
Empowered for action: making open science practical	Efforts to identify and combat antimicrobial resistance in Uganda: a systematic review
AfricaRXiv	Reproducibility in the presentation of research projects
Discussion and Q&A	Open science in Africa project presentation
Break	Break/Q&A
GitLab pages for presentation and portfolios	Networking hour: self-assigned breakout rooms
Networking hour: self-assigned breakout rooms	
Day 3: unconference	Day 4: genomics and reproducibility
Careers in the community	Welcome and housekeeping
Science communication presentation 1	Keynote: environmental genomics
Alternative career paths within open science	Q&A genomics
Science communication presentation 2	Reproducibility in research: presentation
Break	Open science in Africa project
TCC Africa and its partners	Break
Working and studying outside Kenya: experiences and opportunities	Leveraging cloud genomics
	Large-scale computational regulatory genomics
	Closing remarks

our shared goal, eliminating the need to reinvent the wheel. Each phase of the model relied on the collaboration of communities. We benefited from communities and organizations such as Human Heredity and Health Bioinformatics network (H3AbioNet), the Carpentries, Training Center in Communication-Africa, Open Life Science (OLS), and individual experts within our networks, as illustrated in the model below (Figure 6).

Discussion

We have applied the sensitize-train-hack-collaborate model in virtual events to sensitize and empower students and researchers in bioinformatics and open science skills. Though previously developed for in-person events (Mwangi et al., 2021), we demonstrate it is possible to adopt it for virtual events. However, some challenges exist, especially for the hack phase, which works best in person. We also demonstrate the importance of blending open science and bioinformatics awareness and training activities for reproducible and collaborative research. The Global North has successfully defined and adopted bioinformatics and open science practices. The recent pandemic demonstrated the power of having trained personnel with solid skills in bioinformatics and open science as researchers worldwide came together to tackle COVID-19 disease, quickly developing a vaccine (Ball, 2020). Adopting open science and increasing collaboration improved the quality and quantity of research output. Aaron Kessler describes the need to define open science, those who benefit from it, and how its adoption or lack thereof constrains its impact (Kessler et al., 2021). The BOSS events contextualized Open science (Chan et al., 2019); we tailored the activities to a resource-constrained community by designing content, training approach, and supporting participants to facilitate their attendance. The Bioinformatics workshops equipped participants with new tools and techniques, which are rapidly evolving and often take time to diffuse to the global South. We note that early career researchers must be aware of the core competencies they need to generate sound research output, where they can find resources to learn and opportunities to apply their skills. Open science and bioinformatics research differ in the Global North and South, mainly due to the availability of resources and funding. With events being designed and implemented by African researchers and students, we avoid "Trickle-down" science, which does not provide region-specific solutions (Reidpath and Allotey, 2019).

By empowering the participant with pedagogy skills through Carpentries instructor training, we increased the pool of well-trained and skilled research workforce with knowledge of the methods to solve challenges in the region and empower others, thus sustaining the impact of the events.

Creating awareness on the need and benefits of a practice and highlighting how it benefits a particular community increases its adoption. Through the FAIR symposium, participants were sensitized on open science practices. Majority of the participants were from Africa, a testament that researchers in the Global South are interested in and embracing open science and need more information on the benefits and demerits (McKiernan et al., 2016), and what these practices look like in the region. Indeed, we noted that the most impactful sessions were presentations on mapping open science progression in Kenya and Africa, which showed that researchers were interested in those who have already embraced these practices. The presentation on the benefits of open science and how more researchers are adopting these practices helped to demystify some of the misconceptions (Kuchma, 2021), which have slowed the adoption of open science. The eagerness to hear and learn more about open policies by a students would enable them to drive changes where they work and impact decisions made by their departments and advisors (Kathawalla et al., 2021), and enable them to collaborate with other researchers, including from the global North (Allen and Mehler, 2019). The session on publishing in low-cost access journals also garnered much interest, especially from the students in the audience. In some postgraduate programs in Kenya, students must publish in journals as part of the degree requirements (Mwangi et al., 2021). However, they sometimes even lack the funds to conduct their research, let alone

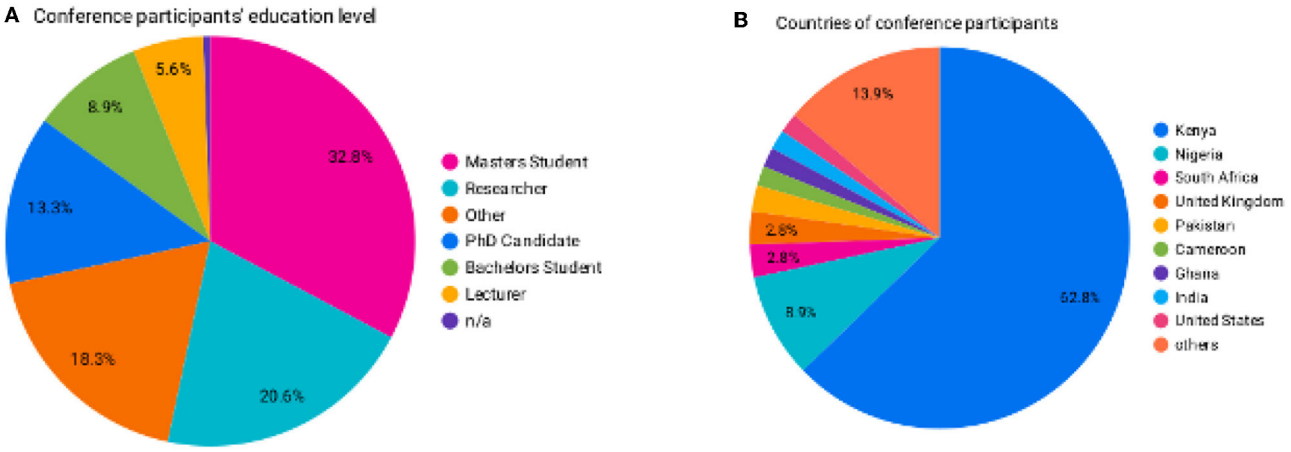


FIGURE 5
(A) The education level of the participants who attended the conference. **(B)** The countries of origin of conference participants.

pay the article processing fees charged by publishers. To ease this burden of publishing, most of the publishers offer waivers to authors from low- and middle-income countries (Mwangi et al., 2021).

The drive to increase the adoption of open science and current bioinformatics techniques in the global North requires empowering them with the skills and tools. For a long time, helicopter science and extractive research has been the norm, where African researchers are only involved in sample collection, with analysis and publication being driven from the global North. However, training activities such as the BOSS workshops, are changing the landscape, generating a pool of trained students and researcher who can adopt open science and bioinformatics in their research. The BOSS workshop audience consisted mainly of undergraduate, and master's students interested in learning new skills and applying them to their work. Undergraduate students have general knowledge, not techniques they can use for analysis therefore hands-on training is crucial to enable them to understand the theoretical concepts they learn. Masters' students are similarly exposed to theoretical content during the first parts of their program and must learn many techniques as they work on their projects (Brazas et al., 2017). Therefore, they are keen to attend such workshops to learn the skills required for their projects. Although teaching is long, broad, and theoretical, workshops delivers skills in relatively short, practical, and focused courses (Schneider et al., 2010). The BOSS workshops applied practical, experiential and project based learning to facilitate greater skills retention, thus empowering the participants to apply the skill in their research (Emery and Morgan, 2017).

After skill acquisition, there needs to be a practical application of these skills where a participant combines knowledge learned and creativity to produce an outcome, thus enhancing skills retention and collaboration (Ahmed et al., 2019).

The BOSS hack phase employed the use of hackathons and collaborative mini-projects. However, we noted challenges in delivering a hackathon virtually, with only one project out of the four selected being completed within the 3 weeks. The participants had different levels of skill and motivation, which

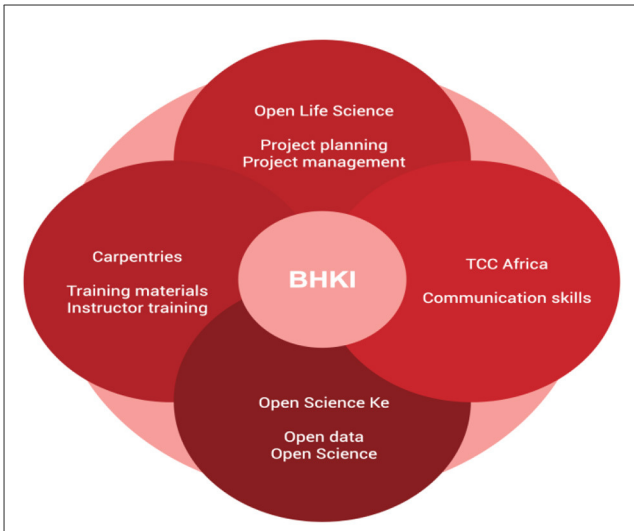


FIGURE 6
The Grassroot Open Science and Bioinformatics communities collaborated during the events of the Bioinformatics and Open Science Skills.

introduced a disconnect in communication (Herrington et al., 2003). They also faced multiple challenges in the initial steps, which proved to be demoralizing and thus affected performance. This phase of the model may have better success through in-person participation, as previously observed (Mwangi et al., 2021). Hackathon and training models developed for resource-constrained regions, should address possible challenges (Jjingo et al., 2022), including offering incentives to participate, such as certification, and support for logistics, such as bandwidth.

Increasing the pool of trainers with pedagogy skills sustains the skills acquisition, and adoption of such practices. Through Carpentries instructor training, the BOSS events enables them to pass on open science practices and tools to new researchers to practice open, reproducible, and collaborative research. There is a need to shift to provide more point-of-need training to equip

those with the skills with teaching skills to competently train and assess learners (Attwood et al., 2019; McGrath et al., 2019). Organizations such as Carpentries, European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) (www.ebi.ac.uk/training/train-trainer), ELIXIR-EXCELERATE (Morgan et al., 2017) have developed "Train the Trainer" (TtT) programs to teach local members and outside their regions, which we leveraged in this study. Training instructors from a pool of already qualified or skilled individuals enable the sustainability of capacity-building programs (Yarber et al., 2015; McGrath et al., 2019).

Conferences present an opportunity for mentorship, collaboration, brainstorming, networking, the connection of early career researchers with established ones, professional development, and an opportunity for co-learning and community building (Lortie, 2020). However, the cost of travel, visa issues, and conference registration costs often limit the participation of researchers from the global South. Therefore, the shift to virtual conferences has increased the diversity of attendees by eliminating some of the barriers of in-person conferences. The BOSS conference had up to 207 registrants from multiple countries and education level, thus enhancing knowledge spillovers (Jaffe et al., 1993). The online nature of the conference also allowed the use of minimal funds to conduct the conference. The audience was also quite diverse from new demographics, allowing participants to connect across borders and disciplines, noting that research and science are inherently transnational cross-border activities. Virtual conferences allow inclusive, accessible, and equitable meetings (Sarabipour, 2020). However, some challenges driven by systemic inequalities persisted, reducing the participation of under-represented communities (Olzmann, 2020). Providing funding for Internet, childcare, and coworking spaces for the BOSS conference improved accessibility and participation. Funding events such as BOSS allowed the elimination of registration costs, which invited more people to participate. Offering incentives such as Internet support, childcare support, and prizes to those attending events also increased participation. Providing these resources increased the participation of students and researchers without funding for conference attendance.

Lessons and recommendations

Despite the challenges, the BOSS events were a great success. We reached a large group of participants with the skills and tools in open science and bioinformatics. Virtual events are low-cost solutions to increasing diversity, representation, and participation in training and conferences. We learned the following lessons from the organization and implementation of virtual events that could benefit other communities interested in hosting events using a similar model.

- Collaborate: we can increase the impact and reach of grassroots communities when they collaborate to leverage time, funds, and expertise.
- Localize and contextualize the events: It is essential to understand the challenges and needs of the community when choosing the content, approach, and technology for the events. Use platforms requiring little training and great features to engage an audience and ensure maximum participation.
- Facilitate participation: Increasing diversity is not a matter of reach but removing barriers to participation. Where resources are available, offer funding to support the internet, childcare, and equipment.
- Continuous evaluation: for a series of events, closely monitor and evaluate the approach, the platforms, the needs, and the impact of the events to facilitate continuous improvements.
- Design for Event Networking: Symposia and conferences are platforms for networking, sharing ideas, and serendipitous encounters that lead to collaborations. It is crucial to design the sessions to include networking hours, chat features, and discussion among the participants.
- Practice what you preach: It was essential to embrace openness by design for open science in the bioinformatics series of events. We shared all materials with an open license and used open-source tools.
- Offer incentives, including certification and prizes, and invite more local speakers in an unconference theme style that participants can relate to, encouraging participation and engagement.

Conclusions

Open science in bioinformatics facilitates reproducibility and collaboration, a skill greatly needed in a heavily computational field. Therefore, it is vital to impart open science skills to bioinformatics students and researchers, especially when training in open science tools accompanies bioinformatics training. Through the Open Science Skills Bioinformatics events, we applied a modified OpenScienceKE framework, which proved to be a suitable model to introduce and train open science and bioinformatics skills. We show that sensitizing and empowering researchers with the skills works best when the participants apply the skills to real-world problems: project-based learning. Furthermore, we have demonstrated how to implement virtual events in resource-constrained settings by providing Internet and equipment support to participants, thus improving accessibility and diversity. Challenges exist, especially with virtual events, such as poor connectivity and high dropout rates. Sharing the recordings after the sessions ensures that participants can catch up on the content, thus leaving no one behind. Virtual events are here to stay, even when the pandemic is over. Therefore, we must understand the opportunities and challenges and tailor the approaches and tools in different contexts.

Data availability statement

The datasets generated for this study can be found in the BHKi GitHub repository: Workshops data—https://github.com/bioinformatics-hub-ke/Boss-workshops while resources for the BOSS; Miniprojects data—https://github.com/bioinformatics-hub-ke/BOSS-miniprojects.

Author contributions

PK wrote the initial draft of the manuscript while being supported and guided by CK, and the rest of the authors made significant contributions to the manuscript. All authors participated in sourcing funding, organizing, and facilitating various activities described in the paper.

Funding

We acknowledge the financial support from the Code for Science and Society events Fund, which made it possible for us to host the Bioinformatics and Open Science Skills series of events.

Acknowledgments

We acknowledge the contributions of all our speakers and workshop facilitators, without whose support these events would not have been possible. We thank our collaborators, the Open Life Sciences, the TCC Africa, PLOS, AfricRXiv and H3ABioNet, and our host institution, the international center of insect physiology and ecology. The contribution of Festus Nyasimi, the chairperson of the Kenya Bioinformatics Hub, is greatly appreciated. The authors thank all the conference organizing committee members, Ruth Nanjala, Laura Ondari, Benson Muite, and Sarah Nyakeri, for their immense contributions to the success of the BOSS Conference.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahmed, A. E., Mpangase, P. T., Panji, S., Baichoo, S., Souilmi, Y., Fadlelmola, F. M., et al. (2019). Organizing and running bioinformatics hackathons within Africa: the H3ABioNet cloud computing experience. *AAS Open Res.* 1, 9. doi: 10.12688/aasopenres.12847.2

Allen, C., and Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biol.* 17, e3000246. doi: 10.1371/journal.pbio.3000246

Aron, S., Chauke, P. A., Ras, V., Panji, S., Johnston, K., and Mulder, N. (2021). The development of a sustainable bioinformatics training environment within the H3Africa bioinformatics network (H3ABioNet). *Front. Educ.* 6, 725702. doi: 10.3389/feduc.2021.725702

Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., and Schneider, M. V. (2019). A global perspective on evolving bioinformatics and data science training needs. *Brief. Bioinformat.* 20, 398–404. doi: 10.1093/bib/bbx100

Ball, P. (2020). The lightning-fast quest for COVID vaccines — and what it means for other diseases. *Nature* 589, 16–18. doi: 10.1038/d41586-020-03626-1

Batley, J., and Edwards, D. (2009). Genome sequence data: management, storage, and visualization. *Biotechniques* 46, 333–336. doi: 10.2144/000113134

Bayat, A. (2002). Bioinformatics. *BMJ* 324, 1018–1022. doi: 10.1136/bmj.324.7344.1018

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., et al. (2021). Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med. Res. Methodol.* 21, 117. doi: 10.1186/s12874-021-01304-y

Bloom, B. S. (1969). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: affective Domain Edition*. Addison-Wesley Longman Ltd. Available online at: https://eclass.uoa.gr/modules/document/file.php/PPP242/Benjamin%20S.%20Bloom%20-%20Taxonomy%20of%20Educational%20Objectives%2C%20Handbook%201_%20Cognitive%20Domain-Addison%20Wesley%20Publishing%20Company%20%281956%29.pdf

Braga, M., Araujo, F., Franco, E., Pinheiro, K., Silva, J., Maués, D., et al. (2021). Bioinformatics on the road: taking training to students and researchers beyond state capitals. *Front. Educ.* 6, 726930. doi: 10.3389/feduc.2021.726930

Brazas, M. D., Blackford, S., and Attwood, T. (2017). Plug gap in essential bioinformatics skills. *Nature* 544, 161. doi: 10.1038/544161c

Casimiro-Soriguer, C. S., Perez-Florido, J., Fernandez-Rueda, J. L., Pedrosa-Corral, I., Guillot-Sulay, V., Lorusso, N., et al. (2021). Phylogenetic Analysis of the 2020 West Nile Virus (WNV) Outbreak in Andalusia (Spain). *Viruses.* 13, 836. doi: 10.3390/v13050836

Chan, L., Okune, A., Hillyer, R., Albornoz, D., and Posada, A. (2019). *Contextualizing Openness: Situating Open Science*. Available online at: https://idrc-crdi.ca/en/book/contextualizing-openness-situating-open-science

Chang, Y., Liu, H., Liu, M., Liao, X., Sahu, S. K., Fu, Y., et al. (2019). The draft genomes of five agriculturally important African orphan crops. *GigaScience.* 8, giy152. doi: 10.1093/gigascience/giy152

Emery, L. R., and Morgan, S. L. (2017). The application of project-based learning in bioinformatics training. *PLoS Comput. Biol.* 13, e1005620. doi: 10.1371/journal.pcbi.1005620

Herrington, J., Oliver, R., and Reeves, T. C. (2003). Patterns of engagement in authentic online learning environments. *Austral. J. Educ. Technol.* 19, 1701. doi: 10.14742/ajet.1701

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Q. J. Econ.* 108, 577–598. doi: 10.2307/2118401

Jjingo, D., Mboowa, G., Sserwadda, I., Kakaire, R., Kiberu, D., Amujal, M., et al. (2022). Bioinformatics mentorship in a resource limited setting. *Brief. Bioinformat.* 23, bbab399. doi: 10.1093/bib/bbab399

Kathawalla, U.-K., Silverstein, P., and Syed, M. (2021). Easing into open science: a guide for graduate students and their advisors. *Collabra Psychol.* 7, 18684. doi: 10.1525/collabra.18684

Kessler, A., Likely, R., and Rosenberg, J. M. (2021). Open for whom? The need to define open science for science education. *J. Res. Sci. Teach.* 58, 1590–1595. doi: 10.1002/tea.21730

Kuchma, I. (2021). *Open Science Myths and Misconceptions.* doi: 10.5281/zenodo.4883635

Langat, S. K., Eyase, F., Bulimo, W., Lutomiah, J., Oyola, S. O., Imbuga, M., et al. (2021). Profiling of RNA viruses in biting midges (Ceratopogonidae) and related diptera from Kenya using metagenomics and metabarcoding analysis. *mSphere.* 6, e0055121. doi: 10.1128/mSphere.00551-21

Lortie, C. J. (2020). Online conferences for better learning. *Ecol. Evol.* 10, 12442–12449. doi: 10.1002/ece3.6923

McGrath, A., Champ, K., Shang, C. A., van Dam, E., Brooksbank, C., and Morgan, S. L. (2019). From trainees to trainers to instructors: sustainably building a national capacity in bioinformatics training. *PLoS Comput. Biol.* 15, e1006923. doi: 10.1371/journal.pcbi.1006923

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., et al. (2016). How open science helps researchers succeed. *ELife* 5, e16800. doi: 10.7554/eLife.16800

Morgan, S. L., Palagi, P. M., Fernandes, P. L., Koperlainen, E., Dimec, J., Marek, D., et al. (2017). The ELIXIR-EXCELERATE train-the-trainer pilot programme: empower researchers to deliver high-quality training. *F1000Research*. 6, 1557. doi: 10.12688/f1000research.12332.1

Mulder, N., Schwartz, R., Brazas, M. D., Brooksbank, C., Gaeta, B., Morgan, S. L., et al. (2018). The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLoS Comput. Biol.* 14, e1005772. doi: 10.1371/journal.pcbi.1005772

Mwangi, K. W., Mainye, N., Ouso, D. O., Esoh, K., Muraya, A. W., Mwangi, C. K., et al. (2021). Open science in Kenya: where are we? *Front. Res. Metrics Anal.* 6, 669675. doi: 10.3389/frma.2021.669675

OECD. (2015). *Making Open Science a Reality*. Paris: OECD.

Okafor, I. A., Mbagwu, S. I., Chia, T., Hasim, Z., Udokanma, E. E., and Chandran, K. (2022). Institutionalizing open science in Africa: limitations and prospects. *Front. Res. Metrics Anal.* 7, 855198. doi: 10.3389/frma.2022.855198

Olzmann, J. A. (2020). Diversity through equity and inclusion: the responsibility belongs to all of us. *Mol. Biol. Cell* 31, 2757–2760. doi: 10.1091/mbc.E20-09-0575

Ras, V., Botha, G., Aron, S., Lennard, K., Allali, I., Claassen-Weitz, S., et al. (2021). Using a multiple-delivery-mode training approach to develop local capacity and infrastructure for advanced bioinformatics in Africa. *PLoS Comput. Biol.* 17, e1008640. doi: 10.1371/journal.pcbi.1008640

Reidpath, D. D., and Allotey, P. (2019). The problem of 'trickle-down science' from the global north to the global south. *BMJ Glob. Health* 4, e001719. doi: 10.1136/bmjgh-2019-001719

Sarabipour, S. (2020). Virtual conferences raise standards for accessibility and interactions. *Elife* 9, e62668. doi: 10.7554/eLife.62668

Sayres, M. A. W., Hauser, C., Sierk, M., Robic, S., Rosenwald, A. G., Smith, T. M., et al. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLoS ONE* 13, e0196878. doi: 10.1371/journal.pone.0196878

Schneider, M. V., Watson, J., Attwood, T., Rother, K., Budd, A., McDowall, J., et al. (2010). Bioinformatics training: a review of challenges, actions and support requirements. *Brief. Bioinformat.* 11, 544–551. doi: 10.1093/bib/bbq021

Teal, T. K., Cranston, K. A., Lapp, H., White, E., Wilson, G., Ram, K., et al. (2015). Data carpentry: workshops to increase data literacy for researchers. *Int. J. Digital Curat.* 10, 135–143. doi: 10.2218/ijdc.v10i1.351

Yarber, L., Brownson, C. A., Jacob, R. R., Baker, E. A., Jones, E., Baumann, C., et al. (2015). Evaluating a train-the-trainer approach for improving capacity for evidence-based decision making in public health. *BMC Health Serv. Res.* 15, 547. doi: 10.1186/s12913-015-1224-2

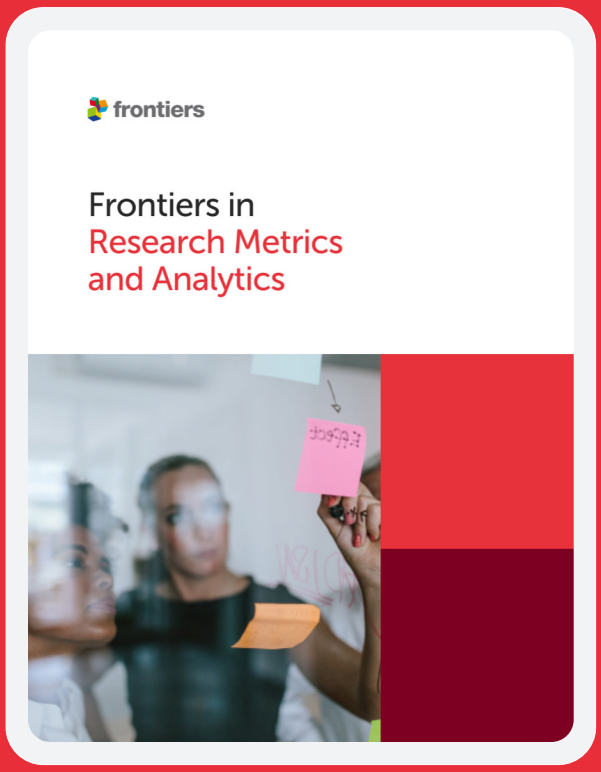
Frontiers in
Research Metrics and Analytics

Explores what drives excellence in methods of analysis

Provides a forum for the study of measuring, evaluating, and improving the efficiency, reliability, and transparency of research and innovation in all areas of scientific inquiry and applications.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact