







RESEARCH ARTICLE

# Considerations for an integrated population health databank in Africa: lessons from global best practices [version 1; peer review: 2 approved]

Jude O. Igumbor <sup>1</sup>, Edna N. Bosire<sup>1</sup>, Marta Vicente-Crespo <sup>1,2</sup>,  
Ehimario U. Igumbor <sup>3,4</sup>, Uthman A. Olalekan<sup>5</sup>, Tobias F. Chirwa <sup>1</sup>,  
Sam M. Kinyanjui<sup>6</sup>, Catherine Kyobutungi<sup>2</sup>, Sharon Fonn<sup>1</sup>

<sup>1</sup>School of Public Health, University of the Witwatersrand, Johannesburg, Gauteng, 2193, South Africa

<sup>2</sup>African Population and Health Research Centre, Nairobi, Kenya

<sup>3</sup>Nigeria Centre for Disease Control, Abuja, Nigeria

<sup>4</sup>School of Public Health, University of the Western Cape, Cape Town, Western Cape, South Africa

<sup>5</sup>Warwick-Centre for Applied Health Research and Delivery (WCAHRD), Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK

<sup>6</sup>KEMRI Wellcome Trust Research Programme, Filifi, Kenya

**V1** First published: 23 Aug 2021, 6:214  
<https://doi.org/10.12688/wellcomeopenres.17000.1>

Latest published: 23 Aug 2021, 6:214  
<https://doi.org/10.12688/wellcomeopenres.17000.1>

## Abstract



**Background:** The rising digitisation and proliferation of data sources and repositories cannot be ignored. This trend expands opportunities to integrate and share population health data. Such platforms have many benefits, including the potential to efficiently translate information arising from such data to evidence needed to address complex global health challenges. There are pockets of quality data on the continent that may benefit from greater integration. Integration of data sources is however under-explored in Africa. The aim of this article is to identify the requirements and provide practical recommendations for developing a multi-consortia public and population health data-sharing framework for Africa.



**Methods:** We conducted a narrative review of global best practices and policies on data sharing and its optimisation. We searched eight databases for publications and undertook an iterative snowballing search of articles cited in the identified publications. The Leximancer software © enabled content analysis and selection of a sample of the most relevant articles for detailed review. Themes were developed through immersion in the extracts of selected articles using inductive thematic analysis. We also performed interviews with public and population health stakeholders in Africa to gather their experiences, perceptions, and expectations of data sharing.

**Results:** Our findings described global stakeholder experiences on research data sharing. We identified some challenges and measures to harness available resources and incentivise data sharing. We

## Open Peer Review

Approval Status  

	1	2
<b>version 1</b> 23 Aug 2021	 <a href="#">view</a>	 <a href="#">view</a>

1. **Ejemai Eboime** , University of Alberta, Edmonton, Canada
2. **Ushotaneffe Useh** , North-West University, Potchefstroom, South Africa

Any reports and responses or comments on the article can be found at the end of the article.

further highlight progress made by the different groups in Africa and identified the infrastructural requirements and considerations when implementing data sharing platforms. Furthermore, the review suggests key reforms required, particularly in the areas of consenting, privacy protection, data ownership, governance, and data access.

**Conclusions:** The findings underscore the critical role of inclusion, social justice, public good, data security, accountability, legislation, reciprocity, and mutual respect in developing a responsive, ethical, durable, and integrated research data sharing ecosystem.

### Keywords

Data sharing, open science, databank, ethics, population health



This article is included in the [KEMRI | Wellcome Trust gateway](#).

**Corresponding author:** Jude O. Igumbor ([jude.igumbor@wits.ac.za](mailto:jude.igumbor@wits.ac.za))

**Author roles:** **Igumbor JO:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Bosire EN:** Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Vicente-Crespo M:** Methodology, Writing – Review & Editing; **Igumbor EU:** Writing – Review & Editing; **Olalekan UA:** Writing – Review & Editing; **Chirwa TF:** Writing – Review & Editing; **Kinyanjui SM:** Writing – Review & Editing; **Kyobutungi C:** Writing – Review & Editing; **Fonn S:** Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This research was funded by the Wellcome Trust (Research Ecosystems in Africa and Asia Priority Area). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2021 Igumbor JO *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Igumbor JO, Bosire EN, Vicente-Crespo M *et al.* **Considerations for an integrated population health databank in Africa: lessons from global best practices [version 1; peer review: 2 approved]** Wellcome Open Research 2021, 6:214 <https://doi.org/10.12688/wellcomeopenres.17000.1>

**First published:** 23 Aug 2021, 6:214 <https://doi.org/10.12688/wellcomeopenres.17000.1>

## Introduction

The public and population health research and development landscape in Africa has seen an increase in publications and the maturation of mostly donor-funded development programmes, research projects and multi-disciplinary capacity building networks<sup>1-9</sup>. These programmes collect and generate data that could be collated, integrated, or triangulated to address the complex and inter-related public and population health challenges in Africa. Health research data collation and sharing programmes are already in place in many high-income countries. Examples include the BigData@Heart platform of the European Union's (EU) Innovative Medicine Initiative<sup>10</sup>, the EU's Horizon 2020 Project and Open Science Cloud<sup>11</sup>, and others<sup>12-15</sup>.

The growth of databanks and repositories has expanded opportunities for data sharing to advance global health. These platforms<sup>16</sup> are setup to generate evidence-driven translation of research<sup>10</sup> which enhance our understanding of and response to public health challenges. This, in turn, can improve public health training and service delivery, and speed up health innovation. Health data integration and use is equally important in strengthening health systems. It can generate evidence-informed solutions; inform the roles and choices of patients and service providers; spur discovery to improve patient care; and help evaluate the outcome of health services and health capacity and research building programmes<sup>17</sup>.

Despite the improvements of the last decades, Africa still lags behind in research and development - contributing less than 2% of global research output<sup>18</sup>. While the reasons are manifold<sup>19</sup>, the situation is compounded by the lack of (or limited) African-led databanks or data repositories platforms. This hampers data sharing, reuse, integration, meta-analyses, and cross-referencing. Digitisation, integration, and information sharing may allow Africa to generate knowledge more rapidly to address its public health challenges.

A vision of an African integrated databank is mindful of related challenges. These include data privacy, malicious use of data, complexities of regulating digital information, fragmented privacy regulations and jurisdictional nuances, and lack of acknowledgement of researchers and scientists<sup>20-25</sup>. Additionally, conventional informed consent and human research ethics committees (RECs) must consider emerging issues of data stewardship such as the longer storage, sharing, re-identification and indeterminate future use of collected data<sup>26-30</sup>.

The main objective of this article is to provide practical recommendations and requirements to support the development of a multi-consortia public and population health data sharing framework for Africa. This research seeks to inform a platform that will harnesses available resources, incentivise data sharing, and optimise the progress made by different research groups in Africa. The review draws on a collection of global best practices and policies. With this research, we address the challenges and misconceptions of data sharing in Africa. The collection of global stakeholder experiences on research data sharing presented here offers essential discussion points

for consideration in developing an integrated population health databank in Africa. This article, therefore, targets all who are impacted by research data sharing or stand to gain from an understanding of the key tenets to consider when sharing research data in the context of privacy, confidentiality, information security and respect of human data and biological specimens.

## Methods

### Narrative review

We undertook a narrative review of publications and policy documents on data sharing in public and population health.

The methodological standards of narrative reviews described by Greenhalgh *et al.*<sup>31</sup> and noted as best suited for exploring broad and complex topics using a constructivist philosophy<sup>32</sup> were followed. Inclusion of policy documents in this review is a common practice under these circumstances<sup>33</sup>. Inclusion of policy documents is also informed by the strong policy foundation of the topic, and the expectation that this review may inform future policies on data sharing. We searched eight databases for publications, namely PubMed, EMBASE, PsycINFO, Joanna Briggs, The Cochrane Library, EBM reviews, Scopus, and Web of Science. We did not set any time frames so as to include historic patterns, which may inform current data sharing practices. Our data search included all articles related to "population health data sharing" and "public health data sharing". We also followed-up articles cited in the papers we identified in our initial search to ensure relevance of the review to our target audience<sup>31,34</sup>. The search process was, therefore, an iterative snowballing exercise.

Our initial search identified 3825 articles that were loaded into Mendeley to remove duplicates. Two independent reviewers (JOI and ENB) evaluated the title and abstract of each article to assess its relevance for inclusion in our review. This approach did not rely on a pre-defined keyword search to identify conceptually and empirically relevant documents. Any disagreements between the reviewers were resolved through discussions among the review team. We followed a qualitative appraisal based on principles of pragmatism, pluralism, historicity, contestation and reflexivity<sup>31,34</sup>. At the end, we identified 655 documents for further review.

The Leximancer software © Version 5 enabled content analysis and selection of a sample of articles for detailed review<sup>35,36</sup>. Leximancer like alternative software (such as Nvivo and MXQDA) are all paid-to-use software with limited trial period. Leximancer identifies lexical co-occurrence of natural language into semantic patterns<sup>37</sup>. It is reproducible and uses an unsupervised machine learning model that is built on Bayesian Theory to predict events based on an observed pattern<sup>35,37</sup>. Leximancer identified seven core themes from the 655 articles selected. We extracted and reviewed articles with the highest co-count and likelihood of containing each theme in their segments. We selected as many as 20 articles per theme based on our reaching saturation after reading on average, the top 15 articles. Our selection of articles also involved full-text screening.

## Interviews with key informants

To ensure that our approach to the literature addressed the concerns and questions of local African stakeholders, we had interviews with 35 key informants from African-led research and capacity building programmes who produce population and public health data that could be included in a shared database. To identify these consortia we took advantage of the range of African-led programmes funded by the Alliance for Accelerating Excellence in Science in Africa (AESAs)<sup>38</sup>. Participants were purposively sampled, which created a diverse group, ranging from basic science and genomics to applied translation science. In-depth interviews of about 60 to 90 minutes were conducted virtually using Microsoft teams. We used an open-ended guide (see *Extended data*<sup>39</sup>) to facilitate the interviews, but the discussions were flexible, with the interviewee responses shaping the discussions. We obtained a written consent to participate in the interviews from the participants. Eleven out of the 35 participants declined being recorded and notes were taken during their interview. Twenty-four interviews were audio-recorded and transcribed, and summary of emerging themes were discussed with the participants at the end of each interview. Summaries from all interviews were compiled into key themes and sub-themes. The findings of interviews presented in this paper are highly consolidated and pose no risk to the expert informants interviewed; therefore, ethical approval was not required to be obtained.

In all, the views expressed in the paper are completely based on review of literature that is available in public domain. The informal and internal consultations with network peers that constituted the interviews were used to position our findings. The consultations were also to ensure the literature review's regional relevance, and to promote objectivity and reflexivity in our analysis and interpretation of findings. The interviews, literature review and initial analysis were conducted by two of the authors (a male and a female) with PhD in Public Health and Medical Anthropology, respectively. They have training and experience in qualitative research, ethics, epidemiology, and data science.

## Results

The outcome of the interviews framed our approach to the meta-synthesis in the narrative review. Key observations from these discussions indicated a strong interest in research data sharing; inadequate awareness and misunderstanding of the ethical, legal, and social implications of data sharing; and pervasive data sharing between researchers based on professional and social networks. We also observed the respondents' perceived lack of capacity for secure and responsible data sharing in the region; notable data access challenges; misconceptions of funders' expectations of data sharing; strong fear of data misuse and exploitation; concerns about insufficient regulation and governance; and inadequate incentives and acknowledgment of data custodians.

Our analysis of the document review suggested five overarching themes: (a) Data sharing context; (b) Laws, regulations, and oversight; (c) Enablers of data sharing; (d) Governance and value-based implementation; and (e) Data infrastructure, quality, storage, and security.

Below, we present global best practice under each of the themes and discuss this in relation to the findings from our interviews with the 35 African researchers, research administrators and ethics committee members. We conclude by making recommendations to support the establishment of an integrated population health databank in Africa.

## Data sharing context

**Databanks and standards.** Databanks or data repositories are being established globally. Notable public health database programmes feeding into repositories in the Global South include the USAID-funded Demographic and Health Surveys (DHS)<sup>40</sup>, UNICEF's Multiple Indicator Cluster Surveys (MICS)<sup>41</sup>, the International Network for the Demographic Evaluation of Populations' (INDEPTH's) Health and Demographic Surveillance System (HDSS)<sup>6</sup> and Human Heredity and Health in Africa (H3Africa)<sup>42</sup>. These platforms offer best practice standards for data sharing. The Public Population in Genomics (P3G) consortium is another global best practice model whose vision is to increase the power of analysis and discovery through greater integration. Similar and complementary protocols are available from Genome-Wide Association Studies (GWAS) Policy and the database of Genotypes and Phenotypes (dbGaP)<sup>43–45</sup>.

Lessons from genomic biobanks offer guidance on starting up future databanks<sup>46,47</sup>. These include ensuring sustainability, managing jurisdictional obstacles, governance, quality management, material transfer agreements, use of technology and intellectual property<sup>47,48</sup>. Our findings are cognisant of nuanced and substantive differences in data types and variations in the ethical and legal contexts of these data.

Africa does not have the kind of robust, integrated databanks or data repositories present in most of the developed world. But there are opportunities to integrate existing data platforms. There is a spread of health and demographic surveillance system sites, routine national surveys, priority disease specific registries and databases, and the proliferation of genomic data repositories in the region<sup>6–9,42</sup>. Other examples include routine DHS, large scale donor funded research and/or development programmes across the continent, country specific survey and administrative datasets, and data emerging from the Developing Excellence in Leadership, Training and Science in Africa (DELTA Africa) programme.

INDEPTH – one of the oldest data platforms in Africa offers good data sharing practices. It provides potential to collate data from member HDSS sites into outputs that enable systematic comparisons<sup>6</sup>. Another example is the H3Africa programme which provides exemplary lessons for an integrated African databank<sup>42</sup>. The H3Africa consortium conducts biannual research priority setting and regular review of operational policies, guidelines, and logistics. These measures are essential for standardisation and quality assurance<sup>42</sup>. In all, Africa has pockets of quality data that may benefit from greater integration.

**Perceived challenges, risks and considerations for data sharing.** Individual willingness to share data is mediated by sociodemographic status, cultural and religious factors<sup>49–54</sup>.

For example, younger people and females are less likely to participate in consenting to data reuse<sup>55</sup>. Fears of loss of privacy or confidentiality breach, commercialisation of data, misuse and abuse are equally concerning<sup>56-59</sup>. These concerns are also driven by insufficient public engagement and low public awareness of research governance, participant protection and risk minimisation measures<sup>54</sup>. This leads to minimal public appreciation of the importance of health research.

Poor communication and use of technical terms may breed mistrust and impede participation and willingness to permit data sharing<sup>60</sup>. The use of language and analogies that are sensitive to the context of research could improve communication and understanding<sup>61</sup>. In addition, studies have raised concerns about participants' understanding, and the quality and extent of information participants should have in order to make informed decisions<sup>62,63</sup>. To deal with this problem, authors recommended improving study participants' knowledge of data sharing<sup>61,63</sup> with tools such as videos<sup>64</sup>, pictures<sup>65</sup> and vignettes<sup>66-69</sup>.

Beyond research participants, our findings highlight that scientists are concerned that the risks of data sharing might outweigh the advantages. This perception is driven by the fear of possible loss of academic advantage and independence; the possibility of their work being misused, misinterpreted or misrepresented; the loss of intellectual property; and an increased workload for administration and data management<sup>70</sup>. If these issues remain unaddressed, the practice of data sharing will remain a dream in Africa. Major funders of public and population health research in Africa expect that data sharing should be the norm<sup>71-77</sup>. In most cases, funders provide global tools for sharing data<sup>78,79</sup>. We, however, found no evidence of donor support in terms of financial resources, capacity building or infrastructure to facilitate an African integrated interdisciplinary data custodial and sharing mechanism.

Other important risks of data sharing include concerns of data quality; poor curation and indexing of datasets; variations in data provenance, metadata and management protocol with implications for data comparison and integration of datasets and databases<sup>80</sup>. Most of these challenges may be addressed through rich collection of metadata of each data set<sup>80,81</sup>.

Relatedly, trust in databanks<sup>82</sup> is dependent on the perceived trustworthiness of the data custodian<sup>83-85</sup>, use of minimum set of information provided<sup>84,86-89</sup>, and the promise of, and belief that privacy will be maintained<sup>84-87,89</sup>. Without these elements there is no public trust.

Factors affecting public attitudes to data sharing have been summarised as sensitivities, controllability, benefits, risks, governance and public attitude<sup>53</sup>.

Internal policies, collaborative agreements and contracts within research networks and specialised fields of public and population health govern data access and sharing are essential elements of data governance<sup>90</sup>. These instruments are, in part, designed to mitigate some of the challenges.

## Laws, regulations, and oversight

**Data protection laws.** As of 2018, only 19 African countries had privacy protection laws<sup>91</sup>. Six others (Kenya, Nigeria, Togo, Tanzania, Uganda and Zimbabwe) had laws in draft stages. An analysis of the privacy protection laws across the continent classified almost all of these laws as moderate to limited<sup>92</sup>. Whatever differences may exist between countries, within-country variations in privacy regulations is equally common<sup>93</sup>. Consequently, countries have developed mechanisms to facilitate lawful application of their, often conflicting and fragmented, privacy regulations<sup>24</sup>.

For African countries without privacy protection regulations, there are global models to explore. These include the UK Data Protection Act of 2018<sup>94</sup> (see principles in **Box 1**) and examples from the African continent<sup>92</sup>. These tools give individuals control of their data through their right to informed consent<sup>56</sup>. They also stipulate special protection for certain types of data including genetic and biometric data<sup>95</sup>.

**Ethics committees.** Ethics committees include research ethics committee (REC), biomedical research ethics committees (BREC) or institutional review board (IRB). In this article, we use the term research ethics committee (REC). These are multidisciplinary, independent groups of individuals appointed to review proposed studies with human participants. The

### Box 1. UK data sharing principles

1. Personal data shall be processed fairly and lawfully and shall not be processed unless – (a) at least one of the conditions in Schedule 2 is met, and (b) in the case of sensitive personal data, at least one of the conditions in Schedule 3 is also met.
2. Personal data shall be obtained only for one or more specified and lawful purposes and shall not be further processed in any manner incompatible with that purpose or those purposes.
3. Personal data shall be adequate, relevant, and not excessive in relation to the purpose or purposes for which they are processed.
4. Personal data shall be accurate and, where necessary, kept up to date.
5. Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.
6. Personal data shall be processed in accordance with the rights of data subjects under this Act.
7. Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.
8. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data

Source: Government of UK Legislation. Data Protection Act 2018. <http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>.

REC<sup>96</sup> must ensure respect for participants; beneficence, as well as justice by protecting their rights, safety, and well-being.

The composition, structure and requirements of RECs vary between countries. Some countries require additional permission or registration to conduct research. However, RECs have a role to play in the transfer of data to a third-party institution by ensuring compliance with data control regulations and privacy protection policies.

Yet, in many countries, RECs are confronted with numerous challenges including lack of legal protection<sup>97</sup>, inability to reach quorum in decision making, inappropriate constitution of REC<sup>97,98</sup> and inefficiency or bias amongst its members<sup>99</sup>. In addition, the growing scope of social implications of data sharing often falls outside the responsibility of RECs whose adjudication is based on presented intention of a particular research project without detailed consideration of broader social impact of the research<sup>50,100,101</sup>.

Fortunately, there are a number of global guidelines to rely on for direction even if most RECs have not kept up with recent developments in research and technology. The Helsinki Declaration remains a major reference document for data security, ethical principles and governance of data sharing<sup>102</sup>. Others include the Australian Guidelines on Human Biobanks and Genetic Research Databases<sup>103</sup>; The OECD Principles and Guidelines for Access to Research Data from Public Funding<sup>104</sup>; the Bermuda Principles<sup>105</sup>; and the Expert Advisory Group on Access (EAGDA) report on Data Access<sup>106,107</sup>. Similar tools have been developed in parts of Africa<sup>108</sup>.

**Consent.** Informed consent is the cornerstone of ethical conduct and regulation of research. Increased digitisation of health data has resulted in easier access to data, and data integration facilitated by greater connectivity via the internet<sup>80</sup>. This calls for more attention to the ethical and legal implications<sup>109</sup>. The universally applicable guidelines for consenting involves three key features: (a) of information to potential research participants needed to make an informed decision; (b) facilitating the understanding of what has been disclosed; and (c) promoting the voluntariness of the decision to participate or not in the research and ensuring respect for participants. Ensuring that the informed consent process fulfils these three requirements can go a long way towards mitigating problems.

For data to be shared for further future use, RECs need to issue waivers permitting the use of de-identified data or broad consent from research participants<sup>110</sup>, as well as contending with emerging considerations of data stewardship such as the longer than usual data storage, sharing, re-identification and indeterminate future use of collected data<sup>26-30</sup>. These approaches have their limitations. For instance, the proliferation of data sources and hubs increases the risk of unlawful re-identification. Different consent options are described in detail in terms of their benefits and risks by Peppercorn *et al.*<sup>111</sup>.

Dynamic consenting allows research participants to opt-out or opt-in at different stages of the research after the original informed consent was issued<sup>112-115</sup>. On the other hand, broad consent impede participants' control of their data<sup>116</sup>. From the participants' perspective, realistic measures to allow dynamic consenting should be detailed in the original consent. Re-contacting participants should of course, follow standard ethical principles including options on communication of findings or participant access to data<sup>117,118</sup>.

Further, it has been suggested that the respect accorded to study participants or groups during primary data collection should be maintained in secondary data storage, sharing and reuse. Elements of respect include privacy protection and confidentiality; autonomy; data security; respect for individuals and group rights; ensuring dignity of participants; and, protection of life, wellbeing and welfare<sup>10,102,112,119</sup>. In this regard, any further use of data should be in line with the scope of original informed consent provided by the research participants. To mitigate likelihood of unknown future use, authors have pointed out that participants must be subjected to appropriate informed consent as discussed above. In the case of specific consent, the intention of the research is clearly stated at the time of data collection including likely future use of the data<sup>112,114</sup>. In the absence of this certainty at the time of data collection, broad consent may be adopted with conditions to protect the research participants<sup>112-114</sup>. Such protection may be offered by RECs or data access committees. It is still incumbent on researchers to provide as much information as possible when broad informed consent is solicited.

Reaching a consensus on data sharing practices and data reuse has not been systematically addressed, particularly in Africa. Other important yet unaddressed issues include public views or perceptions of cross border data transfer<sup>120</sup>. The differences in jurisdictional powers of national governments and other oversight institutions such as RECs seem to be part of the impediments. Other considerations for the deployment of a data sharing platform include identifying data sources/patterns, engagement with leaderships, ethical and regulatory compliance, data management and legal conditions<sup>121</sup>.

Ethics waivers have been given for data reuse in circumstances where it is impossible to obtain informed consent<sup>102,112,114,117</sup>. The RECs determine the reasonability of circumstance for waiver<sup>117,122</sup>. Such waivers should preclude secondary use of data where participants are identifiable<sup>123</sup>. A common example may include the request for ethics waiver to use medical records of readily accessible and regular users of health services such as patients on chronic treatment. Others have cautioned against the negative psychosocial implications of re-contacting people to consent including deceased family members or reliving a past trauma or unintended breach of privacy<sup>120</sup>. Additionally, researchers have argued that data collected with public funds during routine service provision should be maximised for public benefit and so support such waivers<sup>124-127</sup>. Generally, many have favoured use of aggregated data when

individual consent cannot be obtained. In this context, the impact on groups or communities should be considered and similar group anonymity should be ensured if necessary<sup>128</sup>. On the other hand more stringent measures to obtain ethics waivers have also been recommended<sup>55,129–131</sup>.

**Data ownership and custodianship.** Data ownership is very contentious especially when it comes to sharing the data. The data may be held by an individual scientist or collaborative teams; manually or digitally collected or generated; and stored locally or in shared repositories<sup>132</sup>. Other aspects may be related to individuals involved in data collection, and those who store and share data. Interview with DELTAS Africa consortium stakeholders revealed a wide range of perceptions on the issue of data ownership. Many consortium stakeholders argued that the funding bodies were the owners of data and had the responsibility of deciding when and how data should be shared. Others argued that the principal investigators, researchers, governments, or academic and research institutions were primary owners of these data. Few participants, including members of RECs perceived data ownership to encompass study participants and communities where studies are conducted. Given the complexity of data ownership, and that many stakeholders can mount logical arguments as to ownership, scientists have recommended non-exclusive ownership of data. They submit that data ownership should be governed by legal and moral obligations including trust and custodianship with variations in the right of access and utility by different stakeholders<sup>133–135</sup>. They have argued that data ownership should be based on national privacy regulations and permission granted.

**Intellectual property rights.** Closely linked to the issue of data ownership is intellectual property rights. Many researchers we had interviews with voted in support of a system that recognises researchers' or scientists' contributions and their further involvement in the use of their data if possible. Ultimately, it has been argued that this procedure should be guided by local intellectual property laws<sup>104,114,136</sup>. Similarly, databank users are required to report back to the custodians of the databanks all publications and patents emanating from the data provided to them<sup>107,117,119</sup>.

Authors of the reviewed documents have suggested that data sharing and implementation of databanks should be based on the principle of distributive justice by optimising benefits to society, minimising harm and equitable beneficence related to accessing data and emergent health innovations<sup>10,47</sup>. This proposition invokes the principles of transparency and equity by ensuring that benefits are shared as broadly as possible, especially when dealing with vulnerable populations<sup>114,117</sup>. Benefit sharing is extended to include equitable and fair access to the databank. Most databank policies are, however, not limited to non-commercial use given that some commercial uses are aimed at creating public good and the distinction will determine access.

#### Enablers of data sharing

**Trust and transparency.** Gaining and ensuring the trust of individual research participants and the public has been

described as an essential element in building and maintaining databanks<sup>10</sup>. Trust is a by-product of different principles of good research ethics including clear consultations, open communication and recognition of the individual's autonomy<sup>137,138</sup>. In the case of big databanks, authors have suggested that these attributes should be on-going and not a one-time checkbox activity. Maintaining public trust facilitates benefit optimisation, promotes respect, mitigates harm, and enables social justice and priority setting. Trust may be derived from involving the participants and civil society representatives in the design, governance, knowledge translation and beneficence of the databank output<sup>139</sup>. The engagements should also be cross cutting to involve other researchers, policy makers and funders<sup>112,113,140,141</sup>.

Transparency helps to build trust and accountability and may be achieved by allowing inclusive stakeholders access to policy, guidelines, and data sharing operations. Research participants expect a transparent platform to be clear about how data will be shared and with whom<sup>53,142</sup>, the type of research that is to be performed<sup>143</sup>, by whom the research will be performed, information on data sharing and monitoring policies and database governance, conditions framing access to data and data access agreements<sup>144–146</sup>, and any partnerships with the pharmaceutical industry<sup>147</sup>. Patients and research partners are also interested in knowing how involved patients and other human rights advocacy groups will be in providing oversight and supervision of the platform to ensure unbiased access and use of the databank<sup>148</sup>. Transparency may be enhanced by keeping and communicating sufficient records of operational activities including audits logs and trails<sup>86,87,149,150</sup>; notification of study participants when records are accessed<sup>84,86,151</sup>; operating a decentralised data storage system<sup>87</sup>; and use of data for only specified and agreed purpose<sup>86–88,152</sup>.

**Stakeholder and community engagement.** The success of data storage and sharing is dependent on inclusive stakeholder engagement<sup>10</sup>. Engagement facilitates fair negotiation and consensus on thorny issues. Authors recommend that community engagement should start at the beginning of the project. While our list is not exhaustive and may vary with the type of research conducted, some of the key stakeholders to consult or engage with may include the study participants or patients, civic organisations and leaders, government departments heads of relevant parastatals and nongovernmental organisations, academic research administrators, ethicists, established researchers, graduate students, industry representatives, human rights lawyers, clergy, and traditional leaders.

Stakeholder consultation is an important strategy to promote other essential elements of data storage and sharing such as equity, trust, transparency, autonomy and participation<sup>10,109,153</sup>. For example, H3Africa provides a framework for community engagement<sup>154</sup>. The key components in this framework include defining the goals of engagement; defining "the community" or "the public" in research; identifying strategies, models, and methods for community engagement (e.g., consulting gatekeepers, community meetings); identifying who will do the engagement as well as outlining the role and expectations of community engagement.

The Tikanga Framework of New Zealand, aimed at including Maori People in decisions regarding the use of their data, is an example of a flexible system that is responsive to the material circumstances of its target population<sup>96</sup>. Databanks may need to tailor-make their standard operating procedures to address the unique needs of specific groups<sup>155</sup>. It is important to ensure continuous and appropriate interaction with stakeholders.

Engaging marginalised and vulnerable populations is one of the cornerstones of developing an effective databank. Therefore, measures to promote greater participation of these groups are recommended<sup>156</sup>. In addition to the importance of trust, it is suggested that improving the relationship with the public enhances their disposition to information and sample sharing, minimises common concerns and increases public participation<sup>157</sup>. Consequently, authors have recommended that from the onset of projects, researchers should have a clear plan to involve their target community in the development of the implementation and accountability measures including opportunities to learn about the databank, measures to regularly update the public and ways of addressing concerns about the databank<sup>157</sup>.

**Incentivisation of data contributors and users.** In reality, scientist are not as forthcoming with their data as expected<sup>158–163</sup>. Similarly, there are divergent views on the extent of data sharing among researchers and reported variations are contingent on career ranking and years of experience<sup>159,164</sup>. This difference may be associated with professional disciplines. In life sciences, geneticist are more likely to deny others data when compared to non-geneticists<sup>160</sup>. This is due to variances in intra-disciplinary data collection protocols, sharing requirements and expectations. Nationality of researchers was also a factor likely to effect the prevailing local data sharing culture<sup>159</sup>. Some of the reasons why scientists withhold data include funding agreements, collaborative agreements, data sensitivity, privacy, giving up chance to publish, public critique, lack of data repositories and the absence of consent to share<sup>160,165</sup>. The scepticism about the benefits of data sharing is also common among researchers. Furthermore, researchers in low resources countries fear that their data will be exploited by better resourced scientists<sup>161</sup>. Others view data sharing as a threat to intellectual property, professional value and economic benefits<sup>166</sup>. The greater value placed on publications by institutions has the potential to discourage data sharing<sup>164</sup>.

Best practice solutions suggested by authors include human capital and infrastructural development, and financing to promote research data sharing<sup>165,167–169</sup>. Tangible reward in the form of reputational incentives and peer recognition including citation may promote data sharing<sup>158,170</sup>. Increasing visibility of open access data may also promote sharing<sup>158</sup>. Additionally, creating incentives in the form of rewards may promote data sharing by scientists<sup>46,158,171</sup>. One example is the Cochrane-REWARD prize for reducing waste in research<sup>172</sup>.

Data sharing may be more effective if it is a requirement of the funding agreement. This is particularly important as African

scientists view funding agreements as an obstacle to data sharing. Nevertheless, this view is contrary to the expectations of most funders of research in Africa<sup>72–77</sup>. A public list of funded entities and the data they hold could be made available to promote data sharing and reuse. Policy enforcement may not be sufficient to ensure data sharing and there is need to for a cross-institutional community of practice to promote collaboration and sharing<sup>71</sup>.

Network and co-citation analysis may be used to promote the visibility of available datasets to scientists working in similar fields. Such efforts should be supported with a clear policy that addresses the concerns of all stakeholders, including monitoring and reward mechanisms<sup>161,173</sup>.

Nomenclature, metrics, and weighting of data source citation like citation of peer reviewed publications should be considered. This proposition resonates with the San Francisco Declaration on Research Assessment<sup>174</sup>. Further recommendations of how this may be realised are described by Jones *et al.*<sup>71,175</sup>, including the recommendations of DataCite Collaboration<sup>176</sup>. Additional guidance is provided by the Joint Declaration of Data Citation Principles (JDDCP)<sup>177</sup>.

Promoting international collaborations and publications may be seen as added incentives, as it may unlock global recognition and additional funding opportunities<sup>178</sup>. Lastly, open data badges are the only known tested intervention to improve data sharing<sup>171,179</sup>. Expressly, evidence on effective rewards for data sharing remains unknown and under explored.

**Funders' and researchers' position.** Findings from our interviews with African stakeholders showed that most researchers or scientists in Africa were hesitant to share their data largely due to lack of awareness of the benefits of data sharing, similar to findings from reviewed documents. We also found that many researchers, especially in low-and-middle income countries (LMICs) fear of loss of academic advantage/independence; and the possibility that their work may be misused, misinterpreted or misrepresented among many other reasons<sup>161,166</sup>. Some consortium researchers also believed that research funders restricted them from sharing data. Contrary to such beliefs, the Wellcome Trust presents a summary of funders' statements on data sharing as it "expects all of its funded researchers to maximise the availability of research data with as few restrictions as possible"<sup>180</sup>. The summary excluded the more recent USAID's Policy on Development Data<sup>181</sup>, which purports that "data, and the information derived from data, are assets for USAID, its partners, the academic and scientific communities, and the public at large. The value of data used in strategic planning, design, implementation, monitoring, and evaluation of USAID's programs is enhanced when those data are made available throughout the Agency and to all other interested stakeholders, in accordance with proper protection and redaction allowable by law". As such, we recommend proactive advocacy to ensure that the concept of data sharing becomes a mainstream consideration in national discussions of research management and governance<sup>70</sup>.



The above issues may be amenable to the roles and functions of RECs as an unbiased and value-based entity to arbitrate lawful and moral use of data. However, there were questions about whether most members of African ethics review boards are familiar with the concept of data sharing amongst other ethical issues discussed such as broad consenting. This is similar to what we found in our interviews with DELTAS Africa members including REC members. REC participants recommended that their members be trained and provided with opportunities to attend workshops or other platforms that can expose them to new trends on data and data sharing.

### Governance and value-based implementation

**Policies and values.** Most guidelines and regulations in Africa do not provide clear guidance on governance and how data and biological specimens ought to be shared<sup>182,183</sup>. This is particularly critical given that the different actors involved in data sharing may have different perspectives on data. For example, research participants may be concerned about confidentiality, how the data will be used, and how they might benefit. On the other hand, data collectors may want to produce high-quality data, while data users aim to advance science and inform policies. Clear examples can be borrowed from the UK, USA and Canada. All regulations offer opt-out options when using data for research other than the original intention it was collected for, with the UK National Data Guardian's recommendation being more stringent<sup>24,184</sup>. The European Union General Data Protection Regulation of 2016<sup>185</sup> has also been hailed as an effective framework to facilitate regional harmonisation<sup>24</sup>. Sector-specific guidelines have been recommended to promote pragmatic compliance with policy.

Given such differences, there is need for data sharing policies to state clearly when, where, how and which data should be archived and made available.

Lack of clear policies on data sharing may frustrate researchers who want to share data, and provide loopholes for those who are unwilling to share. Thus, in the absence of absolute privacy protection, risk minimisation is the best alternative<sup>58,186</sup>.

Awareness of risks did not always affect willingness to share data when such risks were weighed against expected benefits<sup>53</sup>. Hence, willingness to share data was more likely to become a factor of "privacy – utility trade-off"<sup>187</sup>. Similarly, most privacy protection regulations do not consider privacy as an absolute right of an individual but contingent on its intersection and weighting against other rights<sup>24</sup>, for instance, the imperative to report a notifiable disease or in case of the safety of children and vulnerable people<sup>188</sup>.

Greater integration also poses risk of re-identification, which infringes on participants or patient privacy protection and trust. This is a major concern for people who share data<sup>57,58,133</sup>. Likewise, the willingness to share data decreased with increase in privacy and confidentiality concerns<sup>52</sup>. Criminal prosecution for negligence or wilful breach of privacy as stipulated by

national laws should be considered. Various recommendations for privacy protection have been made including creation of clear laws to govern re-identification, and stronger sanctions and corresponding enforcement protocol for misuse of data<sup>133,189,190</sup>. The use of data without following due process or attribution should be condemned<sup>46</sup>. In all, the risk of re-identification continues to rise and might as well be recognised, regulated, and used to serve public health interest.

**Data anonymisation and re-identification.** The protection and access to data should be reasonable to allow maximisation of the databank. As a consequence, there are limitations to anonymising data<sup>112,117</sup>. Anonymity will not allow linking datasets and growth of the database may depend on re-identify individuals if there is ethical reasonability and lawful approval to re-identify the participants<sup>113,119</sup>. Regardless, the principle of privacy protection must be always upheld, and such measures should be sufficiently described in the protocol for ethics approval. The data reuse options, and protective measures should also be detailed in the informed consent to involve participants in the decision regarding the reuse of their data by the researcher or a third party. These permutations make a fallacy of absolute anonymity. Hence, the growing call to inform participants that absolute anonymity is increasingly impossible to guarantee<sup>107,191,192</sup>. The difficulties of absolute anonymity are well described<sup>193</sup>. It has, for instance, been demonstrated that surnames can be re-identified using gene sequencing data<sup>194</sup>. Special training or augmentation of existing human research ethics curricula on the use of secondary data may be warranted, and certification mandatory in the event of inter-researcher data sharing.

Understanding the differences in maintaining anonymity is essential to guard against infringement of privacy. Thus, distinctions are made between anonymisation<sup>1</sup>, identifiability<sup>2</sup> and re-identifiability<sup>3</sup><sup>137,195</sup>. There is also the concept of pseudo-anonymisation; this involves removing identifiers and replacing them with single or double blinded codes to anonymise the data in a way that will allow authorised re-identification if or when there is ethical or legal imperative<sup>95,196</sup>.

---

<sup>1</sup> "In general, anonymisation refers to the process of removing identifying information such that the remaining data cannot be used to identify any particular individual...Data would not be considered anonymised if there is a serious possibility that an individual could be re-identified, taking into consideration both: (a) the data itself, or the data combined with other information to which the organisation has or is likely to have access and (b) the measures and safeguards (or lack thereof) implemented by the organisation to mitigate the risk of identification."<sup>195</sup>. Anonymisation is also used in refer to de-identified data that cannot be reversed<sup>89,197</sup>. HIPAA defines anonymised data as 'health information that does not identify an individual...there is no reasonable basis to believe that the information can be used to identify an individual...'<sup>189</sup>.

<sup>2</sup>"The degree to which an individual can be identified from one or more datasets containing direct and indirect identifiers"<sup>195</sup>

<sup>3</sup>"The degree to which an individual can be identified from anonymised dataset(s)"<sup>195</sup>

The reality is that patients' data are shared across departments for clinical care and for billing purposes. There is also an increase in clinical audit of patient records for quality improvement of practice and research without individual patient consent or promise of anonymity by researchers<sup>50,198–201</sup>. Similarly, social media is increasingly being used to mine vast biopsychosocial and other personal data, sometimes without authorization or consent of the individuals whose data is being used<sup>202–205</sup>.

Recognition of these realities, complemented by better regulation should mitigate unintended consequences such as stigmatisation of individuals or communities, genetic discrimination, racial stereotyping and discrimination, commercial exploitation of vulnerable groups, legal jeopardy and shaming<sup>120,206,207</sup>.

Various measures to ensure anonymisation of data have been proposed<sup>208</sup>. An essential step is to become aware of possible identifiers, which can be direct or indirect<sup>209</sup>. Malin *et al.* provide re-identification risks assessment and mitigation measures<sup>191</sup>.

Some ethical issues to note in relation to re-identification or computational phenotyping of data without participant consent is that it may constitute an infringement to the principles of autonomy and respect for person, beneficence and justice<sup>210</sup>. This makes re-identification a double-edged sword requiring due consideration. Re-identification without authorisation takes away a person's right to decide – this may extend to inferences or attributions being made about a dataset based on attributes from an unmasked data set. Equally significant is the re-identification and use of data of minors with consent and assent<sup>210,211</sup>. Re-identification or computational phenotyping may create an undue attention to a group or individual in a manner that may incite or perpetuate unfair treatment<sup>212–215</sup>. A lot of these challenges may be addressed by upholding the consent given by patients or study participants, use of appropriate technologies, mechanisms and permission to promote pragmatic dynamic consenting processes<sup>216</sup>. Over regulation of the data should also not become an impediment to robust scientific work<sup>217</sup>.

Some studies have recommended the sharing of random subsets of the database stripped of all possible individual unique identifiers<sup>153</sup> or to use aggregate datasets<sup>218</sup>. Other authors have suggested the inclusion of noise elements in aggregate data to further mask the dataset<sup>191</sup>. The noise elements may be in the form of random value changes, data swapping (switching values in the record), and synthetic data generation (creation of data from attributes of real records without corresponding to any real individual).

**Data access control.** Access to collected data may be open, controlled or hybrid depending on the level of sensitivity of the data and privacy concerns<sup>166,193</sup>. Open data is available for anyone to use without permission. However, controlled access data requires special permission. Controlled data have

higher risk of individual data re-identification and access to it may be made by the data access committee once all safety measures are met. The hybrid model combines both methods with restricted and open access to some data, thus, it carries a lower risk of re-identification of individual participant data. Similarly access control may be centralised in a pooled data system while access may be localised to the custodian in the federated system<sup>166,193</sup>. The different approaches should not negate the principles of autonomy, privacy, public interest and benefit, acknowledgment of data contributors, transparency, accountability and trustworthiness<sup>193</sup>.

Limited awareness and access to databanks available for secondary users may decrease the return on research investment in Africa. Timely access to data is an essential requirement of data sharing governance<sup>219</sup>. Access to and uptake of data should be promoted during stakeholder engagements and collaborative partnerships. This extends to devoting resources to addressing the impediments to data sharing<sup>220</sup>. A review of global recommendations<sup>219</sup> indicates that access to secondary data should be determined by the nature of the material available; the purpose of the request; the need for additional ethics clearance; intellectual property agreements; user fees; ownership of material; conditions of informed consent; assurance of confidentiality; and, material or user restrictions.

As a guide to data access, Desai *et al.*<sup>221</sup> propose the following five 'safes': "safe project (is the use of the data appropriate?); safe people (can researchers be trusted to use it in an appropriate manner?); safe data (is there a disclosure risk in the data itself?); safe setting (does the access facility limit authorised use?); safe output (are the statistical results re-identifiable?)". While the 'safes' provide a quick frame of reference for review, they should of course be used on the backdrop of local regulations, definitions and contexts. Other guides include "10 rules for responsible big data use"<sup>222</sup>, and the seven recommendations of the Caldicott Commission<sup>188,223,224</sup>.

The decision on access to data is also based on its ethical merit, public good, level of risk and mitigation measures proposed<sup>153</sup>. Other elements of the data access agreement may include "specific research objectives; plans for publication; permissions for and monitoring of access to the data; data storage, security, and confidentiality; allowances for copying or remote use, if any; de-identification plans; data destruction protocols; and, identification of parties responsible for data analysis and data security"<sup>153</sup>. Others have included up to 12 months after data release to publish findings of the research<sup>43</sup>.

The agreement should also prohibit users from re-identifying de-identified data without appropriate approval by an ethics committee<sup>43</sup>. Intention to obtain data from other sources that may result in wilful or accidental re-identification should be carefully considered and declared. This act is described

as data linkage and has been described in terms of its process, risks and benefits<sup>225</sup>. There is a growing list of studies that applied various data linkage methodologies to address complex issues<sup>226–230</sup>. There are proposals on how to use anonymised linkage technologies or split file methodologies to protect sensitive information or to de-identify multiple datasets after linkage by a bona fide third party with no conflict of interest<sup>231–233</sup>.

Most data sharing agreements are silent on the consequences of violating data access agreement<sup>234</sup> and rely on national regulations. This too must be explicitly stated in the agreement. Authors suggested that non-compliant users of the databank resources (principal investigators [PIs] and their Co-PIs) should be prohibited from using the databank and reported to authorities in their institutions, funders and other regulatory authorities and databanks<sup>98,235</sup>.

### Data access committees

Access to databanks is controlled by data access committees (DAC). DACs are tasked with the responsibility of reviewing data access requests and serve as oversight committees to approve or disapprove data access applications. The committee may be made up of civic organisation representatives, PIs, funders, other researchers, representatives of the group from whom the data was obtained, journal editors, and ethicists. Their specific roles include acquiring and storing data, ensuring data protection and information privacy, ensuring compliance to research consent agreements, protecting data quality and data donors, and balancing of timely publication with open access to data<sup>134,236–238</sup>. They equally have a fiduciary role to develop inclusive and unambiguous policies needed to execute these responsibilities.

There are two levels of governance of databanks – internal daily operations and external policy administration and stakeholder relations<sup>70</sup>. Governance provides a set of standard operating procedures, and ethical and legal consideration to inform the strategic and operation management of biobanks<sup>239</sup>. These principles also cover issues of funding, internal and external auditing and quality control, standard operation procedures for managing samples or data and ethical and legal consensus on management of samples and data. It is also part of the governance functions to have clear presentation processes of data collation, storage, use, and disclosure including policies and processes of data protection and risks assessment that may need to be updated regularly<sup>83</sup>. Specifically, the governance function of ensuring data protection entails measures to guard against privacy breaches such as unauthorised access to data or security breaches resulting from a deliberate attack on the system leading to loss of control of the dataset in their custody. In addition, governance entails providing a guideline on who, how, when and under what authority datasets can be linked or merged<sup>83</sup>.

Despite the important mandate that DACs play, they are confronted with various challenges, chief among them financial constraints and lack of sufficient oversight mechanisms<sup>240</sup>.

In addition, there is lack of clear definition of the relationship between DACs and biomedical RECs. In response, data custodians have pooled resources to develop a single better resourced DAC. The GA4GH provides a good framework to model from or adapt as necessary<sup>241</sup>.

Moreover, to address inequalities and curtail vested interests, authors have recommended that DACs should be inclusive, global and transparent<sup>242</sup>. This approach may address the issues of trust, transparency, equity, legitimacy, integrity and accountability<sup>173</sup>. In other words, DACs should be constituted to have a full spectrum of its stakeholders. To ensure fairness and effective executions of other fiduciary responsibilities, data access committee should be an independent committee without conflicts of interest and should have mechanisms to evaluate and mitigate its internal risks<sup>240</sup>.

### Data infrastructure, quality, storage and security

**Data quality.** The quality of shared data is important to ensure reproducibility<sup>241,243–247</sup>. Scepticism and self-doubt of quality of research may inhibit some researchers from sharing their data<sup>178</sup>. Data quality is a challenge in Africa due to lack of infrastructure, inadequate skills, and capacity amongst researchers as well as lack of guidelines on how data must be prepared or processed as discussed above. These concerns parallel what we found during our key informant interview with African research stakeholders.

Databanks are required to work with data contributors to establish and continuously implement data quality assurance measures including developing quality threshold indicators for routine review and updating<sup>104,112,117,248</sup>. Studies have reported that data quality assurance should be documented, unbiased, open to review, factual and proportionate<sup>10,104,117,119</sup>. African research may need to focus on generating more high-quality data. The H3Africa routine participatory process<sup>42</sup> may be a model to emulate as it assures control, compliance, and accountability along its data management value chain. While enforcement of data quality may not be enough to facilitate reuse<sup>249</sup>, data seal of approval is additionally offered by repositories guaranteeing researchers that data will be stored in a measure that assures their quality and consistent reuse while ensuring the trustworthiness of digital archives<sup>250,251</sup>.

Regulatory licencing and oversight of databanks could also help ensure quality<sup>252</sup>.

**Data storage and retrieval.** Integration of different datasets during storage may have risks, including re-identification of anonymised data, risk of disclosing other data, misinterpretation of data for various reasons, malicious use of data, harm to the public posed by illegal disclosure and commercialisation<sup>128,253</sup>. Cataloguing data in a consistent manner will promote harmonisation and interoperability<sup>254</sup>. This is further enhanced by using internationally accepted norms and standards to ensure compatibility<sup>104</sup>. Castillion *et al.*<sup>255</sup>

provide a comprehensive list of the requirements for online repository to address some of the common issues on security and utility. The sub items include metadata availability, discoverability, data standardisation, quality assurance, storage, backup, migration, succession plan, legal status, access and terms of use<sup>161,255</sup>.

Most consortia have relied on data integration systems such as the Open Archival Information System (OAIS)<sup>256,257</sup>, which enables the management of organisations and individuals intending to share data. The system offers a guide for developing common terminologies and concepts, architectures and operations of databanks to facilitate uniform and valid content sharing<sup>258</sup>. Detailed description of the complete enterprise system with data security features are described by Winter *et al.*<sup>258</sup>.

To ensure privacy protection, most databanks store anonymised or de-identified data with additional safety and access control measures to secure the data in their custody<sup>24,113,118,259,260</sup>. Strategies on maintaining anonymity have been developed above. To maintain anonymity, some studies have recommended the sharing of random subsets of the database stripped of all possible individual unique identifiers<sup>153</sup> or to use aggregate datasets<sup>218</sup>. Other authors have suggested the inclusion of noise elements in aggregate data to further mask the dataset<sup>191</sup>. The noise elements may be in the form of random value changes, data swapping (switching values in the record), and synthetic data generation (creation of data from attributes of real records without corresponding to any real individual)<sup>191</sup>. To ensure data truthfulness in public health, two general methods of re-identification prevention are used. These are data generalisation and suppression<sup>191</sup>. Under generalisation methods, data is replaced with general values and under the suppression method, unique identifiers are excluded from the data release<sup>261–264</sup>. Details for data de-identification and anonymisation measures for different data and sample types are described in a literature<sup>189,194,265,266</sup>. Other authors have recommended limiting time of access to datasets as well as the data they can access for a clearly defined project<sup>128</sup>. In addition to the mitigation measures, some countries prohibit unauthorised re-identification of shared data<sup>267</sup>.

The diverse datasets and data sources, and the technological advances in data management increase the risk of re-identification. Therefore, case-by-case consideration should be given to different requests by the data access committee and research ethics committee. Pharmaceutical industries for instance, have professional bodies and working groups (such as TransCelebrate<sup>268</sup> and Pharmaceutical Software Users Exchange<sup>269</sup>) that develop and regulate policies and procedures for data de-identification. Tucker *et al.*<sup>260</sup> have summarised best practice approaches to ensure data protection recommended by relevant institutions. In addition, Jones and Ford<sup>253</sup> have proposed models of integrating administrative data with other clinical data and reported

practical applications of the different models together with ethical, legal and social requirements for each model. They distinguish between two models — pooled data and federated data — by where the data is hosted and accessed. With a pooled system, data is accessed through a hosting entity whereas in a federated data model, data may be accessed through the source organisations.

The need for standardisation of data management frameworks that clarify data storage and sharing methodologies is central to both pooled and federated data sharing models. The framework may include standardisation of variable names, codes and storage format<sup>270</sup>. An alternative will be to adopt a standard metadata structure to allow transformation and integration as required by a central data management team constituted by a core team and representative data managers from across the consortia<sup>238</sup>. The core team may be made up of a neutral convening organisation with a governance function including convening stakeholders, quality assurance and oversight, financial management, communication, policy development and execution<sup>238,270,271</sup>.

**Security.** The safety of the data in most countries is protected by national privacy protection regulations, such as those mentioned above, and must meet human research ethical committee standards and approval<sup>272</sup>. These laws mandate the custodians of data to protect it from abuse, unauthorised access and tampering, loss or unlawful disclosure<sup>272</sup>. Privacy protection stipulates a notification obligation in the event of breach of privacy due to unauthorised access, loss or disclosure of information in the care of a legal data custodian<sup>273</sup>.

The three biggest cloud data storage service providers include Amazon, Google and Microsoft<sup>274</sup>. This cloud computing and few service providers come with significant risks ranging from integrity and exploitation of data by the service provider and its employees<sup>222,274–276</sup>, cloud attacks<sup>277</sup>, user identity spoofing<sup>278</sup>, data tampering<sup>279</sup>, denial of service<sup>280</sup>, unlawful access to database and infiltration of the system<sup>278</sup>, as well as re-identification of de-identified data<sup>281</sup>. Lessons from adverse experiences may offer hope to mitigate some of the risks in future<sup>274,282</sup>.

Some proponents of data security favour the establishment of remote access controlled data centres with state of the art monitoring systems to avoid physical transfer of data or unauthorised access or utilisation of datasets with capabilities to provide feedback or alerts on infringements<sup>107,283</sup>. Others have recommended the use of secure encrypted servers for data transfer<sup>153</sup>. They added that such electronic data transfer options should have multifactor authentication steps to access the databank with restriction to downloading or copying the dataset. Methodologies to ascertain the likelihood of re-identification are also evolving with their strengths and limitations<sup>234</sup>. Examples of the methodologies include K-anonymity<sup>261</sup> and unicity<sup>284</sup>.

There are various techniques for ensuring secure sharing of electronic information<sup>285</sup>. These techniques are grouped into two broad categories including the cryptographic and non-cryptographic techniques<sup>286–288</sup>. Cryptographic techniques encrypt stored data over the network and uses authentication techniques requiring decryption keys and verification using digital signatures<sup>285</sup>. These systems are also capable of providing patient control over their data by granting patient encryption and decryption control to allow access users of their choice.

Protection of electronic data is an ongoing process and various mechanisms have been adopted. These include the use of patient encryption<sup>289</sup>, employment of a third party to protect data integrity through layered encryption<sup>290</sup>, data partitioning techniques<sup>291</sup>, digital signatures<sup>292</sup>, hierarchical encryption<sup>293</sup>, the Elliptic Curve Digital Signature Algorithm (ECDSA), a cryptographic algorithm (used by Bitcoin), and many other techniques with their own strengths and limitations<sup>285</sup>. Variant three of the ECDSA is acclaimed to withstand many of the risks already described. The choice of privacy protection techniques adopted should also be made based on its functionality and implication for data accuracy using a bottom-up development approach<sup>294</sup>.

The success of cybersecurity will equally depend on good governance that ensures compliance with safety regulation by all parties.

**Sustainability.** The need for financial sustainability to support capacity and infrastructure for data sharing is underscored<sup>167,169</sup>. Efficient pooling of resources for integrated data sharing platforms and joint funding application for data sharing initiatives by research partnerships have also been recommended<sup>295–298</sup>. Other proposed funding mechanisms include the establishment of foundations or charitable trusts to stimulate donor support towards public benefit, and a model involving a shared cost approach by partnering with governments, non-profit organisations and commercial entities<sup>299</sup>.

Researchers have recommended that the sustainability of the databank must be determined from inception<sup>104,117</sup>. Ensuring sustainability will include consistent application of the policies throughout its lifespan including promoting scientific and ethical integrity<sup>47</sup>. Discontinuation or change of ownership or eventual disposal of data should form part of the sustainability plan<sup>112,117</sup>. Obtaining appropriate liability insurance for a databank may be a way of ensuring its sustainability<sup>252</sup>. There are potential opportunities for public-private-partnerships for public good, which may involve private sector use of public data for research or the integration of private sector data in public data, or public-private partnership for innovation and development<sup>300</sup>. On the other hand the challenges to data sharing for commercial use mostly pertain to issues of social licence and public distrust and limited oversight of

commercial data, data ownership, intellectual property, commercial secrecy, insufficient transparency, and profiteering<sup>300</sup>.

Importantly, ensuring the sustainability of the databank must assume the qualities of a resilient system. Such a system is defined by its capacity to proactively adapt to changes and challenges to its daily operation and sustenance<sup>301</sup>. This may also involve collaborative learning and stakeholder involvement as vital prerequisite pillars<sup>302</sup>. Human capital and its adaptive capacity to such innovation will require digital literacy of platform users as well access to technology<sup>303</sup>. These attributes help to create a system that is flexible, and adaptable to variabilities and improvisations<sup>304</sup>. Moreover, a protocol to develop a resilient system that responds to cross country population health needs are described<sup>301</sup>. Role clarification of the different stakeholder groups specified<sup>121</sup> is equally essential to the sustainability of databanks. Further requirement for system's sustainability and adaptive capacity have been richly described and graded in terms of human capital and financing raking<sup>305–308</sup>.

**Data harmonisation.** There are exemplary data sharing repositories in Africa, but these platforms have different levels of information technology, different data structures and largely operate parallel to each other. Integrating such databases may require a harmonised data sharing platform.

Harmonisation is complex. Townsend<sup>309</sup> argues that it can be achieved through a bottom-up approach. This proposition is premised on consortia and stakeholders' capacity to work together to find common grounds, policies, and solutions. An example is made about the success of GA4GH and P3G consortium, and the same can be said about H3Africa deliberative and accountability mechanisms<sup>42,310,311</sup>.

Other than government agencies, public and population health data in Africa predominantly sits with non-governmental organisations, charities, and research and academic institutions. Furthermore, the repositories may be institutional such as a university; governmental holding of administrative, service delivery or surveillance data; discipline specific repository<sup>193</sup>. These institutions are predominantly donor funded and thus, expected to make data available to initiatives that serve public interest.

There are technical challenges to integrating and managing multi-disciplinary data from diverse jurisdictions. These include data dispersion, provenance and heterogeneity<sup>46</sup>. This triple challenge arises from the thousands of possible data sources across the continent on different public and population health topics varying in scope and scale. These data are also collected using different methodologies, formats and data management protocols<sup>46</sup>. The issue of dispersion may be addressed by harmonising and augmenting routine national survey and encouraging in-country groups and independent researchers to adopt existing tools where necessary

and store data in a secured and legal repository. To reduce heterogeneity, similar methodologies may be promoted among contributors to repository with incentives to promote contribution. The submission of metadata describing data elements used for each project will promote accurate utility and integration. Dealing with these challenges can be done in a manner that does not create unintended ethical breaches such as uncontrolled or unauthorised re-identification or disclosure of participant information. Other challenges and opportunities of an integrated system are presented by Shah and Khan<sup>312</sup> and Jones *et al.*<sup>71</sup>.

## Discussion and Conclusion

This article focused on global data sharing practices, and the development of databanks in Africa. The various documents reviewed, and interviews conducted with African stakeholders, offer insights on key challenges to data sharing and databanks. In addition, this research showcases existing opportunities that may be leveraged to develop a multi-consortia public and population health data sharing platforms in Africa, and similar contexts in LMICs. Specifically, African governments can learn from the mistakes of high-income countries on data sharing practices and tap into their positive and practical strategies that may enhance efficient development of integrated databanks in the region.

There are already, best practice platforms in Africa. Initiatives such as the INDEPTH, H3Africa Consortium and the African Academy of Science's DELTAS programme are developing capacity in several research institutions across the continent. Some of these initiatives not only provide exemplary data sharing guidelines in Africa, but also aim to shift the role of African researchers from being mere data collectors or community brokers to becoming active leaders capable of enhancing scientific growth in Africa<sup>2,5</sup>. Yet, we noted various structural, individual, and contextual challenges that may hinder data sharing in Africa. In addition, it is evident that genomic data sharing dominates the scientific world globally and Africa in particular. There is need to address existing factors that hinder data sharing as discussed above and incorporate genomic data with other public health data to enhance scientific benefits in public and population health.

Establishing an integrated databank in the African region is increasingly becoming a matter of when and not if. Bold regional and global treaties may be needed to ensure safe and secure uptake of digitally available data. This includes the continuous development, monitoring and governance of ethical and operational standards in response to data access and proliferation requirements to protect the privacy, security, safety, and anonymity of data contributors.

The rapid growth in human subject or tissue databanks and sharing facilities gives urgency for national regulatory bodies to create guidelines and policies on data management and sharing<sup>110</sup>. Inadequate, or the absence of, such policy

guidelines is a major setback in most LMICs, and Africa. Development of databanks is also an evolving area with the rising scope, scale and complexity of emerging data and data sources ushering novel questions around ethical principles<sup>10,155,242,313,314</sup>. Additionally, incoherence of national laws and regulations coupled with varying levels of adherence to laws does not always translate to moral use of data nor offer a guarantee for public trust<sup>315</sup>, hence the need for continuous development and oversight.

The implementation of dynamic consent and opt-out options for routine health service users at the point-of-care may be a solution to accessing public data in a manner that respects the autonomy of the patients or research participants. In the absence of an integrated databank, opt-out option remains an important ethical consideration with the rise in clinical audit research studies to measure quality of care<sup>26,316–319</sup>.

Our research's heavy reliance on experience from sharing of genomic data and lack of sufficient African studies in the literature is notable. This was due to the availability of publications on genomic data sharing and limited studies focusing on data sharing experience in Africa. The study does not cover the use of data integration for precision medicine from the Global North, which has its own specific ethical complexities already presented by Browman *et al.*<sup>235</sup>. Furthermore, the findings and recommendations reported in this article, however, do not create a one-size-fit-all solution for Africa. Instead, they provide considerations on how to harness Africa's opportunities for safe and secure optimisation of its available data. Africa lags behind in all essential public engagements required to build integrated databanks, as we found no study exploring the view of African populations on data sharing and databank governance. We suggest the use of various targeted surveys on various groups or researchers working on specific health research such as malaria, HIV, or genomic studies as consultative tool to establish public opinion on data sharing.

There is also a need to reconsider consenting tools and processes to include follow-up clauses and mechanisms including the use of appropriate technologies. To this end, others have suggested the addition of an exclusion clause in the information sheet and consent form<sup>29</sup>. This proposition resonates with recommendations that privacy protection policies should serve all dynamic interests of its stakeholders<sup>53</sup>. This article also recognises the multitude of concurrent policies and regulations governing issues of consent, intellectual property, and confidentiality.

The African Union should consider developing multilateral privacy and data governance policies and framework like existing European Union and OECD treaties on data sharing or other Safe Harbour arrangements described by Dove *et al.*<sup>245</sup>. This may be useful to address jurisdictional barriers and efficient resolution and monitoring of matters of registration,

compliance review, recognition, monitoring and enforcement, public participation, and general operations and guiding principles.

The growth in data science technical expertise on the continent<sup>320</sup>, efficient infrastructure management<sup>321</sup> and proficiency in scaling-up innovations could be harnessed to develop integrated databanks<sup>320</sup>. Policies for data sharing will not be realised without dedicated funding and monitoring mechanisms. Funder requirements for the sharing of data are unethical if this cannot be done safely and meaningfully if the infrastructure and skills to manage shared platforms is not developed. At the research project level, funding to ensure good meta-data is provided to enable meaningful sharing is needed. Investment in the sharing super structure, both technical and human, is required. The opportunity of developing an integrated databank may be best managed through benefit from big ethics structure of safe harbours. We also recommend a hybrid harmonisation approach<sup>322</sup>. Blockchain technologies can be used to control access to data. Key informant interviews with African scientist suggests that most would like to participate in future use of their data if given the opportunity.

Public concerns about data sharing are viewed as conditions for sharing. Fortunately, there is a growing array of mitigation measures to address these concerns in partnership with the community. This takes cognisance of differences in the level of these concerns by socio-demographic characteristics. Fortunately, a lot of the concerns are mutable with greater transparency and communication. Others have noted that healthcare providers are more likely to help individuals appreciate and participate in data sharing initiatives<sup>323</sup>. Further classification into broad groups is made based on their concern about data sharing and who to trust with shared data<sup>323</sup>.

Exploring facilitators and barriers in African populations is paramount to future success particularly in the context of who holds the data, and role of socio-economic, cultural, and religious values in data sharing participation. The information will help establish public communication and in developing a platform that is responsive to the will, aspirations, and concerns of African populations platform. Risks posed by data sharing to different groups need to be explored and measures to increase protection require more investigation<sup>234</sup>.

Other general recommendations are listed below, while specific recommendations to specific challenges and risks are presented in [Table 1](#).

1. Developing a utilitarian integrated multidisciplinary databank for African may be feasible by harnessing the increasing data science technical expertise and strategic collaborations in the continent, together with the proliferation of cloud technology and concomitant reduction in cloud computing infrastructural costs and maintenance burden<sup>320,321</sup>.
2. Overall, Africa is well placed to advance in data integration given the wealth of global lessons to leverage. While there is opportunity to build the databank through integration and harmonisation of existing national surveys, HDSS datasets, biobanks, routine health service and administrative data, disease specific registries and notification systems, there are also lessons from prospective digitally enabled African multi-country surveys to build on<sup>324</sup>.
3. An integrated African public and population health databank may be built on familiar and aptly described health system governance principles<sup>325</sup>. The principles include strategic vision, rule of law, transparency, participation and consensus orientation, ethics, accountability amongst others. These principles are in line with the values for data sharing classified into two groups: substantive (e.g. harm minimizations, social justice and public benefit), and procedural (e.g. transparency engagement and reflexivity)<sup>326</sup>.
4. A hybrid developmental approach that combines the benefits of bottom-up and top-down approaches should be explored.
5. African multi-consortia engagements initiatives may be a starting point to harness big datasets, technical capacities, institutional knowledge, policies, operational guidelines, governance mechanisms, strategic partnerships, and social licences and capital.
6. Our findings support the growing call to rethink the process and requirements for informed consent<sup>26,316-319</sup>. Such efforts should seek to develop mechanisms that may allow a gradual build-up of data with appropriate permission for an integrated database.
7. Considering the wealth of data that already exist and their potential to be integrated to address regional public health challenges, extensive stakeholder engagement may be needed to decide how to manage the consent to use legacy data for future research as well as new approaches to future data collection. Such engagement may include the establishment of an inclusive stakeholder committee to generate recommendations for open dialogues and refinement. Other approaches have been used<sup>49,53</sup>.
8. Interventions should be developed to address known concerns about data sharing especially among underrepresented populations.
9. Attention should be paid to the issue of data quality in Africa through capacity building initiatives. This calls for both encouragement and making the provision of quality data an obligatory requirement<sup>80</sup> with support mechanisms. Additional bioinformatics training or incorporation of relevant skills development into training curriculum is also recommended<sup>327</sup>.

**Table 1.** Specific considerations and recommendations.

Themes and sub-themes	Considerations (challenges and risks)	Recommendations
<b>A) DATABANKS</b>	<ul style="list-style-type: none"> <li>✓ Africa lacks integrated data banks the various data repositories e.g., HDSS sites, H3Africa and DHS are not harmonised<sup>18,42,51</sup></li> <li>✓ There is limited oversight and unclear policies by government institutions and research ethics committees on data sharing and governance of databanks<sup>219</sup>.</li> <li>✓ Ethical, legal and social implications of secondary data sharing are mostly unresolved<sup>11,103,28,329</sup></li> <li>✓ Public fear of loss of privacy or confidentiality breach, data misuse and abuse<sup>56,58,59</sup></li> <li>✓ Poor communication on data use leads to distrust from participants<sup>60,61</sup></li> <li>✓ Insecurity, growing cyber-attacks, fear of using the internet<sup>87-89,149</sup>, and dishonesty due to fear of stigmatisation<sup>85</sup>.</li> <li>✓ Researchers fear of possible loss of academic advantage and independence; loss of intellectual property<sup>70</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ A need to develop integrated and harmonised databanks and frameworks for data sharing in Africa<sup>133,140,330</sup>. Examples could be drawn from the Australian Population Health Network<sup>65</sup>; the Canadian National Data Platform<sup>66</sup>; and the UK's Health Data UK<sup>67</sup>.</li> <li>➤ Develop policies on regulatory oversight and that enables collaborations<sup>47,48</sup>.</li> <li>➤ A need to develop a harmonised agreement that respects the independence of separate entities while promoting robust and efficient cross-disciplinary research within the confines of national and international ethical and legal frameworks<sup>68</sup>.</li> <li>➤ Develop proper governance of databanks, quality management and sustainability<sup>47,48</sup>.</li> <li>➤ Data custodians must adhere to ethical guidelines (e.g., privacy, trustworthiness) in data sharing<sup>83-85</sup>, and use or share the data for public good and social justice<sup>93,149,152,331</sup>.</li> <li>➤ A need to improve on communication to research subjects regarding data sharing using strategies such as modular education approach<sup>96</sup>; use of video<sup>91</sup>, pictures and vignettes<sup>92,93</sup>.</li> <li>➤ Need to conduct a public education on data reuse to promote trust and public participation<sup>78</sup>.</li> <li>➤ A need to collect rich metadata of each data set<sup>80,81</sup>.</li> <li>➤ Other considerations are detailed in Wiehe <i>et al.</i>, including identifying data sources/patterns, engagement with leaderships, ethical and regulatory compliance, etc.<sup>121</sup>.</li> </ul>
<b>B) DATA PROTECTION LAWS AND GUIDELINES</b>	<ul style="list-style-type: none"> <li>✓ Limited to moderate data regulation and enforcement particularly in Africa<sup>92</sup>.</li> <li>✓ Other unaddressed issues include public view or perceptions of cross border data transfer<sup>120</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ African countries without data protection policies must develop data protection policies by learning or borrowing from global models such as the UK Data Protection Act of 2018<sup>94</sup>, and examples from the African continent<sup>92</sup></li> <li>➤ Develop safe harbour privacy protection principles to address cross border regulatory bottlenecks, increase data sharing efficiency, and promote data harmonisation<sup>245</sup>.</li> <li>➤ A need to build capacity of research EC to ensure consistent and efficient application of data sharing regulations<sup>333</sup>.</li> <li>➤ EC must be guided by global ethical guidelines including The Helsinki Declaration which provides guidance on data security, ethical principles and governance of data sharing<sup>102</sup>.</li> <li>➤ Other guidelines include: the Australian Guidelines on Human Biobanks and Genetic Research Databases<sup>103</sup>; The OECD Principles and Guidelines for Access to Research Data from Public Funding<sup>104</sup>; the Bermuda Principles<sup>105</sup>; Fort Lauderdale Agreement<sup>334</sup> among others.</li> <li>➤ The material transfer agreement (MTA) documents should include issues of data provenance, data quality assurance, meta-data and other requirements for accurate interpretation of data, intellectual property, informed consent, security and privacy terms etc.<sup>107,335</sup>.</li> <li>➤ Develop Ethics waiver policies including setting up a central adjudicator of request when re-identification is necessary, and consenting is impractical. Examples include the Confidentiality Advisory Group (CAG) and the Public Benefit and Privacy Panel (PBPP) in England and Scotland respectively<sup>94</sup>.</li> </ul>
<b>Ethics Committees (EC)</b>	<ul style="list-style-type: none"> <li>✓ lack of legal protection<sup>96</sup>.</li> <li>✓ Inability to reach quorum in decision making and inappropriate constitution of ethics committees (EC)<sup>152,153</sup></li> <li>✓ Inefficiency or bias amongst its members<sup>99</sup>.</li> <li>✓ Lack of financial and administrative support to enable it to function smoothly<sup>332</sup>.</li> <li>✓ Social implications of data sharing often falls outside the ECs mandate<sup>50,100,101</sup>.</li> <li>✓ EC members' poor familiarity with secondary data use, including laws governing cross-border transfers may be an impediment to safe data sharing.</li> </ul>	<ul style="list-style-type: none"> <li>➤ A need to build capacity of research EC to ensure consistent and efficient application of data sharing regulations<sup>333</sup>.</li> <li>➤ EC must be guided by global ethical guidelines including The Helsinki Declaration which provides guidance on data security, ethical principles and governance of data sharing<sup>102</sup>.</li> <li>➤ Other guidelines include: the Australian Guidelines on Human Biobanks and Genetic Research Databases<sup>103</sup>; The OECD Principles and Guidelines for Access to Research Data from Public Funding<sup>104</sup>; the Bermuda Principles<sup>105</sup>; Fort Lauderdale Agreement<sup>334</sup> among others.</li> <li>➤ The material transfer agreement (MTA) documents should include issues of data provenance, data quality assurance, meta-data and other requirements for accurate interpretation of data, intellectual property, informed consent, security and privacy terms etc.<sup>107,335</sup>.</li> <li>➤ Develop Ethics waiver policies including setting up a central adjudicator of request when re-identification is necessary, and consenting is impractical. Examples include the Confidentiality Advisory Group (CAG) and the Public Benefit and Privacy Panel (PBPP) in England and Scotland respectively<sup>94</sup>.</li> </ul>



Themes and sub-themes	Considerations (challenges and risks)	Recommendations
<b>B) DATA PROTECTION LAWS AND GUIDELINES</b>  <b>Consenting</b>	<ul style="list-style-type: none"> <li>✓ There are no clear guidelines for conducting informed consent<sup>336</sup>.</li> <li>✓ Complex use of data has made it difficult to differentiate between data collected for routine medical care and data collected for research<sup>49,198-201,237</sup>.</li> <li>✓ Possible risk of patients and research participants being relegated to data donors, and negating the principles of autonomy and self-determination<sup>109</sup>.</li> <li>✓ There are unresolved issues on future use of data, including when participants want to opt in/out of studies<sup>111</sup>.</li> <li>✓ The scope of consenting is also not so clear in longitudinal studies, especially those involving minor<sup>62</sup>, and parents may be reluctant to consent for minors.</li> </ul>	<ul style="list-style-type: none"> <li>➤ It is important to give people the opportunity to negotiate how others use their personal information<sup>338</sup>.</li> <li>➤ Researchers/Investigators must ensure that consenting process is a broad, continuous process and touches on data sharing clauses (data sharing now and, in the future), and ensure waivers permitting the use of de-identified data<sup>110</sup>.</li> <li>➤ Longitudinal studies should have follow-up mechanisms e.g. collecting additional identifiers for participants on the consent form to allow future re-contacting for further consenting<sup>156</sup>.</li> <li>➤ Research ethics committees should contend with emerging considerations of data stewardship such as the longer than usual data storage, sharing, re-identification and indeterminate future use of collected data<sup>28-30</sup>.</li> <li>➤ Researchers must ensure that participants have enough information about their studies and consent options including a consent waiver, dynamic consent to opt in and/or opt out etc.<sup>111</sup>.</li> <li>➤ There is need to adapt existing software to facilitate data governance and participants' control of their data<sup>52,339,340</sup>. Examples include Fast Healthcare Interoperability Resources (FHIR), Sync for Science, Private Access, Patient Health Records (PHR) and Blue Button<sup>33,239</sup>.</li> <li>➤ Data ownership should be governed by legal and moral obligations including trust and custodianship with variations in the right of access and utility by different stakeholders<sup>133-135</sup>.</li> <li>➤ There is a need to adopt a non-exclusive ownership of data – whereby data ownership should be governed by legal and moral obligations.</li> <li>➤ Data custodians must adhere to principles of respect for privacy and autonomy; reciprocity and feedback to stakeholders; acknowledgment and attribution to contributors; and, respect for intellectual property<sup>107,311</sup>.</li> <li>➤ There is need to develop a system and guidelines/templates for Intellectual Property that is guided by local intellectual property laws<sup>104,114,136</sup>.</li> <li>➤ Databank users are required to report back all publications and patents emanating from the data provided to them<sup>107,117,119</sup>.</li> <li>➤ Genomic databases are global public good and all humans should share in, and have access to, the benefits of databases<sup>342</sup>. Similar views are shared in UNESCO's International Declaration on Human Genetic Data<sup>343</sup>. Thus, provide access to databases to anyone who rightfully demonstrates a need to access the data.</li> </ul>
<b>C) ENABLERS OF DATA SHARING</b>  <b>Stakeholder and Community engagement</b>	<ul style="list-style-type: none"> <li>✓ Data sharing may raise several issues to researchers, employers, and funders on: What are the legal rights in data? Who has these rights? And how does one with these rights use them to share data in a way that permits or encourages productive downstream uses?</li> <li>✓ Some data repositories e.g., journals have strict measures that hinder access to data by those who cannot pay for it.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Stakeholder consultation is an important strategy to promote equity, trust, transparency, autonomy and participation in data storage/sharing<sup>10,109,153</sup>.</li> <li>➤ Communication should be done with the required sensitivity to avoid ambiguity and misinterpretation<sup>153</sup>.</li> <li>➤ Community engagement should commence at the beginning of the project, to ensure feasibility and timely risks mitigation with stakeholders' input<sup>109</sup>.</li> <li>➤ The consultation should clarify purpose of the data storage and sharing platform, roles and responsibilities, governance and accountability mechanisms, data protection, types of informed consent, benefit-sharing, intellectual property, and data ownership. Exemplary framework can be drawn from H3Africa<sup>154</sup>.</li> </ul>

Themes and sub-themes	Recommendations
<p><b>C) ENABLERS OF DATA SHARING</b></p>	
<p><b>Trust</b></p> <ul style="list-style-type: none"> <li>✓ Social licence may be misinterpreted as trust, which may be implied as informed consent to use information offered for research<sup>137</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ In the case of big databanks, maintaining trust should be on-going and not a onetime checkbox activity.</li> <li>➤ The engagements should also be cross cutting to involve other researchers, policy makers and funders, and not only research participants and communities<sup>512,153,140,141</sup>.</li> </ul>
<p><b>Respect for study participants/ groups</b></p> <ul style="list-style-type: none"> <li>✓ Issues may include where researchers do not disclose fully to participants on future use of data.</li> <li>✓ Another issue would be not being clear during consenting time whether participants will be recontacted or not.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Use of data should be in line with the scope of original informed consent provided by the research participants.</li> <li>➤ The intention of the research is clearly stated during consenting/at the time of data collection including likely future use of the data<sup>112,114</sup>.</li> <li>➤ In absence of specificities, broad consenting should be done to protect the research participants<sup>112-114</sup>.</li> <li>➤ Elements of respect may include privacy protection and confidentiality; autonomy; data security; respect for individuals and group rights; ensuring dignity of participants; and, protection of life, wellbeing and welfare<sup>10,102,112,119</sup>.</li> <li>➤ Re-contacting participants should of course, follow standard ethical principles including options on communication of findings or participant access to data<sup>117,118</sup>.</li> </ul>
<p><b>Transparency</b></p> <ul style="list-style-type: none"> <li>✓ Providing patients or study participants with insufficient information on how data will be managed or shared<sup>83,86,152,346</sup></li> <li>✓ Unspecified secondary use of data<sup>104</sup></li> <li>✓ Giving multiple users access to data<sup>93,104,105</sup></li> <li>✓ Data misuse, identity theft and sharing data on the internet<sup>99,101,103,119,121</sup>, and centralised database without sufficient safeguards<sup>99,119</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Researchers must ensure participants are informed about how data will be shared and with whom<sup>53,142</sup>.</li> <li>➤ Researchers must disclose to participants about monitoring policies and database governance, conditions framing access to data and data access agreements<sup>144-146</sup>.</li> <li>➤ Also, disclose the role of patients and human rights advocacy groups involvement in providing oversight and supervision of the platform to ensure unbiased access and utilization of the databank<sup>148</sup>.</li> <li>➤ Ensure proper keeping and communicating sufficient records of operational activities including audits logs and trails<sup>86,87,149</sup>.</li> </ul>
<p><b>Incentivization of data contributors and users</b></p> <ul style="list-style-type: none"> <li>✓ funding agreements, collaborative agreements, data sensitivity, privacy, giving up chance to publish, public critique, lack of data repositories and the absence of consent to share<sup>60,165</sup></li> <li>✓ Fear of exploitation especially amongst researchers in low resources countries<sup>161</sup>.</li> <li>✓ Threat to intellectual property, professional value and economic benefits<sup>166</sup>.</li> <li>✓ The greater value placed on publications by institutions may also be discouraging data sharing<sup>164</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ It is important for governments and funders to ensure capital and infrastructural development, and financing to promote research data sharing<sup>65,167-169</sup>.</li> <li>➤ Research institutions and researchers need to promote tangible reward in the form of reputational incentives and peer recognition including citation to enhance data sharing<sup>158,170</sup>.</li> <li>➤ Make data sharing a requirement for project funding, journal publications, university tenure or promotion<sup>212,270</sup>.</li> <li>➤ A need to develop clear data sharing policy that addresses the concerns of all stakeholders, including monitoring and reward mechanisms<sup>161,173</sup>.</li> <li>➤ A need to promote diversity and inclusion of minorities and vulnerable groups<sup>65</sup> and international partners in data sharing<sup>178</sup>.</li> <li>➤ Develop open data badges – which is a tested intervention to improve data sharing<sup>171,179</sup>.</li> </ul>

Themes and sub-themes	Considerations (challenges and risks)	Recommendations
<b>C) ENABLERS OF DATA SHARING</b>		
<b>Funders and researchers' position</b>	<ul style="list-style-type: none"> <li>✓ Most researchers or scientists in Africa are hesitant to share their data largely due to lack of awareness of the benefits of data sharing.</li> <li>✓ Lack of funding and limited provisions for data sharing.</li> <li>✓ Few members of African ethics review boards are familiar with the concept of data sharing amongst other ethical issues discussed such as broad consenting</li> </ul>	<ul style="list-style-type: none"> <li>➤ We recommend proactive advocacy to ensure that the concept of data sharing becomes a mainstream consideration in national discussions of research management and governance<sup>70</sup>.</li> <li>➤ There are policies that illustrate that all data is public good, and all funded research should be shared. This includes the Wellcome Trust<sup>180</sup> and the USAID's Policy on Development Data<sup>181</sup>.</li> <li>➤ A need to train researchers on data management, and the recruitment of dedicated support staff to document data and manage repositories<sup>155,221,253</sup>.</li> </ul>
<b>D) GOVERNANCE AND VALUE-BASED IMPLEMENTATION</b>		
<b>Policies and Values</b>	<ul style="list-style-type: none"> <li>✓ Most guidelines and regulations within Africa do not provide clear guidance on governance and how data and samples ought to be shared<sup>182,183</sup>.</li> <li>✓ Lack of clear policies on data sharing may both frustrate researchers who want to share data and provide loopholes for those who are unwilling to share.</li> <li>✓ Diminished confidence on government custodial of the data<sup>142</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Governments and research institutions in Africa must develop clear guidelines on data sharing and repositories.</li> <li>➤ Create clear laws to govern re-identification and stronger sanctions and corresponding enforcement protocol for misuse of data<sup>133,189,190</sup>.</li> <li>➤ Establish proper governance by providing a guideline on who, how, when and under what authority datasets can be linked or merged<sup>49</sup>.</li> <li>➤ Develop a central policy and inclusive governance structure that promotes collaboration and participants<sup>193,146</sup>.</li> </ul>
<b>Data anonymization and re-identification</b>	<ul style="list-style-type: none"> <li>✓ There is also an increase in clinical audit of patient records for quality improvement practice and research without individual patient consent<sup>50,198-201</sup>.</li> <li>✓ Yet, data anonymization may be challenging when researchers or clinicians want to link medical data to make clinical decisions in future, or recontacting patients to obtain additional information.</li> <li>✓ Growth of the database means anonymity will not allow linking datasets or to re-identify individuals in the database if there is ethical reasonability and lawful approval to re-identify the participants<sup>113,119</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Researchers and data custodians must be aware of possible identifiers, which can be direct or indirect<sup>191,209</sup>.</li> <li>➤ Data controllers must uphold to the consent given by patients or study participants, use of appropriate technologies, mechanisms and permission to promote pragmatic dynamic consenting processes properly described by Kaye <i>et al.</i><sup>216</sup>.</li> <li>➤ Researchers must ensure that details on data reuse and protective measures are clearly stated in the informed consent, and inform participants when absolute anonymity is increasingly impossible to guarantee albeit highly preventable<sup>107,191,192</sup>.</li> <li>➤ It is important to adequately educate researchers and data custodians to ensure data privacy protection compliance as well as signing renewable confidentiality pledges<sup>193</sup>.</li> <li>➤ Data should be de-identified before it is shared<sup>810</sup>.</li> </ul>
<b>Data Access</b>	<ul style="list-style-type: none"> <li>✓ Most data sharing agreements are silent on the consequences of violating data access agreement<sup>234</sup> and rely on national regulations.</li> <li>✓ Limited awareness and access to databanks available for secondary users<sup>219</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Develop clear data access agreements or guidelines on what the application can and cannot do with the data provided<sup>260</sup> as well as consequences of nonadherence to data access agreement<sup>234</sup>.</li> <li>➤ Data access should not negate the principles of autonomy, privacy, public interest and benefit, acknowledgment of data contributors, transparency, accountability and trustworthiness<sup>193</sup>.</li> <li>➤ Promote data access discussions during stakeholder and collaborative partnerships, including resource provisions to addressing the impediments to data sharing<sup>220</sup>.</li> </ul>

Themes and sub-themes	Considerations (challenges and risks)	Recommendations
<b>D) GOVERNANCE AND VALUE-BASED IMPLEMENTATION</b>		
<b>Data access committees (DACs)</b>	<ul style="list-style-type: none"> <li>✓ Financial constraints and lack of sufficient oversight mechanisms<sup>240</sup>.</li> <li>✓ There is lack of clear definition of the relationship between DACs and biomedical research ethics committees (ECs) when conducting evaluations.</li> <li>✓ Insufficient oversight mechanisms<sup>59</sup></li> <li>✓ Inequalities in terms of the composition of DACs- which may exclude important stakeholders<sup>242</sup>.</li> <li>✓ Conflict of interests between DAC members and other stakeholders<sup>42</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ DACs must be provided with adequate funding to perform their roles<sup>240</sup></li> <li>➤ Develop clear guidelines and framework to guide functioning of DACs.</li> <li>➤ A need to adapt to technological, scientific, data security, new data sources and research methodological advances and changes in public sentiments<sup>347,348</sup>.</li> <li>➤ The need to have an oversight over DAC is recommended<sup>95,240</sup>.</li> <li>➤ To address inequalities and curtail vested interests, DACs should be inclusive, global and transparent<sup>42</sup>.</li> <li>➤ DACs should be an independent committee without conflicts of interest or measure to evaluate and mitigate its internal risks<sup>240</sup>.</li> </ul>
<b>E) DATA INFRASTRUCTURE, QUALITY, STORAGE AND SECURITY</b>		
<b>Infrastructure</b>	<ul style="list-style-type: none"> <li>✓ Many African institutions have limited infrastructure (spaces, inadequate equipment/tools, power supply shortages, poor information technology) for data repositories and data sharing<sup>1-3</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ There is a need to develop ICT infrastructure and efficient workflow; harmonised policies, guideline and operating procedure; data access policies and mechanism; and, government regulation and oversight<sup>349</sup>.</li> <li>➤ Other considerations include human and social capital, financial resources and governance<sup>350</sup>.</li> <li>➤ Developing an adaptive information technology enabled system.</li> <li>➤ Ensure adequate financial resources to address the mentioned challenges.</li> </ul>
<b>Data Quality</b>	<ul style="list-style-type: none"> <li>✓ Some of the reasons why scientist do not reuse data include concerns about data quality; lack of awareness of benefits of big data; and, lack of technical capacity to use big data<sup>351</sup>.</li> <li>✓ Scepticism and self-doubt of quality of research may inhibit some researchers from sharing their data<sup>178</sup>.</li> <li>✓ Poor data quality in Africa is due to lack of infrastructure, inadequate skills and capacity amongst researchers as well as lack of guidelines on how data must be prepared or processed.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Data custodians and Databanks must establish high quality threshold indicators for routine review and updating<sup>104,112,117,248</sup>.</li> <li>➤ Data quality assurance should be documented, unbiased, open to review, factual and proportionate<sup>10,104,117,119</sup>.</li> <li>➤ Researchers and data custodians must establish the contextual meaning of data to minimise misinterpretations. Example can be drawn from the H3Africa model<sup>42</sup>.</li> <li>➤ It is important to also offer data seal of approval to guarantee researchers that data will be stored in good quality, and consistent reuse while ensuring the trustworthiness of digital archives<sup>430,251</sup>.</li> <li>➤ Regulatory licencing and oversight of databanks could also help ensure quality and accountability<sup>252</sup>.</li> </ul>
<b>Data storage &amp; Retrieval</b>	<ul style="list-style-type: none"> <li>✓ Identification of anonymised data, increased risk of disclosing other data, misinterpretation of data for various reasons, malicious use of data, harm to the public posed by illegal disclosure and commercialization<sup>128,253</sup>.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Cataloguing data in a consistent manner will promote harmonization and interoperability<sup>254</sup>.</li> <li>➤ African data scientists or custodians must draw from internationally accepted norms and standards to ensure compatibility<sup>104</sup>.</li> <li>➤ Data custodians (e.g. on online platforms) must ensure: metadata availability, discoverability, data standardization, quality assurance, storage, backup, migration, succession plan, legal status, access and terms of use and more shown in the table<sup>161,255</sup>.</li> <li>➤ Develop an integrated system such as the Open Archival Information System (OAIS) for data management and sharing<sup>256,257</sup>.</li> <li>➤ Databanks must store anonymised or de-identified data with additional safety and access control measures<sup>24,113,259,260</sup>, use individual unique identifiers<sup>153</sup> or aggregate datasets<sup>218</sup>.</li> </ul>

Themes and sub-themes	Considerations (challenges and risks)	Recommendations
<p><b>E) DATA INFRASTRUCTURE, QUALITY, STORAGE AND SECURITY</b></p> <p><b>Security</b></p> <ul style="list-style-type: none"> <li>✓ For cloud data- issue of integrity and exploitation of data by service provider and its employees<sup>222,274-276</sup>, cloud attacks<sup>277</sup>,</li> <li>✓ User identity spoofing<sup>278</sup>,</li> <li>✓ Data tampering<sup>279</sup>,</li> <li>✓ Denial of service<sup>280</sup>,</li> <li>✓ Unlawful access to database and infiltration of the system<sup>278</sup>,</li> <li>✓ Danger of re-identification of de-identified data<sup>281</sup>.</li> </ul>	<p>The success of data security (including cybersecurity) will depend on good governance that ensure compliance with safety regulation by all parties.</p> <p>➤ A need to develop policies on data security that mandate the custodians of data to protect it from abuse, unauthorised access and tampering, loss or unlawful disclosure<sup>72</sup>.</p> <p>➤ Privacy protection provide a notification in the event of breach of privacy due to unauthorised access, loss or disclosure of information in the care of a legal data custodian<sup>273</sup>.</p> <p>➤ Establishment of remote access controlled data centres, and good monitoring systems<sup>107,283</sup>.</p>	
<p><b>Sustainability of databanks</b></p> <ul style="list-style-type: none"> <li>✓ Challenges to sustainability include the cost of maintaining a central databank, issues of social licence and public distrust and limited oversight of commercial data, data ownership, intellectual property, commercial secrecy, insufficient transparency, and profiteering<sup>300</sup>.</li> <li>✓ Funding constraints also have implications on data cleaning, analysis, storage, which may ultimately affect the data quality.</li> </ul>	<p>➤ Researchers must plan for sustainability of databank before their studies commence<sup>104,117</sup>.</p> <p>➤ A need for consistent application of data policies throughout its lifespan including promoting scientific and ethical integrity on data<sup>47</sup>.</p> <p>➤ Governments and funders must increase financial sustainability to support capacity and infrastructure for databanks and data sharing<sup>167,169</sup>.</p> <p>➤ There is also a need to invest in human capital<sup>305,306,308</sup>.</p> <p>➤ Other ways of ensuring sustainability of databanks is through obtaining appropriate liability insurance<sup>252</sup>.</p> <p>➤ Public-private-partnership in data management can improve for innovation and development and sustainability of databanks<sup>300</sup>. A good example can be drawn from , the European Union's Data Protection Regulation<sup>300</sup>.</p>	
<p><b>F) DATA HARMONIZATION</b></p> <ul style="list-style-type: none"> <li>✓ Data repositories in Africa are disintegrated. Consortia are often not homogeneously impractical to developing consortium specific data sharing guidelines.</li> <li>✓ Many consortia have specific guidelines which may make it difficult to integrate data.</li> <li>✓ Data repositories in Africa largely sits in research institutions or NGOs or generalist data repository that are not specific to any discipline; and project or programme specific repository<sup>93</sup>.</li> </ul>	<p>➤ Develop an integrated multidisciplinary guideline that is flexible for public and population health. And which will allow multilayer data sharing for public good<sup>10,133</sup></p> <p>➤ Develop stakeholder-centric ecosystems in terms of its principles and policies seeking to efficiently meet the needs of its members<sup>133</sup>.</p> <p>➤ Stakeholders must work together, through a bottom up approach, to find common grounds, policies, and solutions to harmonization challenges<sup>235,309</sup>. Examples include success of GA4GH, P3G and H3Africa<sup>423,10,311</sup>.</p> <p>➤ Develop a flexible guideline/policy interoperability and convergence between partners to facilitate collaboration and platform efficiency<sup>330</sup>.</p>	

## Data availability

### Underlying data

Zenodo: Public and Population Health data sharing in Africa – views of academics and researchers, <https://doi.org/10.5281/zenodo.5155880><sup>352</sup>.

This project contains the following underlying data:

De-identified transcripts of the interviews with the 24 key informants

### Extended data

Zenodo: Interview Guide Used in the Key Informant Interviews: Public and Population Health data sharing in Africa - Views of

Academics and Researchers <https://doi.org/10.5281/zenodo.5168457><sup>39</sup>.

This project contains the following extended data:

- Interview guide use in the key informant interviews

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

## Acknowledgements

We thank the public and population health researchers who participated in the various consultative activities supporting this research.

## References

- Kasprowitz VO, Chopera D, Waddilove KD, et al.: **African-led health research and capacity building- is it working?** *BMC Public Health*. 2020; **20**(1): 1104. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Igumbor JO, Bosire EN, Basera TJ, et al.: **CARTA fellows' scientific contribution to the African public and population Health Research agenda (2011 to 2018)**. *BMC Public Health*. 2020; **20**(1): 1030. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Uthman OA, Wiysonge CS, Ota MO, et al.: **Increasing the value of health research in the WHO African Region beyond 2015 - Reflecting on the past, celebrating the present and building the future: A bibliometric analysis**. *BMJ Open*. 2015; **5**(3): e006340. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nachega JB, Uthman OA, Ho YS, et al.: **Current status and future prospects of epidemiology and public health training and research in the WHO African region**. *Int J Epidemiol*. 2012; **41**(6): 1829–1846. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chirwa TF, Zingoni ZM, Munyewende P, et al.: **Developing excellence in biostatistics leadership, training and science in Africa: How the Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) training unites expertise to deliver excellence [version 2; peer review: 2 approved, 1 approved with reservations]**. *AAS Open Res*. 2020; **3**: 51. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sankoh O, Byass P: **The INDEPTH Network: filling vital gaps in global epidemiology**. *Int J Epidemiol*. 2012; **41**(3): 579–588. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mulder N, Abimiku A, Adebamowo SN, et al.: **H3Africa: Current perspectives**. *Pharmgenomics Pers Med*. 2018; **11**: 59–66. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fauci AS, Eisinger RW: **PEPFAR - 15 years and counting the lives saved**. *N Engl J Med*. 2018; **378**(4): 314–316. [PubMed Abstract](#) | [Publisher Full Text](#)
- ICF: **The Demographic and Health Surveys (DHS) Program**. [Reference Source](#)
- Kalkman S, Mostert M, Gerlinger C, et al.: **Responsible data sharing in international health research: A systematic review of principles and norms**. *BMC Med Ethics*. 2019; **20**(1): 21. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- European Commission: **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 2.1**. 2016; 1–10. [Reference Source](#)
- Sankar PL, Parker LS: **The Precision Medicine Initiative's All of Us Research Program: An agenda for research on its ethical, legal, and social issues**. *Genet Med*. 2017; **19**(7): 743–750. [PubMed Abstract](#) | [Publisher Full Text](#)
- AAAS: **Precision Medicine in China**. *Conn Med*. 2016; **39**(3): 163–165, 184. [Reference Source](#)
- Australian Commission on Safety and Quality in Healthcare: **Operating Principles and Technical Standards for Australian Clinical Quality Registries**. 2008. [Reference Source](#)
- Milne BJ, Atkinson J, Blakely T, et al.: **Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI)**. *Int J Epidemiol*. 2019; **48**(3): 677–677e. [PubMed Abstract](#) | [Publisher Full Text](#)
- The National Academies Press: **Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease**. 2011. [Reference Source](#)
- Institute of Medicine (US) Roundtable on Evidence-Based Medicine: **The Learning Healthcare System**. Washington, D.C.: National Academies Press, 2007. [Reference Source](#)
- Duermeijer C, Amir M, Schoombe L, et al.: **Africa generates less than 1% of the world's research; data analytics can change that**. 2018; 1–15. [Reference Source](#)
- Fonn S, Ayiro LP, Cotton P, et al.: **Repositioning Africa in global knowledge production**. *Lancet*. 2018; **392**(10153): 1163–1166. [PubMed Abstract](#) | [Publisher Full Text](#)
- Healthcare IT News Staff: **The biggest healthcare data breaches of 2018 (so far)**. *Healthcare IT News*. 2018. [Reference Source](#)
- El Emam K, Jonker E, Arbuckle L, et al.: **A Systematic Review of Re-Identification Attacks on Health Data**. *PLoS One*. 2011; **6**(12): e28071. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hagmann J: **Information governance - beyond the buzz**. *Rec Manag J*. 2013; **23**(3): 228–240. [Publisher Full Text](#)
- Smallwood RF: **Information Governance: Concepts, Strategies and Best Practices**. Hoboken, NJ: John Wiley and Sons Ltd, 2014. [Reference Source](#)
- Chan T, Di Iorio CT, De Lusignan S, et al.: **UK National Data Guardian for Health and Care's Review of Data Security: Trust, better security and opt-outs**. *J Innov Health Inform*. 2016; **23**(3): 627–632. [PubMed Abstract](#) | [Publisher Full Text](#)
- HIPAA: **Largest Healthcare Data Breaches of 2017**. *HI PAJ*. [Reference Source](#)
- Erikainen S, Friesen P, Rand L, et al.: **Public involvement in the governance of population-level biomedical research: Unresolved questions and future directions**. *J Med Ethics*. 2020; medethics-2020-106530. [PubMed Abstract](#) | [Publisher Full Text](#)
- Mostert M, Bredenoord AL, Bieshaar MCIH, et al.: **Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach**. *Eur J Hum Genet*. 2016; **24**(7): 956–960. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Knoppers BM, Leroux T, Doucet H, et al.: **Framing Genomics, Public Health Research and Policy: Points to Consider**. *Public Health Genomics*. 2010; **13**(4): 224–234. [PubMed Abstract](#) | [Publisher Full Text](#)
- Joly Y, Dalpé G, So D, et al.: **Fair shares and sharing fairly: A survey of public**

- views on open science, informed consent and participatory research in biobanking. *PLoS One*. 2015; **10**(7): e0129893.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Sanderson SC, Brothers KB, Mercaldo ND, et al.: **Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US.** *Am J Hum Genet*. 2017; **100**(3): 414–427.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Wong G, Greenhalgh T, Westhorp G, et al.: **RAMESES publication standards: meta-narrative reviews.** *BMC Med*. 2013; **11**: 20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Kuhn TS, Hacking I: **The Structure of Scientific Revolutions.** 2013.
33. Greenhalgh T, Robert G, Macfarlane F, et al.: **Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations.** *Milbank Q*. 2004; **82**(4): 581–629.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Greenhalgh T, Robert G, Macfarlane F, et al.: **Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review.** *Soc Sci Med*. 2005; **61**(2): 417–430.  
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Leximancer: **Leximancer User Guide Release 5.0.** 2019.
36. Biroscak BJ, Scott JE, Lindenberger JH, et al.: **Leximancer Software as a Research Tool for Social Marketers: Application to a Content Analysis.** *Soc Mar Q*. 2017; **23**(3): 223–231.  
[Publisher Full Text](#)
37. Wilk V, Soutar GN, Harrigan P: **Tackling social media data analysis: Comparing and contrasting QSR NVivo and Leximancer.** *Qual Mark Res*. 2019; **22**(2): 94–113.  
[Publisher Full Text](#)
38. Pulford J, Abomo P, Liani M, et al.: **DELTA Africa Learning Research Programme : Learning Report No . 3.** 2019; **3**(3).  
[Reference Source](#)
39. Igumbor J: **Interview Guide Use in the Key Informant Interviews: Public and Population Health data sharing in Africa - Views of Academics and Researchers.** *Zenodo*. 2021.  
<http://www.doi.org/10.5281/zenodo.5168457>
40. Corsi DJ, Neuman M, Finlay JE, et al.: **Demographic and health surveys: a profile.** *Int J Epidemiol*. 2012; **41**(6): 1602–1613.  
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Khan S, Hancioglu A: **Multiple Indicator Cluster Surveys: Delivering Robust Data on Children and Women across the Globe.** *Stud Fam Plann*. 2019; **50**(3): 279–286.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Abimiku A, Mayne ES, Joloba M, et al.: **H3Africa Biorepository Program: Supporting Genomics Research on African Populations by Sharing High-Quality Biospecimens.** *Biopreserv Biobank*. 2017; **15**(2): 99–102.  
[Publisher Full Text](#) | [Free Full Text](#)
43. Paltoo DN, Rodriguez LL, Feolo M, et al.: **Data use under the NIH GWAS Data Sharing Policy and future directions.** *Nat Genet*. 2014; **46**(9): 934–938.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Genome-Wide Association Studies: **Genome-Wide Association Studies (GWAS) - Frequently Asked Questions.** 2008; 1–8.  
[Reference Source](#)
45. NIH: **Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS).** *Intellectual Property*. 2008; 1–18.
46. Reichman OJ, Jones MB, Schildhauer MP: **Challenges and opportunities of open data in ecology.** *Science*. 2011; **331**(6018): 703–705.  
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Isasi R, Knoppers BM: **From banking to international governance: Fostering innovation in stem cell research.** *Stem Cells Int*. 2011; **2011**: 498132.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Zika E, Paci D, Rijkers-Defrasne S, et al.: **A European survey on biobanks: Trends and issues.** *Public Health Genomics*. 2011; **14**(2): 96–103.  
[PubMed Abstract](#) | [Publisher Full Text](#)
49. Hassan L, Dalton A, Hammond C, et al.: **A deliberative study of public attitudes towards sharing genomic data within NHS genomic medicine services in England.** *Public Underst. Sci*. 2020; **29**(7): 702–717.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. O'Doherty KC, Christofides E, Yen J, et al.: **If you build it, they will come: Unintended future uses of organised health data collections Donna Dickenson, Sandra Soo-jin Lee, and Michael Morrison.** *BMC Med Ethics*. 2016; **17**: 54.  
[Publisher Full Text](#)
51. Moodley K, Sibanda N, February K, et al.: **"It's my blood": ethical complexities in the use, storage and export of biological samples: perspectives from South African research participants.** *BMC Med Ethics*. 2014; **15**: 4.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Garrison NA, Sathe NA, Antommarrina AHM, et al.: **A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States.** *Genet Med*. 2016; **18**(7): 663–671.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Shabani M, Bezuidenhout L, Borry P: **Attitudes of research participants and the general public towards genomic data sharing: A systematic literature review.** *Expert Rev Mol Diagn*. 2014; **14**(8): 1053–1065.  
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Aitken M, De St Jorre J, Pagliari C, et al.: **Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies.** *BMC Med Ethics*. 2016; **17**(1): 73.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
55. Hill EM, Turner EL, Martin RM, et al.: **"Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study.** *BMC Med Res Methodol*. 2013; **13**(1): 72.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Kalkman S, Van Delden J, Banerjee A, et al.: **Patients' and public views and attitudes towards the sharing of health data for research: A narrative review of the empirical evidence.** *J Med Ethics*. 2019; medethics-2019-105651.  
[PubMed Abstract](#) | [Publisher Full Text](#)
57. Aitken M, McAteer G, Davidson S, et al.: **Public Preferences regarding Data Linkage for Health Research: A discrete choice experiment.** *Int J Popul Data Sci*. 2018; **3**(1): 429.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. Howe N, Giles E, Newbury-Birch D, et al.: **Systematic review of participants' attitudes towards data sharing: A thematic synthesis.** *J Heal Serv Res Policy*. 2018; **23**(2): 123–133.  
[PubMed Abstract](#) | [Publisher Full Text](#)
59. Lysaght T, Ballantyne A, Xafis V, et al.: **"Who is Watching the Watchdog?": Ethical Perspectives of Sharing Health-related Data for Precision Medicine in Singapore.** *BMC Med Ethics*. 2020; **21**(1): 118.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
60. Mutenherwa F, Wassenaar DR, De Oliveira T: **Ethical issues associated with HIV molecular epidemiology: A qualitative exploratory study using inductive analytic approaches.** *BMC Med Ethics*. 2019; **20**(1): 67.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
61. Ndebele PM, Wassenaar D, Munalula E, et al.: **Improving understanding of clinical trial procedures among low literacy populations: an intervention within a microbicide trial in Malawi.** *BMC Med Ethics*. 2012; **13**(1): 29.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
62. Burstein MD, Robinson JO, Hilsenbeck SG, et al.: **Pediatric data sharing in genomic research: Attitudes and preferences of parents.** *Pediatrics*. 2014; **133**(4): 690–697.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
63. Robinson JO, Slashinski MJ, Wang T, et al.: **Participants' recall and understanding of genomic research and large-scale data sharing.** *J Empir Res Hum Res Ethics*. 2013; **8**(4): 42–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
64. Fureman I, Meyers K, McLellan AT, et al.: **Evaluation of a video-supplement to informed consent: injection drug users and preventive HIV vaccine efficacy trials.** *AIDS Educ Prev*. 1997; **9**(4): 330–341.  
[PubMed Abstract](#)
65. Corneli AL, Sorenson JR, Bentley ME, et al.: **Improving participant understanding of informed consent in an HIV-prevention clinical trial: A comparison of methods.** *AIDS Behav*. 2012; **16**(2): 412–421.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
66. Verheggen FW, Van Wijmen FC: **Informed consent in clinical trials.** *Health Policy*. 1996; **36**(2): 131–153.  
[PubMed Abstract](#) | [Publisher Full Text](#)
67. Lindegger G, Milford C, Slack C, et al.: **Beyond the checklist: Assessing understanding for HIV vaccine trial participation in South Africa.** *J Acquir Immune Defic Syndr*. 2006; **43**(5): 560–566.  
[PubMed Abstract](#) | [Publisher Full Text](#)
68. Lindegger G, Richter LM: **HIV vaccine trials: Critical issues in informed consent.** *S Afr J Sci*. 2000; **96**(6): 313–317.  
[PubMed Abstract](#)
69. Hellman S, Hellman DS: **Sounding board of mice not men: Problems of the randomised clinical trial.** *N Engl J Med*. 1991; **324**(22): 1585–1589.
70. Rani M, Buckley BS: **Systematic archiving and access to health research data: rationale, current status and way forward.** *Bull World Health Organ*. 2012; **90**(12): 932–939.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
71. Jones KH, Heys SM, Daniels H, et al.: **Exploring barriers and solutions in advancing cross-centre population data science.** *Int J Popul Data Sci*. 2019; **4**(1): 1109.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
72. National Institute of Health: **NIH Data Sharing Policy.** 2003.  
[Reference Source](#)
73. Wellcome Trust: **Data, software and materials management and sharing policy.** 2017; 1–2.  
[Reference Source](#)
74. SHERPA - Juliet: **Research funders' open access policies.** 2008.  
[Reference Source](#)
75. Rollando P, Parc C, Naudet F, et al.: **[Data sharing policies of clinical trials funders in France].** *Therapie*. 2020; **75**(6): 527–536.  
[PubMed Abstract](#) | [Publisher Full Text](#)
76. University of Cambridge: **Funders' Policies.** (accessed Dec. 16, 2020).  
[Reference Source](#)
77. F. I. of M: **Committee on the Outcome and Impact Evaluation of Global HIV/AIDS Programs Implemented Under the Lantos-Hyde Act of 2008; Board on Global Health; Board on Children, Youth, and Families; Institute**

- of Medicine. Committee on the Outcome and Impact Evaluation, *Evaluation of PEPFAR*. 2013.  
[Reference Source](#)
78. Simon GE, Shortreed SM, Coley RY, *et al.*: **Assessing and Minimizing Re-identification Risk in Research Data Derived from Health Care Records.** *EGEMS (Wash DC)*. 2019; 7(1): 6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  79. National Institute of Health: **The National Institute of Mental Health Data Archive.** (accessed Jan. 03, 2020).  
[Reference Source](#)
  80. Sielemann K, Hafner A, Pucker B: **The reuse of public datasets in the life sciences: Potential risks and rewards.** *PeerJ*. 2020; 8: e9954.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  81. Tenopir C, Rice NM, Allard S, *et al.*: **Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide.** *PLoS One*. 2020; 15(3): e0229003.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  82. Moon LA: **Factors influencing health data sharing preferences of consumers: A critical review.** *Heal Policy Technol*. 2017; 6(2): 169–187.  
[Publisher Full Text](#)
  83. Ancker JS, Edwards AM, Miller MC, *et al.*: **Consumer Perceptions of Electronic Health Information Exchange.** *Am J Prev Med*. 2012; 43(1): 76–80.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  84. Park H, Lee SJ, Kim Y, *et al.*: **Patients' perceptions of a health information exchange: A pilot program in South Korea.** *Int J Med Inform*. 2013; 82(2): 98–107.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  85. Teixeira PA, Gordon P, Camhi E, *et al.*: **HIV patients' willingness to share personal health information electronically.** *Patient Educ Couns*. 2011; 84(2): e9–e12.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  86. Caine K, Hanania R: **Patients want granular privacy control over health information in electronic medical records.** *J Am Med Inform Assoc*. 2013; 20(1): 7–15.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  87. Dhopeswarkar RV, Kern LM, O'Donnell HC, *et al.*: **Health Care Consumers' Preferences Around Health Information Exchange.** *Ann Fam Med*. 2012; 10(5): 428–434.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  88. Luchenski S, Balasanthiran A, Marston C, *et al.*: **Survey of patient and public perceptions of electronic health records for healthcare, policy and research: Study protocol.** *BMC Med Inform Decis Mak*. 2012; 12(1): 40.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  89. Patel VN, Dhopeswarkar RV, Edwards A, *et al.*: **Low-income, ethnically diverse consumers' perspective on health information exchange and personal health records.** *Inform Health Soc Care*. 2011; 36(4): 233–252.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  90. Walport M, Brest P: **Sharing research data to improve public health.** *Lancet*. 2011; 377(9765): 537–539.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  91. Consumers International: **The state of data protection rules around the world: A briefing FOR CONSUMER ORGANISATIONS.** 2018.  
[Reference Source](#)
  92. DLA Piper: **Data Protection Laws of the World Handbook.** *Nonscholar*. 2020; (accessed Dec. 29, 2020).  
[Reference Source](#)
  93. Staunton C, Adams R, Botes M, *et al.*: **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*. 2019; 109(7): 468–470.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  94. Government of UK Legislation: **Data Protection Act 2018.** 2018.  
[Reference Source](#)
  95. Raza S, Hall A: **Genomic medicine and data sharing.** *Br Med Bull*. 2017; 123(1): 35–45.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  96. McBeth S: **Access to Linked Administrative Data Through an Indigenous Cultural Lens.** *Int J Popul Data Sci*. 2020; 5(5).  
[Publisher Full Text](#)
  97. **Ethical issues facing medical research in developing countries.** Gambia Government/Medical Research Council Joint Ethical Committee. *Lancet*. 1998; 351: 286–7.  
[PubMed Abstract](#)
  98. Bhatt A: **Ethics committee composition.** *Perspect Clin Res*. 2012; 3(4): 146–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  99. Cleaton-Jones P, Wassenaar D: **Protection of human participants in health research - a comparison of some US Federal Regulations and South African Research Ethics guidelines.** *South Afr Med J*. 2010; 100(11): 712–6.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  100. Council of Europe: **Details of Treaty no. 164.** 1999.  
[Reference Source](#)
  101. The Biotechnology Centre of Oslo: **ACT 2008-06-20 no. 44: Act on medical and health research (the Health Research Act).** 2008.  
[Reference Source](#)
  102. World Medical Association: **Declaration of Helsinki - Ethical Principles of Medical Research Involving Human Subjects.** 2013.  
[Reference Source](#)
  103. Office of Population Health Genomics: **Guidelines for human biobanks, genetic research databases and associated data.** 2010; 1–37.  
[Reference Source](#)
  104. Pilat D, Fukasaku Y: **OECD Principles and Guidelines for Access to Research Data from Public Funding.** *Data Sci J*. 2007; 6: OD4–OD11.  
[Publisher Full Text](#)
  105. Powlledge TM: **Revisiting Bermuda.** *Genome Biol*. 2003; 4: spotlight-20030311–01.  
[Publisher Full Text](#)
  106. Expert Advisory Group on Data Access: **Governance of Data Access.** The Wellcome Trust, London. 2015.  
[Reference Source](#)
  107. Mascalzoni SW, Deborah ES, Dove YR, *et al.*: **International Charter of principles for sharing bio-specimens and data.** *Eur J Hum Genet*. 2015; 23(6): 721–728.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  108. Dhali A, Mahomed S, Sanne I: **Biobanks and human health research: Balancing progress and protections.** *South African J Bioeth Law*. 2015; 8(2): 55.  
[Publisher Full Text](#)
  109. Mouton Dorey C, Baumann H, Biller-Andorno N: **Patient data and patient rights: Swiss healthcare stakeholders' ethical awareness regarding large patient data sets - a qualitative study.** *BMC Med Ethics*. 2018; 19(1): 20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  110. Nansumba H, Ssewanyana I, Tai M, *et al.*: **Role of a regulatory and governance framework in human biological materials and data sharing in National Biobanks: Case studies from Biobank Integrating Platform, Taiwan and the National Biorepository, Uganda [version 2; peer review: 2 approved].** *Wellcome Open Res*. 2020; 4: 171.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  111. Peppercorn J, Shapira I, Deshields T, *et al.*: **Ethical aspects of participation in the Database of Genotypes and Phenotypes of the National Center for Biotechnology Information: The Cancer and Leukemia Group B Experience.** *Cancer*. 2012; 118(20): 5060–5068.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  112. Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization: **International Ethical Guidelines for Health-related Research Involving Humans.** CIOMS, Switzerland. 2016.  
[Reference Source](#)
  113. Nuffield Council on Bioethics: **The Collection, Linking and Use of Data in Health Care and Biomedical Research: Ethical Issues.** London: Nuffield Council on Bioethics. 2016.  
[Reference Source](#)
  114. World Medical Association: **Declaration of Taipei - Research on Health Databases, Big Data and Biobanks.** 2018.  
[Reference Source](#)
  115. Duchange N, Darquy S, d'Audiffret D, *et al.*: **Ethical management in the constitution of a European database for leukodystrophies rare diseases.** *Eur J Paediatr Neurol*. 2014; 18(5): 597–603.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  116. Kaye J: **The Tension Between Data Sharing and the Protection of Privacy in Genomics Research.** *Annu Rev Genomics Hum Genet*. 2012; 13(1): 415–431.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  117. OECD: **Recommendation of the Council on Human Biobanks and Genetic Research Databases.** 2009.  
[Reference Source](#)
  118. Baker DB, Kaye J, Terry SF: **Governance Through Privacy, Fairness, and Respect for Individuals.** *EGEMS (Wash DC)*. 2016; 4(2): 1207.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  119. Global Alliance for Genomics and Health (GA4GH): **Framework for Responsible Sharing of Genomic and Health-Related Data.** 2014.  
[Reference Source](#)
  120. Eckstein L, Chalmers D, Critchley C, *et al.*: **Australia: regulating genomic data sharing to promote public trust.** *Hum Genet*. 2018; 137(8): 583–591.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  121. Wiehe SE, Rosenman MB, Chartash D, *et al.*: **A Solutions-Based Approach to Building Data-Sharing Partnerships.** *EGEMS (Wash DC)*. 2018; 6(1): 20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  122. BRAINS (Brain Imaging in Normal Subjects) Expert Working Group, Shenkin SD, Pernet C, *et al.*: **Improving data availability for brain image biobanking in healthy subjects: Practice-based suggestions from an international multidisciplinary working group.** *Neuroimage*. 2017; 153: 399–409.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  123. Nicholas N, Nicholas S: **Understanding confidentiality and the law on access to medical records.** *Obstet Gynaecol Reprod Med*. 2010; 20(5): 161–163.  
[Publisher Full Text](#)
  124. Hawkes N: **Cameron promotes new partnership between research, industry, and the NHS.** *BMJ*. 2011; 343: d7956.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  125. Cassell J, Young A: **Why we should not seek individual informed consent for**



- participation in health services research. *J Med Ethics*. 2002; **28**(5): 313–317. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
126. Rawlins M, Academy of Medical Sciences: **A new pathway for the regulation and governance of health research**. *Acad Med Sci*. 2011; 121. [Reference Source](#)
127. Miller FG: **Research on medical records without informed consent**. *J Law Med Ethics*. 2008; **36**(3): 560–566. [PubMed Abstract](#) | [Publisher Full Text](#)
128. Shepherd E, Sexton A, Duke-Williams O, et al.: **Risk identification and management for the research use of government administrative data**. *Rec Manag J*. 2019; **30**(1): 101–123. [Publisher Full Text](#)
129. Metcalfe C, Martin RM, Noble S, et al.: **Low risk research using routinely collected identifiable health information without informed consent: Encounters with the patient information advisory group**. *J Med Ethics*. 2008; **34**(1): 37–40. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
130. Iversen A, Liddell K, Fear N, et al.: **Consent, confidentiality, and the Data Protection Act**. *BMJ*. 2006; **332**(7534): 165–169. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
131. Hansson MG: **Need for a wider view of autonomy in epidemiological research**. *BMJ*. 2010; **340**: c2335. [PubMed Abstract](#) | [Publisher Full Text](#)
132. Borgman CL: **Research Data: Who Will Share What, with Whom, When, and Why?** *SSRN Electron J*. 2010. [Publisher Full Text](#)
133. Deverka PA, Majumder MA, Villanueva AG, et al.: **Creating a data resource: What will it take to build a medical information commons?** *Genome Med*. 2017; **9**(1): 84. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
134. Rosenbaum S: **Data governance and stewardship: Designing data stewardship entities and advancing data access**. *Health Serv Res*. 2010; **45**(5 Pt 2): 1442–1455. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
135. Ford E, Kazempour Y, Cooper MJF, et al.: **Media content analysis of general practitioners' reactions to care data expressed in the media: What lessons can be learned for future NHS data-sharing initiatives?** *BMJ Open*. 2020; **10**(9): e038006. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
136. Chokshi DA, Parker M, Kwiatkowski DP: **Data sharing and intellectual property in a genomic epidemiology network: Policies for large-scale research collaboration**. *Bull World Health Organ*. 2006; **84**(5): 382–387. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
137. Xafis V, Schaefer GO, Labude MK, et al.: **An Ethics Framework for Big Data in Health and Research**. *Asian Bioeth Rev*. 2019; **11**(3): 227–254. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
138. Lysaught MT: **Respect: Or, how respect for persons became respect for autonomy**. *J Med Philos*. 2004; **29**(6): 665–680. [PubMed Abstract](#) | [Publisher Full Text](#)
139. Kuehn BM: **IOM Outlines Framework for Clinical Data Sharing, Solicits Input**. *JAMA*. 2014; **311**(7): 665. [PubMed Abstract](#) | [Publisher Full Text](#)
140. Auffray C, Balling R, Barroso I, et al.: **Making sense of big data in health research: Towards an EU action plan**. *Genome Med*. 2016; **8**(1): 71. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
141. Allen C, Des Jardins TR, Heider A, et al.: **Data Governance and Data Sharing Agreements for Community-Wide Health Information Exchange: Lessons from the Beacon Communities**. *EGEMS (Wash DC)*. 2014; **2**(1): 1057. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
142. Lemke AA, Wolf WA, Hebert-Beirne J, et al.: **Public and biobank participant attitudes toward genetic research participation and data sharing**. *Public Health Genomics*. 2010; **13**(6): 368–377. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
143. Richter G, Borzikowsky C, Lieb W, et al.: **Patient views on research use of clinical data without consent: Legal, but also acceptable?** *Eur J Hum Genet*. 2019; **27**(6): 841–847. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
144. McCormick N, Hamilton CB, Koehn CL, et al.: **Canadians' views on the use of routinely collected data in health research: A patient-oriented cross-sectional survey**. *CMAJ Open*. 2019; **7**(2): E203–E209. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
145. O'Brien EC, Rodriguez AM, Kum HC, et al.: **Patient perspectives on the linkage of health data for research: Insights from an online patient community questionnaire**. *Int J Med Inform*. 2019; **127**: 9–17. [PubMed Abstract](#) | [Publisher Full Text](#)
146. Colombo C, Roberto A, Krljeza-Jeric K, et al.: **Sharing individual participant data from clinical studies: A cross-sectional online survey among Italian patient and citizen groups**. *BMJ Open*. 2019; **9**(2): e024863. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
147. Darquy S, Moutel G, Lapointe AS, et al.: **Patient/family views on data sharing in rare diseases: Study in the European LeukoTreat project**. *Eur J Hum Genet*. 2016; **24**(3): 338–343. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
148. McCormack P, Kole A, Gainotti S, et al.: **'You should at least ask'. the expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research**. *Eur J Hum Genet*. 2016; **24**(10): 1403–1408. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
149. Dimitropoulos L, Patel V, Scheffler SA, et al.: **Public attitudes toward health information exchange: Perceived benefits and concerns**. *Am J Manag Care*. 2011; **17**(12 Spec No.): SP111–6. [PubMed Abstract](#)
150. Patel VN, Dhopeshwarkar RV, Edwards A, et al.: **Consumer support for health information exchange and personal health records: A regional health information organization survey**. *J Med Syst*. 2012; **36**(3): 1043–1052. [PubMed Abstract](#) | [Publisher Full Text](#)
151. Schwartz PH, Caine K, Alpert SA: **Patient Preferences in Controlling Access to Their Electronic Health Records: a Prospective Cohort Study in Primary Care**. *J Gen Intern Med*. 2015; **30** Suppl 1(Suppl 1): S25–30. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
152. Marquard JL, Brennan PF: **Crying wolf: Consumers may be more willing to share medication information than policymakers think**. *J Healthc Inf Manag*. 2009; **23**(2): 26–32. [PubMed Abstract](#)
153. Dawson L, Benbow N, Fletcher FE, et al.: **Addressing Ethical Challenges in US-Based HIV Phylogenetic Research**. *J Infect Dis*. 2020; **222**(12): 1997–2006. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
154. Tindana P: **H3Africa Guidelines for Community Engagement**. 2017. [Reference Source](#)
155. Alison Paprica P, Sutherland E, Smith A, et al.: **Essential requirements for establishing and operating data trusts: Practical guidance co-developed by representatives from fifteen canadian organizations and initiatives**. *Int J Popul Data Sci*. 2020; **5**(1): 1353. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
156. Antommaria AHM, Brothers KB, Myers JA, et al.: **Parents' attitudes toward consent and data sharing in biobanks: A multisite experimental survey**. *AJOB Empir Bioeth*. 2018; **9**(3): 128–142. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
157. Overby CL, Maloney KA, Alestock TD, et al.: **Prioritizing approaches to engage community members and build trust in biobanks: A survey of attitudes and opinions of adults within outpatient practices at the university of Maryland**. *J Pers Med*. 2015; **5**(3): 264–279. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
158. Fecher B, Friesike S, Hebing M, et al.: **A reputation economy: how individual reward considerations trump systemic arguments for open access to data**. *Palgrave Commun*. 2017; **3**(1): 17051. [Publisher Full Text](#)
159. Piwowar HA, Chapman WW: **Public sharing of research datasets: A pilot study of associations**. *J Informetr*. 2010; **4**(2): 148–156. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
160. Kim Y: **Institutional and Individual Influences on Scientists' Data Sharing Behaviors**. 2013. [Reference Source](#)
161. Terr RF, Littler K, Olliaro PL: **Sharing health research data – the role of funders in improving the impact [version 2; peer review: 3 approved]**. *F1000Res*. 2018; **7**: 1641. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
162. Karczewski KJ, Tatonetti NP, Manrai AK, et al.: **Methods to ensure the reproducibility of biomedical research**. *Pacific Symp Biocomput*. 2017; **22**(212679): 117–119. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
163. Iqbal SA, Wallach JD, Khoury MJ, et al.: **Reproducible Research Practices and Transparency across the Biomedical Literature**. *PLoS Biol*. 2016; **14**(1): e1002333. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
164. Damalas D, Kalyvioti G, Sabatella EC, et al.: **Open data in the life sciences: The 'Selfish Scientist Paradox'**. *Ethics Sci Environ Polit*. 2018; **18**(1): 27–36. [Publisher Full Text](#)
165. van Panhuis WG, Paul P, Emerson C, et al.: **A systematic review of barriers to data sharing in public health**. *BMC Public Health*. 2014; **14**(1): 1144. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
166. Piwowar HA, Bechic MJ, Bilofsky H, et al.: **Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers**. *PLoS Med*. 2008; **5**(9): e183. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
167. Editorial: **Vital statistics**. *Nature*. 2013; **494**(7437): 281. [PubMed Abstract](#) | [Publisher Full Text](#)
168. United States: **Global Health: Challenges in Improving Infectious Disease Surveillance Systems: Report to Congressional Requesters**. 2001. [Reference Source](#)
169. AbouZahr C, Cleland J, Coullare F, et al.: **The way forward**. *Lancet*. 2007; **370**(9601): 1791–1799. [PubMed Abstract](#) | [Publisher Full Text](#)
170. Bourdieu P: **Homo Academicus**. *Les Éditions de Minuit*. 1984. [Reference Source](#)
171. Rowhani-Farid A, Allen M, Barnett AG: **What incentives increase data sharing**

- in health and medical research? A systematic review. *Res Integr Peer Rev*. 2017; **2**(1): 4.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
172. Cochrane: **The Cochrane-REWARD prize for reducing waste in research**. *Cochrane*. 2019.  
[Reference Source](#)
173. Pisani E, AbouZahr C: **Sharing health data: good intentions are not enough**. *Bull World Health Organ*. 2010; **88**(6): 462–466.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
174. Raff JW: **The San Francisco Declaration on Research Assessment**. *Biol Open*. 2013; **2**(6): 533–534.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
175. Jones KH, Heys S, Tingay KS, *et al.*: **The Good, the Bad, the Clunky: Improving the Use of Administrative Data for Research**. *Int J Popul Data Sci*. 2019; **4**(1): 587.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
176. Brase J: **DataCite - A Global Registration Agency for Research Data**. In *2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*. 2009; 257–261.  
[Publisher Full Text](#)
177. Data Citation Synthesis Group: **Joint Declaration of Data Citation Principles**. 2014.  
[Publisher Full Text](#)
178. Hate K, Meherally S, More NS, *et al.*: **Sweat, Skepticism, and Uncharted Territory: A Qualitative Study of Opinions on Data Sharing among Public Health Researchers and Research Participants in Mumbai, India**. *J Empir Res Hum Res Ethics*. 2015; **10**(3): 239–250.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
179. Kidwell MC, Lazarević LB, Baranski E, *et al.*: **Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency**. *PLoS Biol*. 2016; **14**(5): e1002456.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
180. Wellcome Trust: **Summary of Funders' data sharing policies**. 2012.  
[Reference Source](#)
181. USAID: **ADS Chapter 579: USAID Development Data**. 2020.  
[Reference Source](#)
182. Abayomi A, Christoffels A, Grewal R, *et al.*: **Challenges of biobanking in South Africa to facilitate indigenous research in an environment burdened with human immunodeficiency virus, tuberculosis, and emerging noncommunicable diseases**. *Biopreserv Biobank*. 2013; **11**(6): 347–354.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
183. Ramsay SA, de Vries M, Soodyall J, *et al.*: **Ethical issues in genomic research on the African continent: experiences and challenges to ethics review committees**. *Hum Genomics*. 2014; **8**(1): 15.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
184. Birnbaum D, Borycki E, Karras BT, *et al.*: **Addressing Public Health informatics patient privacy concerns**. *Clin Gov An Int J*. 2015; **20**(2): 91–100.  
[Publisher Full Text](#)
185. European Union: **General Data Protection Regulation (GDPR)**. *Euratom*. 2016; **2001**: 20–30.  
[Reference Source](#)
186. Mazor KM, Richards A, Gallagher M: **Stakeholders' views on data sharing in multicenter studies**. *J Comp Eff Res*. 2017; **6**(6): 537–547.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
187. McGuire AI, Hamilton JA, Lunstroth R, *et al.*: **DNA data sharing: research participants' perspectives**. *Genet Med*. 2008; **10**(1): 46–53.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
188. National Data Guardian for Care and Health: **Review of Data Security, Consent and Opt-Outs**. 2016.  
[Reference Source](#)
189. The Office for Civil Rights (OCR), Malin B: **Guidance Regarding Methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule**. *Health Information Privacy*. 2012; 1–32.  
[Reference Source](#)
190. NCVHS: **R42 - Recommendations on De-identification of Protected Health Information under HIPAA**. 2017; **188**: 1–17.  
[Reference Source](#)
191. Malin B, Loukides G, Benitez K, *et al.*: **Identifiability in biobanks: Models, measures, and mitigation strategies**. *Hum Genet*. 2011; **130**(3): 383–392.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
192. Wang S: **A community effort to protect genomic data sharing, collaboration and outsourcing**. *NPJ Genomic Med*. 2017; **2**(1): 33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
193. Xafis V, Labude MK: **Openness in Big Data and Data Repositories: The Application of an Ethics Framework for Big Data in Health and Research**. *Asian Bioeth Rev*. 2019; **11**(3): 255–273.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
194. Gymrek M, McGuire AI, Golan D, *et al.*: **Identifying personal genomes by surname inference**. *Science*. 2013; **339**(6117): 321–324.  
[PubMed Abstract](#) | [Publisher Full Text](#)
195. Personal Data Protection Commission Singapore: **Advisory Guidelines on the PDPA for Selected Topics**. *Personal Data Protection Commission Singapore*. Singapore, 2018.  
[Reference Source](#)
196. Global Alliance for Genomics and Health (GA4GH): **Global Alliance for Genomics and Health: Privacy and Security Policy**.
197. Garfinkel SL: **De-identification of personal information**. Gaithersburg, MD, 2015.  
[Reference Source](#)
198. Mbuthia D, Molyneux S, Njue M, *et al.*: **Kenyan health stakeholder views on individual consent, general notification and governance processes for the re-use of hospital inpatient data to support learning on healthcare systems**. *BMC Med Ethics*. 2019; **20**(1): 3.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
199. Fiscella K, Tobin JN, Carroll JK, *et al.*: **Ethical oversight in quality improvement and quality improvement research: New approaches to promote a learning health care system Ethics in Biomedical Research**. *BMC Med Ethics*. 2015; **16**(1): 63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
200. Mittelstadt B, Fairweather B, Shaw M, *et al.*: **Ethical issues in Patient Safety Research: Interpreting existing guidance**. *World Heal Organ*. 2013; 41.  
[Reference Source](#)
201. Faden R, Kass N, Whicher D, *et al.*: **Ethics and informed consent for comparative effectiveness research with prospective electronic clinical data**. *Med Care*. 2013; **51**(8 SUPPL.3): S53–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
202. Christofides E, Muise A, Desmarais S: **Information disclosure and control on Facebook: are they two sides of the same coin or two different processes?** *Cyberpsychol Behav*. 2009; **12**(3): 341–345.  
[PubMed Abstract](#) | [Publisher Full Text](#)
203. Acquisti A, Gross R: **Predicting Social Security numbers from public data**. *Proc Natl Acad Sci U S A*. 2009; **106**(27): 10975–10980.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
204. Jernigan C, Mistree BFT: **Gaydar: Facebook friendships expose sexual orientation**. *First Monday*. 2009; **14**(10).  
[Publisher Full Text](#)
205. Kahn JP, Vayena E, Mastroianni AC: **Opinion: Learning as we go: Lessons from the publication of Facebook's social-computing research**. *Proc Natl Acad Sci USA*. 2014; **111**(38): 13677–13679.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
206. Zarate OA, Brody JG, Brown P, *et al.*: **Balancing Benefits and Risks of Immortal Data: Participants' Views of Open Consent in the Personal Genome Project**. *Hastings Cent Rep*. 2016; **46**(1): 36–45.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
207. Goho SA: **The legal implications of report back in household exposure studies**. *Environ Health Perspect*. 2016; **124**(11): 1662–1670.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
208. Keerie C, Tuck C, Milne G, *et al.*: **Data sharing in clinical trials - practical guidance on anonymising trial datasets**. *Trials*. 2018; **19**(1): 25.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
209. Hrynaskiewicz I, Norton ML, Vickers AJ, *et al.*: **Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers**. *BMJ*. 2010; **340**: c181.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
210. Cato KD, Bocking W, Larson E: **Did I tell you that? Ethical issues related to using computational methods to discover non-disclosed patient characteristics**. *J Empir Res Hum Res Ethics*. 2016; **11**(3): 214–219.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
211. Currie J: **"Big data" versus "Big brother": On the appropriate use of large-scale data collections in pediatrics**. *Pediatrics*. 2013; **131**(Suppl 2): S127–32.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
212. Grant JM, Mottet LA, Tanis J, *et al.*: **National Transgender Discrimination Survey Report on Health and Health Care**. *Natl Cent Transgender Equal*. 2010; **5**: 23.  
[Reference Source](#)
213. Kosenko K, Rintamaki L, Raney S, *et al.*: **Transgender patient perceptions of stigma in health care contexts**. *Med Care*. 2013; **51**(9): 819–822.  
[PubMed Abstract](#) | [Publisher Full Text](#)
214. van Ryn M: **Research on the provider contribution to race/ethnicity disparities in medical care**. *Med Care*. 2002; **40**(1 Suppl): 1140–51.  
[PubMed Abstract](#) | [Publisher Full Text](#)
215. Van Ryn M, Burke J: **The effect of patient race and socio-economic status on physicians' perceptions of patients**. *Soc Sci Med*. 2000; **50**(6): 813–828.  
[PubMed Abstract](#) | [Publisher Full Text](#)
216. Kaye J, Whitley EA, Lund D, *et al.*: **Dynamic consent: A patient interface for twenty-first century research networks**. *Eur J Hum Genet*. 2015; **23**(2): 141–146.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
217. Kaye J, Hawkins N: **Data sharing policy design for consortia: Challenges for sustainability**. *Genome Med*. 2014; **6**(1): 4.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
218. Isasi R, Knoppers BM, Andrews PW, *et al.*: **Disclosure and management of research findings in stem cell research and banking: Policy statement**. *Regen Med*. 2012; **7**(3): 439–448.  
[PubMed Abstract](#) | [Publisher Full Text](#)
219. Kaye J, Moraia LB, Mitchell C, *et al.*: **Access Governance for Biobanks: The**

- Case of the BioSHaRE-EU Cohorts.** *Biopreserv Biobank.* 2016; **14**(3): 201–206.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
220. Dyke SO, Hubbard TJ: **Developing and implementing an institute-wide data sharing policy.** *Genome Med.* 2011; **3**(9): 60.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
221. Desai T, Ritchie F, Welpton R: **Five Safes: Designing Data Access for Research.** *Econ Work Pap Ser.* 2016; **1601**: 28.  
[Reference Source](#)
222. Zook M, Barocas S, Boyd D, et al.: **Ten simple rules for responsible big data research.** *PLoS Comput Biol.* 2017; **13**(3): e1005399.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
223. Caldicott Committee: **Report on the review of patient-identifiable information.** London: NHS Executive, 1999; 1–137.
224. Department of Health: **To Share or Not to Share? Information Governance Review.** Department of Health, 2013.  
[Reference Source](#)
225. Holman CD, Bass AJ, Rosman DL, et al.: **A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system.** *Aust Heal Rev.* 2008; **32**(4): 766–777.  
[PubMed Abstract](#) | [Publisher Full Text](#)
226. Kariminia A, Butler TG, Corben SP, et al.: **Extreme cause-specific mortality in a cohort of adult prisoners–1988 to 2002: a data-linkage study.** *Int J Epidemiol.* 2007; **36**(2): 310–316.  
[PubMed Abstract](#) | [Publisher Full Text](#)
227. Young TK, Kliewer E, Blanchard J, et al.: **Monitoring disease burden and preventive behavior with data linkage: Cervical cancer among Aboriginal people in Manitoba, Canada.** *Am J Public Health.* 2000; **90**(9): 1466–1468.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
228. Fischbacher CM, Bhopal R, Povey C, et al.: **Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: Myocardial infarction in South Asians.** *BMC Public Health.* 2007; **7**: 142.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
229. Veugelers PJ, Yip AM, Kephart G: **Proximate and contextual socioeconomic determinants of mortality: Multilevel approaches in a setting with universal health care coverage.** *Am J Epidemiol.* 2001; **154**(8): 725–732.  
[PubMed Abstract](#) | [Publisher Full Text](#)
230. Jutte DP, Roos LL, Brownell MD: **Administrative record linkage as a tool for public health research.** *Annu Rev Public Health.* 2011; **32**: 91–108.  
[PubMed Abstract](#) | [Publisher Full Text](#)
231. Schnell R, Bachteler T, Reiher J: **Privacy-preserving record linkage using Bloom filters.** *BMC Med Inform Decis Mak.* 2009; **9**(1): 41.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
232. Wolfson M, Wallace SE, Masca N, et al.: **DataSHIELD: Resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data.** *Int J Epidemiol.* 2010; **39**(5): 1372–1382.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
233. Audrey S, Brown L, Campbell R, et al.: **Young people's views about consenting to data linkage: Findings from the PEARL qualitative study.** *BMC Med Res Methodol.* 2016; **16**(1): 34.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
234. Boronow KE, Perovich LJ, Sweeney L, et al.: **Privacy risks of sharing data from environmental health studies.** *Environ Health Perspect.* 2020; **128**(1): 17008.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
235. Browman GP, Vollmann J, Virani A, et al.: **Improving the quality of 'personalized medicine' research and practice: Through an ethical lens.** *Per Med.* 2014; **11**(4): 413–423.  
[PubMed Abstract](#) | [Publisher Full Text](#)
236. Kaye J, Heeney C, Hawkins N, et al.: **Data sharing in genomics—re-shaping scientific practice.** *Nat Rev Genet.* 2009; **10**(5): 331–335.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
237. O'Brien SJ: **Stewardship of human biospecimens, DNA, genotype, and clinical data in the GWAS era.** *Annu Rev Genomics Hum Genet.* 2009; **10**: 193–209.  
[PubMed Abstract](#) | [Publisher Full Text](#)
238. McGuire AL, Basford M, Dressler LG, et al.: **Ethical and practical challenges of sharing data from genome-wide association studies: The eMERGE Consortium experience.** *Genome Res.* 2011; **21**(7): 1001–1007.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
239. Healthit.gov: **What is a personal health record.** 2013; 1–2.
240. Shabani M, Knoppers BM, Borry P: **From the principles of genomic data sharing to the practices of data access committees.** *EMBO Mol Med.* 2015; **7**(5): 507–509.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
241. Knoppers BM: **Framework for responsible sharing of genomic and health-related data.** *Hugo J.* 2014; **8**(1): 3.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
242. Kaye J, Terry SF, Juengst E, et al.: **Including all voices in international data-sharing governance.** *Hum Genomics.* 2018; **12**(1): 13.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
243. Welch EW, Shin E, Long J: **Potential effects of the Nagoya Protocol on the exchange of non-plant genetic resources for scientific research: Actors, paths, and consequences.** *Ecol Econ.* 2013; **86**: 136–147.  
[Publisher Full Text](#)
244. Baker DB, Kaye J, Terry SF: **Privacy, Fairness, and Respect for Individuals.** *EGEMS (Washington, DC).* 2016; **4**(2): 7.  
[Publisher Full Text](#)
245. Dove ES, Knoppers BM, Zawati MH: **An ethics safe harbor for international genomics research?** *Genome Med.* 2013; **5**(11): 99.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
246. Antman EM, Benjamin EJ, Harrington RA, et al.: **Acquisition, Analysis, and Sharing of Data in 2015 and Beyond: A Survey of the Landscape.** *J Am Heart Assoc.* 2015; **4**(11): e002810.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
247. Sugano S: **International code of conduct for genomic and health-related data sharing.** *Hugo J.* 2014; **8**(1): 1.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
248. Rodriguez H, Snyder M, Uhlén M, et al.: **Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: The Amsterdam Principles.** *J Proteome Res.* 2009; **8**(7): 3689–3692.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
249. Pasquetto IV, Randles BM, Borgman CL: **On the Reuse of Scientific Data.** *Data Sci J.* 2017; **16**: 8.  
[Publisher Full Text](#)
250. Dillo I, De Leeuw L: **Ten Years Back, Five Years Forward: The Data Seal of Approval.** *Int J Digit Curation.* 2015; **10**(1): 230–239.  
[Publisher Full Text](#)
251. Dillo I, de Leeuw L: **Data Seal of Approval: Certification for sustainable and trusted data repositories.** 2014; 20.  
[Reference Source](#)
252. Al-Tabba A, Al-Omari A, Al-Hussaini M: **Appraisal of the Jordanian Law for Data Sharing in Stem Cell Research: In the Light of the «gA4GH Framework» for Innovative Cancer Care.** *Proceedings - 2018 1st International Conference on Cancer Care Informatics, CCI 2018.* 2019; 232–235.  
[Publisher Full Text](#)
253. Jones KH, Ford DV: **Population data science: advancing the safe use of population data for public benefit.** *Epidemiol Health.* 2018; **40**: e2018061.  
[PubMed Abstract](#) | [Free Full Text](#)
254. Lea NC, Nicholls J, Dobbs C, et al.: **Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research.** *JMIR Med Inform.* 2016; **4**(2): e22.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
255. Castillon G, Castilloux AM, Moride Y: **Development of Standards for Online Repositories.** Wellcome Trust, London, 2017.  
[Publisher Full Text](#)
256. Consultative Committee For Space Data Systems: **Reference Model for an Open Archival Information System.** *J Arch Organ.* 1997; **7**: 48.
257. Giaretta D: **Introduction to OAIS Concepts and Terminology.** In *Advanced Digital Preservation.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011; 13–30.  
[Publisher Full Text](#)
258. Winter A, Stäubert S, Ammon D, et al.: **Smart Medical Information Technology for Healthcare (SMITH).** *Methods Inf Med.* 2018; **57**(S 01): e92–e105.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
259. Banzi R, Bertele V, Demotes-Mainard J, et al.: **Fostering EMA's transparency policy.** *Eur J Intern Med.* 2014; **25**(8): 681–684.  
[PubMed Abstract](#) | [Publisher Full Text](#)
260. Tucker K, Branson J, Dilleen M, et al.: **Protecting patient privacy when sharing patient-level data from clinical trials.** *BMC Med Res Methodol.* 2016; **16** Suppl 1(Suppl 1): 77.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
261. Sweeney L: **k-anonymity: A model for protecting privacy.** *Int J Uncertainty Fuzziness Knowledge-Based Syst.* 2002; **10**(5): 557–570.  
[Publisher Full Text](#)
262. Samarati P: **Protecting respondents identities in microdata release.** *IEEE Trans Knowl Data Eng.* 2001; **13**(6): 1010–1027.  
[Publisher Full Text](#)
263. Bayardo RJ, Agrawal R: **Data Privacy through Optimal k-Anonymization.** In *21st International Conference on Data Engineering (ICDE' 05).* 2005; 217–228.  
[Publisher Full Text](#)
264. El Emam K, Dankar FK: **Protecting Privacy Using k-Anonymity.** *J Am Med Informatics Assoc.* 2008; **15**(5): 627–637.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
265. Kushida CA, Nichols DA, Jadrnicek R, et al.: **Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies.** *Med Care.* 2012; **50** Suppl(Suppl): S82–101.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
266. Wang R, Li YF, Wang X, et al.: **Learning your identity and disease from research papers: information leaks in genome wide association study.** 2009; 534–544.  
[Publisher Full Text](#)
267. Office for Civil Rights, HHS: **Standards for privacy of individually identifiable health information. Final rule.** *Fed Regist.* 2002; **67**(157): 53181–53273.  
[PubMed Abstract](#)

268. TransCelerate Biopharma Inc: **About TransCelerate Biopharma Inc.** 2017. [Reference Source](#)
269. Pharmaceutical Users Software Exchange: **The Global Healthcare Data Science Community.** (accessed Jan. 25, 2021). [Reference Source](#)
270. Bacon E, Budney G, Bondy J, *et al.*: **Developing a Regional Distributed Data Network for Surveillance of Chronic Health Conditions: The Colorado Health Observation Regional Data Service.** *J Public Health Manag Pract.* 2019; **25**(5): 498–507. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
271. CHORDS: **The Colorado Health Observation Regional Data Service (CHORDS).** (accessed Dec. 26, 2020). [Reference Source](#)
272. Davidson A: **Australian Privacy Principles.** In *Social Media and Electronic Commerce Law.* Cambridge University Press. 2018; 418–435.
273. Federal Register of Legislation: **Privacy Amendment (Notifiable Data Breaches) Act 2017.** *Privacy Act 1988.* 2017; (12): 1–26. [Reference Source](#)
274. Wachter RM, Cassel CK: **Sharing Health Care Data with Digital Giants: Overcoming Obstacles and Reaping Benefits while Protecting Patients.** *JAMA.* 2020; **323**(6): 507–508. [PubMed Abstract](#) | [Publisher Full Text](#)
275. Copeland R: **Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans.** *Wall Str J.* 2019. [Reference Source](#)
276. Schneble CO, Elger BS, Shaw DM: **Google's Project Nightingale highlights the necessity of data science ethics review.** *EMBO Mol Med.* 2020; **12**(3): e12053. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
277. Yuhanna N: **Your Enterprise Database Security Strategy 2010.** *Forrester Res.* 2009. [Reference Source](#)
278. Metri P, Sarote G: **Privacy Issues and Challenges in Cloud computing.** *J Adv Eng Sci.* 2011; **5**: 1–6.
279. Zissis D, Lekkas D: **Addressing cloud computing security issues.** *Futur Gener Comput Syst.* 2012; **28**(3): 583–592. [Publisher Full Text](#)
280. Lounis A, Hadjidi A, Bouabdallah A, *et al.*: **Healing on the cloud: Secure cloud architecture for medical wireless sensor networks.** *Futur Gener Comput Syst.* 2016; **55**: 266–277. [Publisher Full Text](#)
281. Sweeney L, Yoo JS, Perovich L, *et al.*: **Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study.** *Technol Sci.* 2017; **2017**: 2017082801. [PubMed Abstract](#) | [Free Full Text](#)
282. Ross C: **Google, Mayo Clinic strike sweeping partnership on patient data.** 2019. (accessed Jan. 04, 2020). [Reference Source](#)
283. OECD: **Recommendation of the Council on Health Data Governance.** OECD/LEGAL/0433." 2019. [Reference Source](#)
284. De Montjoye YA, Radaelli L, Singh VK, *et al.*: **Identity and privacy. Unique in the shopping mall: On the reidentifiability of credit card metadata.** *Science.* 2015; **347**(6221): 536–539. [PubMed Abstract](#) | [Publisher Full Text](#)
285. Jayaraman I, Stanislaus Panneerselvam A: **A novel privacy preserving digital forensic readiness provable data possession technique for health care data in cloud.** *J Ambient Intell Humaniz Comput.* 2020; **12**: 4911–4924. [Publisher Full Text](#)
286. Fan L, Buchanan W, Thummler C, *et al.*: **DACAR platform for eHealth services cloud.** In *Proceedings - 2011 IEEE 4th International Conference on Cloud Computing, CLOUD 2011.* 2011; 219–226. [Publisher Full Text](#)
287. Yang SQY, Y X, X L: **US Patent Application no. 14/143,552.** 2014.
288. Wu R, Ahn GJ, Hu H: **Secure sharing of electronic health records in clouds.** *Collab. 2012 - Proc 8th Int Conf Collab Comput Networking Appl Work.* 2012; 711–718. [Publisher Full Text](#)
289. Mashima D, Ahamad M: **Enhancing accountability of Electronic Health Record usage via patient-centric monitoring.** *IHI' 12 - Proc 2nd ACM SIGHIT Int Heal Informatics Symp.* 2012; 409–418. [Publisher Full Text](#)
290. Wang H, Wu Q, Qin B, *et al.*: **FRR: Fair remote retrieval of outsourced private medical records in electronic health networks.** *J Biomed Inform.* 2014; **50**: 226–233. [PubMed Abstract](#) | [Publisher Full Text](#)
291. Yang JJ, Li JQ, Niu Y: **A hybrid solution for privacy preserving medical data sharing in the cloud environment.** *Futur Gener Comput Syst.* 2015; **43–44**: 74–86. [Publisher Full Text](#)
292. Kaletsch A, Sunyaev A: **Privacy engineering: Personal health records in cloud computing environments.** In *International Conference on Information Systems 2011, ICIS 2011.* 2011; 2213–2223. [Reference Source](#)
293. Akinyele JA, Lehmann CU, Green MD, *et al.*: **Self-Protecting Electronic Medical Records Using Attribute-Based Encryption.** *ePrint IACR org.* 2010; 1–20. [Reference Source](#)
294. Bennati S, Pournaras E: **Privacy-enhancing aggregation of Internet of Things data via sensors grouping.** *Sustain Cities Soc.* 2018; **39**: 387–400. [Publisher Full Text](#)
295. Chan M, Kazatchkine M, Lob-Levy J, *et al.*: **Meeting the demand for results and accountability: A call for action on health data from eight global health agencies.** *PLoS Med.* 2010; **7**(1): 1. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
296. World Health Organization-WHO: **Framework and standards for country health information systems.** World Health. 2008; 63. [Reference Source](#)
297. Murray CJ: **Towards good practice for health statistics: lessons from the Millennium Development Goal health indicators.** *Lancet.* 2007; **369**(9564): 862–873. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
298. Boerma JT, Stansfield SK: **Health statistics now: are we making the right investments?** *Lancet.* 2007; **369**(9563): 779–786. [PubMed Abstract](#) | [Publisher Full Text](#)
299. Ravid R: **Standard operating procedures, ethical and legal regulations in BTB (Brain/Tissue/Bio) banking: What is still missing?** *Cell Tissue Bank.* 2008; **9**(3): 151–167. [PubMed Abstract](#) | [Publisher Full Text](#)
300. Ballantyne A, Stewart C: **Big Data and Public-Private Partnerships in Healthcare and Research: The Application of an Ethics Framework for Big Data in Health and Research.** *Asian Bioeth Rev.* 2019; **11**(3): 315–326. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
301. Aase K, Guise V, Billett S, *et al.*: **Resilience in Healthcare (RiH): A longitudinal research programme protocol.** *BMJ Open.* 2020; **10**(10): e038779. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
302. Wiig S, Aase K, Billett S, *et al.*: **Defining the boundaries and operational concepts of resilience in the resilience in healthcare research program.** *BMC Health Serv Res.* 2020; **20**(1): 330. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
303. Berkowsky RW, Czaja SJ: **Challenges associated with online health information seeking among older adults.** *Aging, Technology and Health.* Elsevier, 2018; 31–48. [Publisher Full Text](#)
304. Righi AW, Saurin TA, Wachs P: **A systematic literature review of resilience engineering: Research areas and a research agenda proposal.** *Reliab Eng Syst Saf.* 2015; **141**: 142–152. [Publisher Full Text](#)
305. Ćwiklicki M, Klich J, Chen J: **The adaptiveness of the healthcare system to the fourth industrial revolution: A preliminary analysis.** *Futures.* 2020; **122**: 102602. [Publisher Full Text](#)
306. Gupta J, Termeer C, Klostermann J, *et al.*: **The Adaptive Capacity Wheel: a method to assess the inherent characteristics of institutions to enable the adaptive capacity of society.** *Environ Sci Policy.* 2010; **13**(6): 459–471. [Publisher Full Text](#)
307. Grambsch A, Menne B: **Adaptation and adaptive capacity in the public health context.** *Clim Chang Hum Heal - Risks responses.* 2003; 220–236. [Reference Source](#)
308. Lemos MC, Agrawal A, Johns O, *et al.*: **Building adaptive capacity to climate change in less developed countries.** *Clim Sci Serv Soc Res Model Predict priorities.* 2011.
309. Townend D: **Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream?** *Hum Genet.* 2018; **137**(8): 657–664. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
310. Biobank Lexicon: **Home | Public Population Project in Genomics and Society.** 2013. [Reference Source](#)
311. Knoppers BM, Chisholm RL, Kaye J, *et al.*: **A P3G generic access agreement for population genomic studies.** *Nat Biotechnol.* 2013; **31**(5): 384–385. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
312. Shah SM, Khan RA: **Secondary use of electronic health record: Opportunities and challenges.** *IEEE Access.* 2020; **8**: 136947–136965. [Publisher Full Text](#)
313. Chen H, Pang T: **A call for global governance of biobanks.** *Bull World Health Organ.* 2015; **93**(2): 113–7. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
314. Rudan I, Marušić A, Campbell H: **Developing biobanks in developing countries.** *J Glob Health.* 2011; **1**(1): 2–4. [PubMed Abstract](#) | [Free Full Text](#)
315. Carter P, Laurie GT, Dixon-Woods M: **The social licence for research: Why care.data ran into trouble.** *J Med Ethics.* 2015; **41**(5): 404–409. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
316. Dhai A, Mahomed S: **Biobank research: Time for discussion and debate.** *S Afr*

- Med J.* 2013; **103**(4): 225–227.  
[PubMed Abstract](#) | [Publisher Full Text](#)
317. Manson NC, O'Neill O: **Rethinking informed consent in bioethics.** Cambridge University Press, 2007.  
[Publisher Full Text](#)
318. Knoppers BM, Chadwick R: **Human Genetic Research: Emerging Trends in Ethics.** *Nat Rev Genet.* 2005; **6**(1): 75–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
319. Elger BS, Caplan AL: **Consent and anonymization in research involving biobanks: Differing terms and norms present serious barriers to an international framework.** *EMBO Rep.* 2006; **7**(7): 661–666.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
320. Sookhak M, Gani A, Talebian H, et al.: **Remote data auditing in cloud computing environments: A survey, taxonomy, and open issues.** *ACM Comput Surv.* 2015; **47**(4): 1–34.  
[Publisher Full Text](#)
321. Kuo AMH: **Opportunities and Challenges of Cloud Computing to Improve Health Care Services.** *J Med Internet Res.* 2011; **13**(3): e67.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
322. MCDERMOTT AM, HAMEL LM, STEEL D, et al.: **HYBRID HEALTHCARE GOVERNANCE FOR IMPROVEMENT? COMBINING TOP-DOWN AND BOTTOM-UP APPROACHES TO PUBLIC SECTOR REGULATION.** *Public Adm.* 2015; **93**(2): 324–344.  
[Publisher Full Text](#)
323. Milne R, Morley KI, Howard H, et al.: **Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia.** *Hum Genet.* 2019; **138**(11–12): 1237–1246.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
324. Hyder AA, Selig H, Ali J, et al.: **Integrating capacity development during digital health research: a case study from global health.** *Glob Health Action.* 2019; **12**(1): 1559268.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
325. Siddiqi S, Masud TI, Nishtar S, et al.: **Framework for assessing governance of the health system in developing countries: Gateway to good governance.** *Health Policy.* 2009; **90**(1): 13–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
326. Schaefer GO, Tai ES, Sun S: **Precision Medicine and Big Data: The Application of an Ethics Framework for Big Data in Health and Research.** *Asian Bioeth Rev.* 2019; **11**(3): 275–288.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
327. Soranno PA, Bissell EG, Cheruvellil KS, et al.: **Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse.** *Gigascience.* 2015; **4**(1): 28.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
328. Lawlor RT: **Biobanks in low resource contexts.** In *Biobanking of Human Biospecimens: Principles and Practice*, ARC-Net Applied Research on Cancer Centre, University of Verona, Verona, Italy: Springer International Publishing, 2017; 169–198.  
[Publisher Full Text](#)
329. Sgaier SK, Jha P, Mony P, et al.: **Public health. Biobanks in developing countries: needs and feasibility.** *Science.* 2007; **318**(5853): 1074–1075.  
[PubMed Abstract](#) | [Publisher Full Text](#)
330. Isasi RM: **Policy Interoperability in Stem Cell Research: Demystifying Harmonization.** *Stem Cell Rev Rep.* 2009; **5**(2): 108–115.  
[PubMed Abstract](#) | [Publisher Full Text](#)
331. Ancke JS, Silver M, Miller MC, et al.: **Consumer experience with and attitudes toward health information technology: a nationwide survey.** *J Am Med Inform Assoc.* 2013; **20**(1): 152–156.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
332. Bhat S, Hegde TT: **The costs of institutional review boards.** *N Engl J Med.* 2005; **353**(3): 315–7; author reply 315–7.  
[PubMed Abstract](#)
333. Australian Law Reform Commission: **Essentially yours: the protection of human genetic information in Australia.** 2003.  
[Reference Source](#)
334. Wellcome Trust: **Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility Report of a meeting organized by the Wellcome Trust and held on 14-15 January 2003 at Fort Lauderdale, USA.** 2003.  
[Reference Source](#)
335. Fortier I, Burton PR, Robson PJ, et al.: **Quality, quantity and harmony: the DataShAPER approach to integrating data across bioclinical studies.** *Int J Epidemiol.* 2010; **39**(5): 1383–1393.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
336. Saylor BP: **Beyond informed consent.** *Med Econ.* 2010; **87**(4): 25, 30, 32.  
[PubMed Abstract](#)
337. Califf RM, Robb MA, Bindman AB, et al.: **Transforming Evidence Generation to Support Health and Health Care Decisions.** *N Engl J Med.* 2016; **375**(24): 2395–2400.  
[PubMed Abstract](#) | [Publisher Full Text](#)
338. Bradwell P, Gallagher N: **FYI: The New Politics of Personal Information.** 2007; 1–79.  
[Reference Source](#)
339. Kaufman DJ, Murphy-Bollinger J, Scott J, et al.: **Public Opinion about the Importance of Privacy in Biobank Research.** *Am J Hum Genet.* 2009; **85**(5): 643–654.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
340. Platt J, Bollinger J, Dvoskin R: **Public preferences regarding informed consent models for participation in population-based genomic research.** *Genet Med.* 2014; **16**(1): 11–18.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
341. National Science Board and National Science Foundation: **Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century.** *Natl Sci Board.* 2005; **87**.  
[Reference Source](#)
342. Human Genome Organisation (HUGO), Ethics Committee: **Statement on human genomic databases, December 2002.** *J Int Bioethique.* 2003; **14**(3–4): 207–10.  
[PubMed Abstract](#)
343. Abbing HDC, UNESCO: **International Declaration on Human Genetic Data.** *Eur J Health Law.* 2004; **11**(1): 93–107.  
[PubMed Abstract](#) | [Publisher Full Text](#)
344. Adhikari CP, Pell C, Cheah PY: **Community engagement and ethical global health research.** *Glob Bioeth.* 2020; **31**(1): 1–12.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
345. Hamed S, Klingberg S, Mahmud AJ, et al.: **Researching health in diverse neighbourhoods: critical reflection on the use of a community research model in Uppsala, Sweden.** *BMC Res Notes.* 2018; **11**(1): 612.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
346. O'Donnell HC, Patel V, Kern LM, et al.: **Healthcare consumers' attitudes towards physician and personal use of health information exchange.** *J Gen Intern Med.* 2011; **26**(9): 1019–1026.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
347. Murtagh MJ, Minion JT, Turner A, et al.: **The ECO OUTER methodology for stakeholder engagement in translational research.** *BMC Med Ethics.* 2017; **18**(1): 24.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
348. Wilson RC, Butters OW, Clark T, et al.: **Digital methodology to implement the ECO OUTER engagement process [version 2; peer review: 2 approved].** *F1000Res.* 2016; **5**: 1307.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
349. Wickramasinghe NS, Fadlalla AMA, Geisler E, et al.: **A framework for assessing e-health preparedness.** *Int J Electron Healthc.* 2005; **1**(3): 316–34.  
[PubMed Abstract](#) | [Publisher Full Text](#)
350. Srivastava SK: **Adoption of Electronic Health Records: A roadmap for India.** *Healthc Inform Res.* 2016; **22**(4): 261–269.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
351. Denk F: **Don't let useful data go to waste.** *Nature.* 2017; **543**(7643): 7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
352. Igumbor: **Public and Population Health data sharing in Africa - views of academics and researchers (Version 1).** *Zenodo.* 2021.  
<http://www.doi.org/10.5281/zenodo.5155880>

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 14 February 2022

<https://doi.org/10.21956/wellcomeopenres.18772.r47220>

© 2022 Useh U. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Ushotanefe Useh** 

Lifestyle Diseases Research Entity, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa

This review sought to provide practical recommendations and requirements to support the development of a multi-consortia public and population health data sharing framework for Africa. This research seeks to inform a platform that will harnesses available resources, provide incentive, data sharing, and optimize the progress made by different research groups in Africa.

This work was a narrative of over 655 documents, publications and policies on data sharing in public and population health. The authors also reviewed data protection laws and regulations as they affect different countries in the continent of Africa. They also provided recommendation in Table 1. Among those cited were those from South Africa, Kenya, Nigeria, Togo, Tanzania, Uganda and Zimbabwe.

The study design is appropriate with both narrative review and interview of 35 key informants from African-led research capacity building programmes. Detailed procedure is provided in the study. A detailed Table and with specific consideration is provided.

More than 9 recommendations were provided by the authors. These ranged from developing an integrated data bank, infrastructure development to capacity building initiatives.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Sociology of Health and Illness and Lifestyle Diseases

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 07 September 2021

<https://doi.org/10.21956/wellcomeopenres.18772.r45702>

© 2021 Eboreime E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Ejema Eboreime** 

Department of Psychiatry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

This is very interesting and important article that attempts to identify the requirements and provide practical recommendations for developing a multi-consortia public and population health data-sharing framework for Africa. I believe this situates and compliments the ongoing debates about registration and repositories for global health research.

I have a few minor comments for the authors' consideration.

**Methods:**

It is clear that the authors conducted a multi-methods study (a narrative review and key informant interviews). However, this is not stated clearly in either method sections of the abstract and the main text. My initial thought as I read the article was that this was a narrative review, but as I progressed with reading, I found that interviews were also conducted. I believe stating clearly that two approaches were used, before describing both approaches separately would give readers an early understanding of the research approach used. Further to this, a line or two on how triangulation was conducted would be valuable.

**Discussion:**

I think a line or two that incorporates this discussion into the interesting (similar) debates around

global health research registration and repositories would be an awesome idea. There have very recently been questions as to the necessity or absurdity of this approach in global health initiated by an editorial of mine in the *BMJ Global Health*<sup>1</sup>. This editorial sparked a commentary that highlighted challenges with centralized repositories for community-oriented research which make it an absurdity<sup>2</sup>. It would be interesting for the author to throw some thoughts on how their findings are situated in these debates.

### References

1. Eboime EA, Abimbola S: A priori registration of global health research-necessity or absurdity?. *BMJ Glob Health*. 6 (5). [PubMed Abstract](#) | [Publisher Full Text](#)
2. Mathias K, Nagesh S, Varghese S, Qadeer I, et al.: The absurdity of research registration for community-oriented knowledge coproduction. *BMJ Glob Health*. 2021; 6 (8). [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Health systems, Global health research, Implementation science

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Comments on this article

Version 1

Reader Comment 01 Feb 2022



**Kobus Herbst**, DSI-MRC South Africa Population Research Infrastructure Network, South Africa

This is a very comprehensive review of the literature pertaining to population health databanks from an African perspective. The authors have done a very thorough job. The table summarising their extensive review is very helpful to guide the reader to the findings and recommendations on a particular topic of interest.

This work showcase the contention of the authors that African researchers are not mere data collectors but active research leaders capable of enhancing scientific growth in Africa.

I share the opinion of the authors that the establishment of integrated databanks in Africa are due and possible.

**Competing Interests:** No competing interests were disclosed.

---