REVIEW

The evolving role of research ethics committees in the era of open data

S Mahomed,[1,2] BCom, LLB, LLM, PhD; **M Labuschaigne,**[1] BA Hons, MA, D Litt, LLB, LLD

[1] *Department of Jurisprudence, Unisa, Pretoria, South Africa*
[2] *Steve Biko Centre for Bioethics, University of the Witwatersrand, Johannesburg, South Africa*

Corresponding author: *S Mahomed (mahoms1@unisa.ac.za)*

While open science gains prominence in South Africa with the encouragement of open data sharing for research purposes, there are stricter laws and regulations around privacy – and specifically the use, management and transfer of personal information – to consider. The Protection of Personal Information Act No. 4 of 2013 (POPIA), which came into effect in 2021, established stringent requirements for the processing of personal information and has changed the regulatory landscape for the transfer of personal information across South African borders. At the same time, draft national policies on open science encourage wide accessibility to data and open data sharing in line with international best practice. As a result, the operation of research ethics committees (RECs) in South Africa is affected by the conflicting demands of the shift towards open science on the one hand, and the stricter laws protecting participants' personal information and the transfer thereof, on the other. This article explores the continuing evolving role of RECs in the era of open data and recommends the development of a data transfer agreement (DTA) for the ethical management of personal health information, considering the challenges that RECs encounter, which centres predominantly on privacy, data sharing and access concerns following advances in genetic and genomic research and biobanking.

S Afr J Bioethics Law 2022;15(3):80-83. https://doi.org/10.7196/SAJBL.2022.v15i3.822

The global drive towards inter-connectedness, supported by the notions of open science, open access and open data, will have a significant impact on data sharing for research purposes. Data sharing is credited with accelerating scientific breakthroughs and facilitating the progress of research.[1] In addition to extending the scope of scientific potential, societal benefits from open access to data are also considerable, depending on the types and scale of data made available to the public, including the improvement of social welfare as society gains from accessible information and refining public perceptions of transparency.[2]

To align South Africa (SA) with this growing trend towards open science, the Draft National Open Science policy which encourages open science, open data and open access, was approved for stakeholder consultation in the first quarter of 2022.[3] Similarly, in 2021, the Draft National Data and Cloud policy[4] was published with a vision of transforming SA into a data-driven digital economy. Although both policies encourage open data sharing in line with international best practice, the same period (2021) saw the enactment of SA's Protection of Personal Information Act No. 4 of 2013 (POPIA), which establishes minimum requirements for the processing of personal information, including the grounds for lawful processing. However, POPIA's enactment was met with concerns regarding its interpretation for research (specifically health research) purposes,[5-7] as well as its application to already established research practices in SA.

The operation of research ethics committees (RECs) in SA is affected by the conflicting demands of the shift towards open science on the one hand, and the stricter laws around protecting participants' personal information and the transfer thereof, on the other. POPIA requires an added principles-based assessment when personal information is shared between institutions, both locally and across borders. A recent paper which explored the topic focused on RECs functioning from a global and national perspective while outlining the critical role that they play in reviewing health research proposals when human biological materials are transferred between institutions, as well as RECs as a party to the national SA material transfer agreement (MTA) template.[8] The present article aims to take these discussions one step further and provides an overview of the continuing evolving role of RECs considering SA's new privacy framework. It offers an outline of POPIA, focusing on the legal requirements when personal information/data are shared between local and international institutions. It also considers health data breaches and the challenges that RECs currently face in the context of reviewing protocols involving data sharing and big data. The paper concludes with assessing the value of DTAs in regulating the transfer of data, which would in turn strengthen the operation of RECs regarding legal and ethical compliance in this context.

Managing data transfers under POPIA

The underlying tension between achieving open data vis-à-vis the strict privacy protections regarding the processing of personal information in POPIA should be addressed to achieve the right balance that will not hamper the progress of research. RECs are increasingly required to reconcile the demands of open science with legal privacy protections and to comply with ethical and legal norms when data are transferred. Whereas the protection of personal

information under POPIA appears to be suffused with individual autonomy and self-determination of the data subject, open science strives to serve the greater collective.

POPIA regulates how personal information is collected, used, stored, shared and generally processed from the moment of collection to destruction. Chapter 3, Part A, of POPIA provides for eight conditions that need to be satisfied when personal information is processed. A research institution or researcher is responsible to ensure that personal information is lawfully processed in accordance with these conditions of POPIA and that participants' constitutional rights to privacy are not infringed. National transfers of personal information between institutions require compliance with POPIA, as well as REC approval and participant consent.[9] International transfers of personal information are regulated by section 72 of POPIA and may occur under five circumstances, of which the following three are relevant when international transfer for research purposes takes place:

- when the recipient in the foreign country is subject to a law, binding corporate rules or binding agreement that provides for an adequate level of protection that upholds principles that are substantially similar for the processing of personal information (S72(1)(a));
- when the participant consents to the transfer (S72(1)(b)); or
- when the transfer is for the benefit of the participant and where consent is not reasonably practicable to obtain, recognising that if consent were possible, the research participant would likely provide it (S72(1)(e)).

Where consent of the data subject (research participant) is relied upon as a ground for international transfer of personal information for research purposes, this will only be possible where the participant is provided with details of the third party with whom the personal information will be shared, the risks associated with that sharing and the opportunity to withdraw consent at any time (section 11(2)(b)).

However, as withdrawal may not always be possible after the personal information is shared outside SA and details of the third party or subsequent risks associated with the sharing may not always be known when the initial consent is obtained, the practical application of this ground is questionable.

Alternatively, section 72(1)(e) indicates that international transfers of personal information may take place where the transfers are for the benefit of each individual participant, which implies that a decision to this effect would need to be made by each individual participant. This ground would almost certainly be impossible as a basis for transfers when large data sets are shared outside SA.

It therefore appears that international transfers are most likely to occur on the grounds of section 72(1)(a), i.e. when the recipient in the foreign country is subject to a law, binding corporate rules or binding agreement that provides for an adequate level of protection that upholds principles that are substantially similar for the processing of personal information. However, as the Information Regulator has not yet provided guidance on which countries have similar levels of privacy protections to SA, or criteria to consider when making such an assessment, a binding contractual agreement, e.g. a data transfer agreement (DTA), seems to be the most practical solution to safeguard personal information that is shared across borders.[9] Therefore, while national transfers of data must take place by complying with the processing requirements as set out in POPIA, REC approval and in accordance with participant consent, international transfers of data outside SA can occur where there is a binding DTA in place. Among setting out the conditions and purposes for which the data will be shared, a DTA should also set out the safety mechanisms in place to protect participant privacy after transfer. This point then prompts the next issue for discussion, namely the risk of informational harms to participants and providing country researchers and institutions.

Recent health data breaches

The emphasis on protecting personal information has highlighted risks which now transcend from physical and psychological to informational.[10] Informational harms and healthcare data breaches have seen rapid growth, particularly during the COVID-19 pandemic, where the increase in patient visits to hospital amplified their risk of exposure to security threats.[11]

POPIA provides in section 22 that a data breach occurs when there are reasonable grounds to suspect that a data subject's personal information has been accessed or acquired by any unauthorised person or entity, requiring the mandatory reporting of the breach by the responsible party to the Information Regulator as soon as reasonably possible.

Data breaches have far-reaching effects for affected individuals and companies, ranging from the loss of control over their personal data, limitation of their rights, discrimination, identity theft or fraud, financial loss, unauthorised reversal of pseudonymisation, and damage to reputation, to loss of confidentiality of personal data protected by professional secrecy.[12] In addition to the significant economic or social disadvantage (public embarrassment; stigmatisation) to individuals, the reputational risks, as well as huge financial impact on affected companies, are severe. IBM Security, who examined the financial impact of data breaches globally, reported that these incidents, cost SA in 2022 (up to July 2022) around USD3.6 million in total, or ZAR 55 885 200 million.[13]

Some examples of local health data breaches during the time of the pandemic include those reported by Life Healthcare Group, the second-largest private hospital operator in S A during June 2020;[14] and by Experian, a business and credit information services agency, involving to date SA's biggest data breach ever, exposing the personal information of approximately 24 million South Africans and 793 749 business entities.[15] An analysis of data breaches between 2015 and 2019, recorded by the Privacy Rights Clearinghouse, shows that 76.59% of all recorded data breaches occurred in the healthcare sector, constituting three times as many breaches as those recorded in the education, finance, retail and government sectors combined.[16]

Personal health information is more valuable on the black market than financial data, because hacked health information has a longer shelf-life after the breach, specifically for identity or financial theft. Healthcare establishments usually have large databases, making them attractive targets for hackers.[16]

Current challenges that RECs face

The role of RECs in the transfer of personal information goes beyond legal compliance and should consider individual participant, community and social concerns which may not be explicitly outlined in legal provisions.[17] Also of concern to RECs is the ethical management of research data, including big data (which is according to Ferretti *et al.*[18] 'any research relying on large datasets, made of data heterogeneous in source, processed at high speed, and analysed through novel

computational techniques') or large volumes of data,[19] mainly because traditional ethics oversight procedures and practices, specifically in the field of biomedical and health research,[20] may no longer be adequate.[21]

Artificial intelligence (AI) and computational methods to analyse and harness data have played a more prominent role during COVID-19 in the improvement of individual and public health.[22] These benefits notwithstanding, AI challenges traditional research principles, such as data privacy, informed consent, scientific validity of research, risk assessment and the distribution of benefits,[23] not to mention the epistemic demands that it poses regarding the assessment of scientific validity, technological reliability, accountability, fairness and transparency.[24] AI also removes the need for the participation of research participants in research, as retrospective data processing does not require the participation of research participants.[25]

Ferretti *et al*.[18] identify several ethical challenges for RECs regarding big data research oversight, many of which also apply to the review of protocols involving data transfer and sharing. Firstly, the lack of specific normative standards for the ethics review of big data studies and data sharing; epistemic challenges experienced by RECs, particularly their inadequate experience and expertise regarding the use and transfer of data; normative challenges relating to the scope of ethical reflection due to the inadequacy of traditional tools used to assess biomedical research, including the absence of clear criteria for evaluation of big data studies; and finally, the need for reform, such as capacity building and data literacy for REC members, as well as complementing RECs with data-focused oversight bodies.

Kaye[26] rightly speculates on the extent of the impact of global data sharing on the social contract underpinning research participation and related ethical governance mechanisms, not to mention its effect on informed consent and the right of withdrawal of participants. For example, methods utilised for de-identification and aggregation for the protection of privacy may in fact make it difficult to trace and remove individually derived data.[26] In the context of biobanking, absolute withdrawal, for the public good, may lead to an unnecessary loss of samples. Moreover, in the case of human genetic research involving more advanced sequencing technology, de-identification is problematical because of the uniquely identifiable nature of genetic information, and more so when this is shared with other researchers where reliance is placed on encryption to anonymise or de-identify sequence information.[27]

Most of the challenges that RECs experience appear to be more focused on the protection of privacy of research participants, rather than finding ways to balance the protection of participants with the promotion of research and science, which in turn relies on increased data sharing. Some institutions have introduced data access committees to provide a new tier of oversight[26] and approve the sharing of data for specific projects. Despite the benefits of these committees, they may not be the ideal instrument to monitor data use by secondary or further researchers once information has been shared via managed accessed repositories.[28] From a regulatory perspective, terms of reference or uniform standard operating procedures, including their role vis-à-vis research ethics committees, are still to be clarified.

Data transfer agreements as a guide for RECs

Currently, SA has a national Material Transfer Agreement template[29] that is used when samples and data leave the country for research purposes. RECs as a party to the SA MTA were discussed in our first paper on the topic, which considered their roles when human biological materials (HBMs) are transferred for health research purposes.[8] However, the SA MTA does not adequately deal with data, which is probably because it was published before POPIA came into effect. Following the strict privacy considerations outlined in POPIA, it is crucial that personal information/data are appropriately regulated when transfers outside our borders are contemplated.

In conjunction with guidance from the Code of Conduct for Research being developed by the Academy of Science of South Africa (ASSAf),[30] a DTA which details the terms under which data will be transferred, stored and used, specifying rights and obligations for both the data supplier and the recipient,[31] would be useful in enabling the ethico-legal management of data sharing, while respecting the value of data sharing in the era of open science.

Questions that need to be considered include what provisions should be in place to govern the transfers of data outside of SA. Forecasts predict that by 2025 the global data sphere will grow to 175 zettabytes from 33 zettabytes in 2018.[32] Unsurprisingly, with rapid advances in data analysis techniques and accessibility to data, it is no longer difficult to establish a connection to the person from whom the data originated.[33] We could therefore find ourselves in a situation (in the not too distant future) where most types of data have the possibility of being personal information and therefore subject to privacy regulations. Thus, if non-personal information has the possibility of becoming personal information through analysis techniques, should all transfers of data be regulated, or should we limit regulation to personal information only? The following provisions incorporated into a DTA that sets out the conditions for transfer could assist with safeguarding personal information of participants and be used as a guide by RECs:

- whether the necessary ethics approvals are in place for data to be shared
- that the data are being shared for research purposes only
- details of the type/s and amount of data being transferred
- the category of risk associated with the data – for example, data generated from children, vulnerable groups or special personal information being transferred outside SA for processing by third parties on a large scale would be significantly more risky than personal information that is not categorised as special personal information and which will not be processed on a large scale
- whether the data have been appropriately de-identified, including the methods used to achieve de-identification and whether there is a risk of re-identification
- details of planned sharing arrangements with third parties
- if the data will be shared on open-access platforms
- the security measures in place to safeguard access to the data and prevent unauthorised access to the data, including how security breaches will be mitigated
- who holds intellectual property rights, should the data generate results that are capable of intellectual property protections
- whether there will be any direct benefits to the provider or direct or indirect benefits to participants or the participant community
- whether the research participants' consent is in line with the provisions set out in the DTA.

The development of a national DTA should be a participatory process that would require input from relevant key stakeholders. Driving the process should be the responsibility of the National Department of Health and the Department of Science and Innovation, with input from other government departments, where relevant. At an institutional level, the legal departments and research offices, in collaboration with the institutional REC, should further refine and calibrate the national DTA for institutional use. As the national DTA framework should provide guidance, it would be left to institutions to tailor the framework to fit relevant institutional needs. Once implemented, the monitoring of compliance with the DTA will be the responsibility of the RECs, in conjunction with the research offices or legal departments of institutions.

Conclusion

The current regulatory framework on the protection of personal information provides limited guidance on the sharing of personal health information or data. The sharp increase in local and international data breach incidents points to the urgent need to strengthen the legal framework for data sharing and transfer in SA, including providing ethically sound practices, flexible infrastructure, and appropriate governance policies. The challenges that RECs encounter centre predominantly on privacy, data sharing and access concerns following advances in genetic and genomic research and biobanking. This article recommends the development of a DTA for the ethical management, including transfer and sharing of personal health information in SA. Such a DTA should not only recognise the different priorities and values of a range of stakeholders but should be underpinned by participatory and procedurally fair processes that will give effect to equitable sharing that is in the benefit of research participants, communities, researchers and national research institutions involved in the collection and sharing of data.

Declaration. None.
Acknowledgements. None.
Author contributions. Equal contributions.
Funding. None.
Conflicts of interest. None.

1. 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;4. https://doi.org/10.1186/s41073-017-0028-9.
2. UNESCO. UNESCO Recommendation on Open Science. 16 November 2021. https://www.unesco.org/en/natural-sciences/open-science (accessed 1 August 2022).
3. SabinetLaw. National Open Science Policy on the Cards. 10 March 2022. https://legal.sabinet.co.za/articles/national-open-science-policy-on-the-cards/ (accessed 29 July 2022).
4. Government of South Africa. Draft National Policy on Data and Cloud, Government Notice 306. Pretoria: Government Gazette 44389 of 1 April 2021.
5. Thaldar DW, Townsend B. Genomic research and privacy: A response to Staunton et al. S Afr Med J 2020;110(3):172-174. https://doi.org/10.7196/SAMJ.2020.v110i3.14431
6. 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 2013. S Afr Med J 2019;109(7):468. https://doi.org/10.7196/SAMJ.2019.v109i7.14148
7. Thaldar DW, Townsend BA. Exempting health research from the consent provisions of POPIA. PER J 2021;24. https://doi.org/10.17159/1727-3781/2021/v24i0a10420
8. Mahomed S, Labuschaigne M. The role of research ethics committees in South Africa when human biological materials are transferred between institutions S Afr J Bioethics Law 2019;12(2):84-87. https://doi.org/10.7196/SAJBL.2019.v12i2.685
9. Mahomed S, Loots G, Staunton C. The role of data transfer agreements in ethically managing data sharing for research in South Africa. S Afr J Bioethics Law 2022;15(1):26-30. https://doi.org/10.7196/SAJBL.2022.v15i1.807
10. Feretti A, Ienca M, Sheehan M, et al. Ethics review of big data research: What should stay and what should be reformed? BMC Med Ethics 2021;22:51 https://doi.org/10.1186/s12910-021-00616-4
11. Alkinoon M, Choi SJ, Mohaisen D. Measuring Healthcare Data Breaches. In: Kim H, editor. Information Security Applications. WISA 2021. Lecture Notes in Computer Science, vol 13009. Cham: Springer, 2021. https://doi.org/10.1007/978-3-030-89432-0_22
12. Examples regarding personal data breach notification. Guidelines 01/2021 at par 6. https://edpb.europa.eu/system/files/2022-01/edpb_guidelines_012021_pdbnotification_adopted_en.pdf (accessed 2 August 2022).
13. IBM Security. Cost of a data breach report 2022. https://www.ibm.com/downloads/cas/XZNDGZKA (accessed 2 August 2022).
14. Mungadze S. Life Healthcare reveals damage caused by data breach. IT Web. 31 August 2020. https://www.itweb.co.za/content/rW1xLv59YPGvRk6m (accessed 2 August 2022).
15. Muncaster P. Experian Data Breach Hits 24 Million Customers. Info Security. 30 August 2020. https://infosecurity-magazine.com/news/experian-data-breach-24-million/ (accessed 2 August 2022).
16. Healthcare Data Breach Statistics (December 2021). HIPAA Journal. https://www.hipaajournal.com/healthcare-data-breach-statistics/ (accessed 2 August 2022).
17. Shabani M, Chassang G, Marelli L. The Impact of the GDPR on the Governance of Biobank Research. In: Slokenberga S, Tzortzatou O, Reichel J, editors. GDPR and Biobanking. Law, Governance and Technology Series, vol 43. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-49388-2_4
18. Ferretti A, Ienca M, Velarde MR, Hurst S, Vayena E. The challenges of big data for research ethics committees: A qualitative Swiss study. J Empir Res Human Res Ethics 2022;17(1-2):129-143 at 129.
19. Leonelli S. Scientific research and big data. In: Stanford Encyclopedia of Philosophy, Summer 2020 ed. Zalta EN, editor. https://plato.stanford.edu/archives/sum2020/entries/science-big-data/
20. Rennie S, Buchbinder M, Juengst E, Brinkley-Rubinstein L, Blue C, Rosen DL. Scraping the web for public health gains: Ethical considerations from a 'big data' research project on HIV and incarceration. Public Health Ethics 2020;13(1):111-121. https://doi.org/10.1093/phe/phaa006
21. Samuel G, Chubb J, Derrick G. Boundaries between research ethics and ethical research use in artificial intelligence health research. J Empir Res Human Res Ethics 2021;16(3):325-337. https://doi.org/10.1177/ 15562646211002744
22. Blasimme A, Vayena E. The ethics of AI in biomedical research, patient care and public health. In: Dubber MD, Pasquale F, Das S, editors. Oxford Handbook of Ethics of Artificial Intelligence. Oxford: Oxford University Press; 2019 at 718.
23. Price WN, Cohen IG. Privacy in the age of medical big data. Nature Med 2019;25(1):37-43. https://doi.org/10. 1038/s41591-018-0272-7
24. Friesen P, Douglas-Jones R, Marks M, et al. Governing AI-driven health research: Are IRBs up to the task? Ethics Human Res 2021;43(2):35-42. https://doi.org/10.1002/eahr.500085
25. Metcalf J, Crawford K. Where are human subjects in big data research? The emerging ethics divide. Big Data Soc 2016;3(1):1-14. https://doi.org/10.1177/2053951716650211
26. Kaye J. The tension between data sharing and the protection of privacy in genomics research. Ann Rev Genomics Human Gen 2012;13(1):415-431. https://doi.10.1146/annurev-genom-082410-101454
27. Greenbaum D, Sboner A, Mu XJ, Gerstein M. Genomics and privacy: Implications of the new reality of closed data for the field. PLoS Comput Biol 2011;7(12):e1002278. https://doi.org/10.1371/journal.pcbi.1002278.
28. Johnson AD, Leslie R, O'Donnell CJ. Temporal trends in results availability from genome-wide association studies. PLoS Gen 2011;7(9):e1002269. https://doi.org/10.1371/journal.pgen.1002269
29. South African Government. Material Transfer Agreement for Human Biological Materials. Government Notice 719, Government Gazette 41781, 20 July 2018.
30. Mahomed S, Staunton C. Ethico-legal analysis of international sample and data sharing for genomic research during COVID-19: A South African perspective. BioLaw Law, Genetics Genomics: An unfolding relationship 2021;261-276.
31. Mello MM, Triantis G, Stanton R, Blumenkranz E, Studdert DM. Waiting for data: Barriers to executing data use agreements. Science 2020;367(6474):150-152. https://doi.org/10.1126/science.aaz7028.
32. Reinsel D, Gantz J, Rydning J. The digitisation of the world from edge to core. IDC White Paper, Doc#US44413318, 2018. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf#:~:text=IDC%20predicts%20that%20the%20Global%20Datasphere%20will%20grow,of%20that%20capacity%20supplied%20from%20the%20HDD%20industry (accessed 30 July 2022).
33. Finck M, Pallas F. They who must not be identified – distinguishing personal from non-personal data under the GDPR. Int Data Privacy Law 2020;10(1):11-36. https://doi.org/10.1093/idpl/ipz026

Accepted 16 November 2022.