

Enhancing data governance in collaborative research: Introducing SA DTA 1.1

D Thaldar, PhD¹; M Botes, LLD; L Swales, PhD; P Esselaar, LLM

School of Law, University of KwaZulu-Natal, Durban, South Africa

Corresponding author: D Thaldar (ThaldarD@ukzn.ac.za)

Background. The SA DTA was updated to better serve the South African research community by providing clarity on exactly when – at what stage during research – institutions hold rights to the data they generate in collaborative research contexts where raw data are received and integrated with other data. SA DTA 1.1 introduces significant enhancements in data governance, focusing on the explicit definition and management of ‘inferential data’.

Objectives. To introduce SA DTA 1.1 and demonstrate its practical application in real-world research contexts.

Methods. Using a descriptive design, the study presents two practical case studies: one involving two universities and another a university and a pharmaceutical company. The first case focuses on identifying genetic markers for neurodegenerative diseases, while the second addresses genetic markers for chronic diseases.

Results. In the first case study, the entity conducting the analysis generated inferential data independently, thereby gaining sole rights in such data. In the second case study, both collaborating entities contributed to the data analysis, leading to joint rights in the inferential data. These findings demonstrate the flexibility and clarity provided by SA DTA 1.1 in managing data ownership and intellectual property rights.

Conclusion. SA DTA 1.1 significantly advances data governance by providing clear and adaptable guidelines for data rights in collaborative research. It supports ethical and efficient data sharing, protecting researchers’ interests and fostering global scientific innovation.

S Afr J Bioethics Law 2024;17(2):e2300. <https://doi.org/10.7196/SAJBL.2024.v17i2.2300>

In the dynamic realm of global scientific research, the efficient movement and management of data across borders are pivotal for driving medical and scientific advancements.^[1] Acknowledging the need for a structured approach to handle these complexities, we first introduced our data transfer agreement template for the South African (SA) research community (SA DTA) and an accompanying explanatory memorandum in early 2023.^[2] This initiative followed the release of a beta version for public comment in October 2022, which we detailed in an article in this journal, outlining our drafting approach.^[3]

The development of the SA DTA was motivated by a clear need: to empower the SA research community with a legally robust framework that facilitates the safe, ethical and efficient transfer of data.^[4] The initial framework, which embraced an empowerment approach, prioritised high-quality legal drafting, and voluntary adoption of the SA DTA, rather than mandating its use through legislation. This approach was welcomed by the scientific research community, recognising the diverse needs and capacities of institutions, and providing them with a tool to protect their interests while participating in global scientific collaborations.^[5]

Since its launch, the SA DTA has garnered constructive feedback from various stakeholders, prompting us to refine the template and explanatory memorandum. This updated version, dubbed SA DTA 1.1,^[6] introduces significant enhancements in key areas such as data ownership and intellectual property rights. These updates aim to provide clearer guidelines and greater flexibility, ensuring that the template can effectively govern a wide array of

research scenarios and partnerships. In this article, we introduce SA DTA 1.1, detailing the revisions and their implications for the research community, thus continuing our commitment to fostering a supportive environment for data-driven research collaborations.

Key updates in SA DTA 1.1

The updates to SA DTA 1.1 focus on data ownership and intellectual property rights. These updates were prompted and informed by stakeholder feedback that underscored the need for clarity on when exactly during the research process a research institution would have rights in the data that are generated, especially in collaborative research contexts.

The cornerstone of these updates is the new definition of ‘inferential data’. It is defined as follows:

“Inferential data” means data that arise not merely from the cleaning, ordering, or reformatting of the project data, or the combination thereof with other data, but from analysis of the project data that generates new knowledge or hypotheses that were not explicitly contained in the project data or its combination with other data.’

This definition is the fulcrum of the revised data ownership and intellectual property rights clauses. Essentially, the provider of the project data remains the owner thereof, but the recipient is the owner of the inferential data that are generated; if the inferential data are generated through collaborative analysis, they are owned jointly. Where intellectual property rights are applicable, the same applies.

This delineation ensures that the contributions of each party are acknowledged and protected under the terms of SA DTA 1.1, while concerns about the fair distribution of rights and responsibilities in research collaborations are sufficiently addressed. The updated SA DTA template explicitly states that inferential data, once created, confer ownership to the creator or joint creators, thereby eliminating ambiguities about the allocation of rights and enhancing the agreement's clarity and fairness.

In the next sections, we illustrate the operation of SA DTA 1.1 with two case studies.

Practical case study 1

In this case study, we examine a collaborative research project between two SA universities, University X, the data provider, and University Y, the data recipient. The project focuses on identifying genetic markers associated with neurodegenerative conditions such as Alzheimer's and Parkinson's diseases, using genomic data combined with medical records. We trace the technical steps involved in processing and analysing the data, emphasising data ownership and the division of responsibilities at each stage.

Data provision and initial handling

University X collects raw genomic data and associated medical records from participants who have consented for their data to be used specifically in this research. University X holds sole ownership of these raw data. University X provides these data to University Y, which undertakes the responsibility for further processing and analysis. Despite transferring the data, University X retains ownership of the original data set as per the provisions of SA DTA 1.1.

Data preparation and standardisation

University Y takes on the task of preparing and standardising the received data. This includes cleaning the data by removing incomplete or outlier data points, correcting identifiable errors, and standardising data formats to ensure consistency and accuracy across the data set. These activities, conducted by University Y, are essential for making the data suitable for advanced analysis but do not alter the ownership status of the original data, which remains with University X.

Analysis

University Y applies advanced statistical models and machine learning algorithms to the prepared data set to identify potential genetic markers linked to neurodegenerative diseases. This stage is crucial, as it involves generating new knowledge or hypotheses from the existing data, thereby creating inferential data as defined in SA DTA 1.1. Since these inferential data are derived solely through the efforts of University Y without collaborative input from University X in the analysis phase, ownership of the inferential data resides with University Y, according to SA DTA 1.1.

Validation and clinical implications

The hypothesised genetic markers identified by University Y are then subjected to further validation through additional experimental studies or clinical trials to confirm their relevance and accuracy. This validation process may involve the collection of new data, which, along with the validated inferential data, are owned by University Y, as they stem from its research efforts.

Publication and data management

The results of the research, including the validated genetic markers, are prepared for publication. University Y handles the drafting and publication process, reflecting its sole contribution to the inferential data generation. All rights related to these data and any potential applications are managed solely by University Y, in line with SA DTA 1.1.

Data management post-project

Upon completion of the project, University Y retains the inferential data and any additional data collected during the validation phase. University X retains ownership of the original raw data and may use them in further studies, complying with the original consent terms and the Protection of Personal Information Act No. 4 of 2013 (POPIA).

Practical case study 2

In this case study, we explore a research collaboration between University X, an SA university, and Company Z, an international pharmaceutical company. The project focuses on identifying genetic markers for chronic diseases such as diabetes and cardiovascular diseases using a combination of genomic data and medical records.

Data preparation and standardisation

As with the first case study, University X collects raw genomic data and associated medical records from participants who have consented for their data to be used in this research. University X holds sole ownership of these raw data. The initial technical step of data preparation and standardisation is conducted jointly with Company Z. This phase involves cleaning the data by removing incomplete or outlier data points, correcting identifiable errors, and standardising data formats to ensure consistency and accuracy across the data set. Despite Company Z's involvement, ownership remains with University X, as these activities are preparatory and do not involve generating new knowledge or inferential data.

Data integration

Following preparation and standardisation, University X and Company Z collaboratively manage the integration of the genomic data with medical records. This process results in a comprehensive data set linking genetic information with health outcomes, vital for subsequent analysis. Ownership remains with University X, as this stage, similar to case study 1, does not yet involve the creation of inferential data, per SA DTA 1.1.

Data annotation and enrichment

In this phase, similar to the previous case, the data set is supplemented with gene functions, known disease markers, and other relevant metadata sourced from scientific databases. Company Z leads this enhancement with its bioinformatics capabilities. This enrichment process prepares the data for deeper analysis but does not alter the ownership, which continues to reside with University X, as these activities do not generate inferential data.

Analysis

In this crucial phase, Company Z and University X collaborate to apply advanced statistical models and machine learning algorithms to the

enriched data set, mirroring the joint effort detailed in the first case study. Their goal is to identify complex patterns that may indicate potential genetic markers linked to chronic diseases. As this analysis generates inferential data – new knowledge derived from the analysis – the rights to the hypothesised genetic markers will be jointly held by University X and Company Z.

Validation and clinical implications

Similar to the previous study, the potential genetic markers undergo validation through experimental or clinical studies to confirm their accuracy. This step might involve collecting new data, which, together with the validated inferential data, remain jointly owned by University X and Company Z owing to their collaborative analysis efforts.

Publication and data management

The results, including the validated genetic markers, are prepared for publication with both University X and Company Z involved in co-authorship. The intellectual property rights related to the inferential data and any developments that arise from them are managed according to SA DTA 1.1, ensuring equitable benefits from commercial exploitation.

Data management post-project

At the project's conclusion, guidelines from SA DTA 1.1 dictate that University X retains ownership of the original and combined data sets, while the inferential data remain jointly owned. Both parties may retain copies for future research, adhering to the terms of participant consent and complying with the POPIA.

Discussion

The case studies presented in this article illustrate the practical implementation and operational efficacy of SA DTA 1.1 in diverse collaborative research settings, highlighting its flexibility and adaptability in addressing complex data governance challenges. Through these examples, the refined provisions of SA DTA 1.1 are demonstrated to effectively streamline the process of data ownership and intellectual property rights management, making the agreement template a robust tool for modern research collaborations.

Comparative insights between case studies

In the first case study with University X and University Y, data ownership remains clear and straightforward. University X retains ownership of the original raw data, while University Y, which performs the analysis, claims ownership of the inferential data generated. This scenario underscores SA DTA 1.1's ability to assign data ownership based on contribution to the inferential data, ensuring that contributions from the data-analysing party are justly recognised and rewarded.

The second case study involving University X and Company Z highlights a scenario of collaboration from the outset of the project, including data preparation and analysis. This joint effort leads to co-ownership of inferential data, reflecting SA DTA 1.1's capacity to facilitate equitable sharing of rights when multiple parties contribute significantly to the generation of new knowledge. This case study provides a contrast to the first, where a single entity took a more dominant role in the analysis phase.

The Exchange Control Regulations and the impact of the IPR Act

It is necessary to note a caveat at this stage. As set out in clause 14.5 of the SA DTA 1.1, SA has restrictions on the transfer of intellectual property as contained in the Exchange Control Regulations and also has important qualifications to intellectual property ownership as contained in the Intellectual Property Rights from the Publicly Financed Research and Development Act 51 of 2008 (IPR Act). These pieces of legislation may impact on how parties can exercise their rights in data. It is especially important for foreign research partners, such as Company Z in the second case study, to properly acquaint themselves with these laws.

Discussion of enhanced clarity and fairness

Both case studies emphasise how SA DTA 1.1 clarifies and simplifies the management of data ownership and intellectual property rights. By establishing that ownership of inferential data is directly tied to the creation of new knowledge, SA DTA 1.1 eliminates ambiguities that previously might have led to disputes or confusion. This clarity is particularly vital in collaborative research environments where multiple parties contribute variously to data generation, processing and analysis.

Broader implications and future directions

The clarity provided by SA DTA 1.1 in these case studies ensures that researchers can engage in collaborations without ambiguity regarding data ownership, thereby reducing potential conflicts and fostering more productive and ethical research partnerships. By effectively delineating rights based on the generation of inferential data, SA DTA 1.1 encourages a more nuanced understanding of contributions within research collaborations, which is crucial for advancing scientific research and developing new technologies.

Furthermore, the feedback loop from stakeholders, as illustrated by the evolution from SA DTA to SA DTA 1.1, reflects a responsive and adaptive approach to data governance. This engagement highlights our ongoing commitment to refining the agreement template as new challenges emerge, ensuring that it remains relevant and effective in a rapidly evolving research landscape.

Conclusion

SA DTA 1.1 marks a significant advancement in the governance of research data in SA and beyond. Building on its predecessor, this updated agreement template embodies a commitment to enhancing the legal framework that supports the safe, ethical and efficient transfer of data across the research community. By introducing updates in critical areas such as data ownership and intellectual property rights, SA DTA 1.1 ensures that researchers can navigate the complexities of data sharing with clarity and confidence.

The case studies presented in this article illustrate the practical application of SA DTA 1.1, demonstrating its effectiveness in addressing real-world challenges faced by researchers in collaborative environments. These scenarios show how the agreement adapts to various degrees of collaboration and contribution, ensuring that all parties involved receive fair recognition and protection for their efforts. This adaptability is a core strength of SA DTA 1.1, making it a valuable tool for a wide array of institutional needs and specific research contexts.

Furthermore, SA DTA 1.1 enhances the empowerment approach by providing not only a professionally drafted template but also a detailed explanatory memorandum that helps institutions, especially those with limited legal resources, to engage in international collaborations without compromising their autonomy or the protection of their data. This approach not only fosters a supportive environment for data-driven research collaborations but also promotes the sharing of knowledge in a manner that respects the rights and privacy of all parties involved. In this way, SA DTA 1.1 is more than just a data transfer agreement template; it is a testament to the ongoing effort to foster a research environment where data can be shared responsibly and productively.

Declaration. None.

Acknowledgements. The authors acknowledge the use of ChatGPT4 to write summaries, and to improve the language and readability of this article.

Author contributions. All the authors contributed to SA DTA 1.1. With regard to this marker article, DT authored the first draft; all the authors revised the article for critically important intellectual content, including final approval of the version to be published.

Funding. The authors are grateful for support from the US National Institute of Mental Health and the US National Institutes of Health (award no. U01MH127690). The content of this article is solely the authors' responsibility and does not necessarily represent the official views of the US National Institute of Mental Health or the US National Institutes of Health.

Conflicts of interest. The authors declare that there are no conflicts of interest regarding the publication of this article. MB is an attorney and researcher at BioLAWgic; PE is an attorney in private practice; LS is legal counsel for Centrica PLC in the UK; and DT has a private advocate's practice. These affiliations have not influenced the research or conclusions presented in this article.

1. Colcelli V. Applying national law in cross-border research activity. In: Colcelli V, Cippitani R, Brochhausen-Delius C, Arnold R, eds. *GDPR Requirements for Biobanking Activities Across Europe*. Cham, Switzerland: Springer Nature, 2023;39-49. https://doi.org/10.1007/978-3-031-42944-6_5
2. Swales L, Ogendi P, Botes M, et al. A data transfer agreement template for South Africa. Zenodo 2023. <https://doi.org/10.5281/zenodo.7537396>
3. Swales L, Botes M, Donnelly DL, Thaldar D. Towards a data transfer agreement for the South African research community: The empowerment approach. *S Afr J Bioethics Law* 2023;16(1):13-18. <https://doi.org/10.7196/SAJBL.2023.v16i1.827>
4. Townsend B. The lawful sharing of health research data in South Africa and beyond. *Inform Commun Techn Law* 2022;31(1):17-34. <https://doi.org/10.1080/13600834.2021.1918905>
5. 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>
6. Thaldar D, Botes M, Swales L, Esselaar P. (contributors to version 1.1) A data transfer agreement template for South Africa, version 1.1. Zenodo 2024. <https://doi.org/10.5281/zenodo.11355144>

Received 7 June 2024. Accepted 21 June 2024.