Words matter: Using the Framework Method to analyse the definitions of 'anonymise', 'pseudonymise', 'de-identify', and 'not de-identified'

M van Niekerk, BProc, BOccTher, MSc Med (Bioethics and Health Law), PhD candidate

Department of Occupational Therapy, School of Therapeutic Sciences, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding author: *M van Niekerk (matty.vanniekerk@wits.ac.za)*

A significant number of students handle personal information (PI) during work-integrated learning and supervised research, requiring compliance with both South African (SA) law and internationally derived research ethics instruments. Where these instruments use different terminology, students face barriers to understanding, reflecting not only linguistic differences but materially distinct legal standards and regulatory consequences.

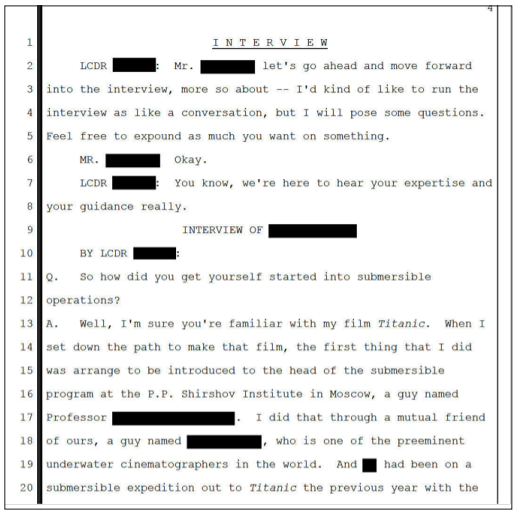
Using the Framework Method across seven analytical dimensions, this study examined the conceptual equivalence of key PI protection terms across four sources: the Oxford English Dictionary, the European Union General Data Protection Regulation (GDPR), the Council for the International Organizations of Medical Sciences (CIOMS) ethical guidelines for human research, and SA's Protection of Personal Information Act 4 of 2013 (POPIA).

Three findings emerged. First, GDPR's 'anonymisation' and POPIA's 'de-identification', although similar in stated purpose, apply different threshold tests: GDPR uses a probability test (whether re-identification is reasonably likely), while POPIA applies a capacity test (whether re-identification is possible by any reasonably foreseeable method). A data set compliant under GDPR may therefore not satisfy POPIA. Second, the instruments diverge on regulatory scope: GDPR treats anonymisation as an exit from regulation, whereas POPIA imposes ongoing obligations on de-identified data, including an explicit re-identification prohibition absent from GDPR. Third, CIOMS and POPIA are more conceptually compatible with each other than either is with GDPR, yet CIOMS employs GDPR-tradition terminology, creating misleading signals of equivalence for SA students.

These divergences reflect fundamentally different conceptions of PI and the rationale for its protection. Students, educators and institutional governance structures must address these distinctions explicitly, rather than assuming terminological equivalence across instruments.

Keywords: de-identification, anonymisation, POPIA, work-integrated learning, research ethics, Framework Method, conceptual equivalence, GDPR

S Afr J Bioethics Law 2026;19(1):e4276. https://doi.org/10.7196/SAJBL.2026.v19i1.4276



I N T E R V I E W

LCDR [redacted]: Mr. [redacted] let's go ahead and move forward into the interview, more so about -- I'd kind of like to run the interview as like a conversation, but I will pose some questions. Feel free to expound as much you want on something.

MR. [redacted]: Okay.

LCDR [redacted]: You know, we're here to hear your expertise and your guidance really.

INTERVIEW OF [redacted]

BY LCDR [redacted]:

Q. So how did you get yourself started into submersible operations?

A. Well, I'm sure you're familiar with my film *Titanic*. When I set down the path to make that film, the first thing that I did was arrange to be introduced to the head of the submersible program at the P.P. Shirshov Institute in Moscow, a guy named Professor [redacted]. I did that through a mutual friend of ours, a guy named [redacted], who is one of the preeminent underwater cinematographers in the world. And [redacted] had been on a submersible expedition out to *Titanic* the previous year with the

Fig. 1. Redacted interview with 'Deep Sea Explorer'.[1]

Who is Mr [Redacted]? Can he be reidentified using any reasonably foreseeable means, now or in the future? Is the Lieutenant Commander (LCDR) similarly identifiable?

The transcript in Fig. 1 has had names redacted, yet most readers will immediately recognise the interviewee. This example illustrates a central problem in data protection: removing the most obvious identifier (the name) does not constitute de-identification/anonymisation if sufficient contextual information remains to make re-identification trivial. What defines sufficiently de-identified differs between different instruments, such as the General Data Protection Regulation (GDPR)[2] of the European Union (EU) and the South African (SA) Protection of Personal Information Act 4 of 2013 (POPIA),[3] extending to reasons beyond different terminology alone.

Despite personal information (PI) processing increasingly being regulated worldwide,[4] a notable population of students process members of the public's PI during work-integrated learning (WIL) opportunities and for supervised research purposes. In health sciences this occurs during clinical rotations in public and private hospitals and clinics; in education, during teaching practicums at schools. The PI processing is an intentional and necessary design feature of WIL, without

which students cannot achieve the competencies their programmes require. It therefore carries full legal and ethical obligations under applicable data protection legislation.

In SA, those obligations arise from POPIA.[3] Where students process PI for research purposes, a second layer of obligation arises under human research ethics instruments such as the Council for the International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Health-Related Research Involving Humans[5] and the Declaration of Helsinki.[6] These two legal and ethical governance systems are not identical and cannot be collapsed into each other: POPIA is domestic legislation with enforcement consequences, while research ethics instruments are normative instruments developed by international bodies and embedded in the legal and regulatory language of their jurisdictions of origin. This situation creates a structural barrier to understanding for SA students. Research ethics guidelines such as the CIOMS guidelines, for example, use words such as 'anonymise' and 'pseudonymise' that originate in the GDPR,[2] whereas POPIA[3] refers to 'de-identify' and 'not de-identified'. These terms are not synonymous.[7] A student who applies a GDPR-derived understanding of 'anonymised' in a POPIA context makes a material interpretation error, because for a non-law student the terminology gives no signal that different standards apply.

The literature continues to illustrate considerable confusion about 'de-identify' and 'anonymise', and whether they are equivalent and/or can be used interchangeably.[7-10] A scoping review of biomedical literature found that approximately 32% of researchers use de-identification and anonymisation as synonyms, a further 32% treat them as strictly different processes, and fewer than half of reviewed articles provided any formal definition at all.[9] Persistent inconsistency in how electronic health record data are described across research institutions has been documented in multi-centre research contexts.[8] It has further been argued that the field requires urgent consensus on anonymisation terminology, noting that initiatives designed to facilitate cross-border data sharing, e.g. DataSHIELD,[11] are undermined when human research ethics committees (HRECs) in different countries cannot establish equivalency between the data labels researchers use.[11] While it has rightly been pointed out that using either of these terms is not a matter of preference but of adherence to jurisdiction and legislated definitions,[7] scholars outside of the legal fraternity, e.g. general researchers and healthcare practitioners/educators, are unlikely even to know that a term has a legislated definition, especially terms originating outside of their country. The case for explicit terminological contextualisation in student training is therefore strong.

This article is positioned at the intersection of law and higher education, rather than being a doctrinal legal analysis. It treats legislative texts, regulatory instruments and dictionary definitions as sources of data to be compared systematically using the Framework Method.[12,13] The aim is conceptual clarity for a non-specialist audience: students, educators, supervisors, and the institutional governance structures of WIL programmes who must navigate these instruments in practice without the benefit of legal training. Conceptual clarity in this context means understanding what each term actually requires, how those requirements differ across instruments, and why they differ. The 'why' is not a question of academic interest only: terms are embedded in legal and regulatory traditions that reflect different historical experiences and foundational rights philosophies. Understanding the traditions is necessary to enable non-law students to interpret what those terms mean in practice.

The article addresses the following research question: to what extent are the key PI protection terms used in internationally derived research ethics instruments conceptually equivalent to their counterparts in POPIA, and what are the implications of any divergence for students processing PI during WIL and for research purposes? To answer this question, the Framework Method was applied to four sources: Oxford English Dictionary (OED) definitions,[14] the GDPR, the CIOMS guidelines, and POPIA. The aim is not to determine which framework is superior, but to map where the instruments converge, where they diverge, and what the practical consequences of that divergence are for SA students and the institutions responsible for their training.

Methods

The Framework Method is a structured qualitative approach for systematically analysing how multiple sources conceptualise the same phenomenon across a set of predefined analytical dimensions.[12,13] Originally developed by Ritchie and Spencer in 1994 for applied policy research, it has been widely adopted in health and social research contexts where the aim is to produce interpretable, transferable findings from a defined set of sources rather than to generate grounded theory from open-ended data.[13]

Sources

Four sources were selected for the analysis using the Framework Method. The OED was selected as the authoritative record of English-language usage over time, providing a baseline for how these terms function outside formal regulatory contexts and allowing assessment of the gap between colloquial and legislated meanings.[14] The GDPR was selected because prominent international research ethics instruments, including the CIOMS guidelines, are embedded in GDPR-tradition terminology. Engagement with the GDPR is therefore necessary to interpret what those instruments mean when they use terms such as 'anonymise' and 'pseudonymise'.[2] The CIOMS guidelines for health-related research involving humans were selected because they represent the most comprehensive international research ethics framework applicable to health sciences research and are widely used in SA research ethics governance.[5] POPIA was selected as the domestic legislative framework governing all PI processing in SA, including processing by students during WIL and research.[3]

It is acknowledged that this selection does not capture the full range of relevant instruments. The US Health Insurance Portability and Accountability Act (HIPAA), and data protection instruments from other African jurisdictions, represent important comparators that fall outside the scope of this analysis. This is a considered scoping decision rather than an oversight: the four sources selected represent the instruments most directly relevant to an SA student processing PI during WIL or research, and adding further sources would require a substantially larger study. The GDPR-CIOMS-POPIA triangle captures the most consequential terminological tensions for this audience.

Analytical dimensions

The seven analytical dimensions were identified deductively through iterative close reading of the four source texts, informed by reviewer feedback on an earlier version of the manuscript and by engagement with the comparative legal literature: definition, legal status, threshold, reversibility, scope consequence, re-identification treatment, and jurisdictional origin. Each dimension was selected on the basis that it captures a practically consequential aspect of how the instruments conceptualise identifier removal; aspects that could not be collapsed into each other without loss of analytical precision. The dimensions were applied systematically to each source in turn before cross-source comparison was undertaken.

Results and interpretation

The findings of the Framework Method analysis are presented in a matrix (Table 1) that plots each of the four sources against each of the seven dimensions.

The Framework Method analysis produced three interpretive findings, each building on the last. The first establishes that terminological overlap conceals material legal divergence; the second identifies where that divergence is structurally deepest; and the third explains why the divergence is particularly consequential for the specific audience this article addresses.

Finding 1: Overlapping terminology describes materially different legal standards

The most immediate and consequential finding of the framework analysis is that the four sources use terms that appear equivalent but impose different requirements. This is not a question of imprecise language. Rather, it is an inherent feature of how data protection terminology has developed across jurisdictions with different foundational orientations.

The clearest example is the relationship between anonymisation (GDPR) and de-identification (POPIA). At the level of stated purpose, the two concepts are equivalent: both require that identifiers are removed so thoroughly that re-identification is impossible. Edgcumbe *et al*.[7] confirm this legislative equivalence, and it is not disputed here. The divergence lies not in what the two concepts aim to achieve, but in the test applied to determine whether the aim of de-identification has been met.

The GDPR applies a probability test: PI is anonymised if re-identification is not reasonably likely, considering all means reasonably likely to be used at the time of assessment (Recital 26).[2] POPIA applies a capacity test: PI is de-identified only if there is no reasonably foreseeable method by which re-identification could occur.[3] The distinction is not immaterial. The GDPR asks whether re-identification is probable now; POPIA asks whether re-identification is possible at any foreseeable point. These foundational differences have practical consequences that extend beyond terminology. When research data move between a GDPR jurisdiction and SA, they move between different foundational conceptions of what PI is and what its protection means.

Consider a healthcare dataset from which names and identity numbers have been removed, but which retains rare diagnoses, geographical location, and dates of treatment. Under the GDPR, a data controller might legitimately conclude that re-identification with current technology and reasonable effort is not probable; the PI is anonymised and exits regulatory scope entirely. Under POPIA, the same dataset may not qualify as de-identified, because the combination of rare diagnosis, location and date creates a foreseeable pathway to re-identification through cross-referencing with hospital records or community knowledge. The PI remains within POPIA's regulatory scope regardless of how thoroughly names, identity numbers and other identifiers have been removed. This divergence cannot be resolved by contractual mechanisms or adequacy decisions alone: it requires that persons handling the PI understand the foundational difference, not merely the surface terminology.[7,17]

A student who has internalised the GDPR-tradition model, whether from CIOMS guidance, international literature or prior training, is likely to apply the probability test unknowingly. In an SA context, that student is making a significantly incorrect legal judgment, not a linguistic error. The Framework Method analysis makes this barrier to understanding visible.

Finding 2: The instruments diverge most sharply on reversibility, scope consequence, and re-identification treatment, which the pseudonymisation gap compounds

The threshold divergence identified in finding 1 might suggest that the GDPR and POPIA are essentially the same concept expressed with different levels of stringency. The second finding corrects this inference: the frameworks differ structurally, and that structural difference reflects fundamentally different foundational orientations towards what PI is and why its protection matters.

The GDPR flows from Article 8 of the EU Charter of Fundamental Rights, which establishes data protection as a fundamental right in its own right, grounded in human dignity and individual autonomy.[2] The foundational logic is that PI is an extension of the person: its protection is inherent, not contingent on demonstrable harm.[18] This produces a binary, exit-orientated model: once PI meets the anonymisation threshold, it is no longer PI, the GDPR ceases to apply,[2,18] and the classification should be stable. Anonymisation is the mechanism by which PI moves permanently outside regulatory scope. This binary logic also shapes the GDPR's threshold test: the question is whether re-identification is *reasonably likely*, which is a probability assessment about current conditions.

POPIA does not operate in this way, and the reason is historical as much as legal. SA's apartheid history, which was defined by systematic state surveillance, racial classification, and the use of PI as an instrument of oppression,[19-21] created a constitutional imperative for robust privacy protection grounded not in individual autonomy but in a dignity-based rights tradition shaped by *ubuntu*.[22-24] PI that was not dangerous in one political context became a tool of oppression in another, which contributes to why de-identified PI does not exit POPIA's scope in the same clean manner as anonymised PI exits the GDPR. Section 37 explicitly prohibits re-identification as a distinct act, retaining regulatory interest in de-identified PI and treating de-identification not as a permanent stable state but as a condition that must be actively maintained. This is not a drafting variation, but a different conception of what data protection requires, grounded in context.

The pseudonymisation gap deepens this structural difference. The GDPR's binary anonymisation model requires an intermediate

Table 1. Framework analysis matrix: Conceptualisation of personal identifier removal across four sources, including pseudonymisation

Analytical dimension	OED[14]	GDPR[2] (EU)	CIOMS guidelines[5] (international)	POPIA[3] (SA)
1a. Definition of anonymise	Removes identifying information or proof of identity. No statutory force. Usage spans 'calling someone anonymous' to 'removing data identifiers' – the meaning has shifted substantially over time (see supplementary material available online at XXXXXX).	Information that does not relate to an identified or identifiable natural person, or PI rendered anonymous so the data subject is not, or no longer, identifiable (Recital 26).	PI from which identifying information has been removed or altered so that the individual cannot be identified. CIOMS acknowledges that true anonymisation is increasingly difficult to achieve in practice.	To delete any information that identifies the data subject, or that can be used or linked by reasonably foreseeable methods to identify the data subject (section 1).
1b. Definition of pseudonymise (for comparison)	**Historically near-equivalent to anonymise in OED usage** *Key OED finding:* The dictionary definitions of anonymise and pseudonymise converge on the act of concealment. It is the legislated definitions that pull them sharply apart. A student relying on colloquial or dictionary understanding has no reliable signal that these terms carry materially different legal consequences.	**Sharply distinguished from anonymisation by the GDPR** *The GDPR distinction is categorical:* Anonymisation = exit from scope; pseudonymisation = risk reduction within scope.	**Coded data** *Practical consequence for SA students:* CIOMS guidance on anonymisation may be read as applying to what POPIA would call de-identification, while CIOMS guidance on coded data (akin to pseudonymisation) has no clear POPIA statutory equivalent.	**Pseudonymisation does not appear in POPIA's statutory text** *This gap is particularly consequential for WIL and research contexts* where full de-identification is often not feasible. It should be noted that some sources do introduce pseudonymisation[15] or pseudo-anonymisation[16] into a POPIA context, but this is not true to the Act.
2. Legal/normative status	Descriptive only. No legal force. Reflects general usage, not regulatory standards. Colloquial use frequently diverges from legislated meaning.	Binding EU law. Enforceable by national supervisory authorities. Non-compliance triggers administrative fines up to EUR20 million or 4% of global annual turnover.	Normative guidelines. Not binding legislation. Adopted voluntarily by institutions and required by research ethics committees as a condition of ethics approval.	Binding SA statute. Enforced by the Information Regulator. Non-compliance may trigger administrative fines of up to ZAR10 million, civil claims, or criminal liability including statutory vicarious liability for institutions.
3. Threshold/test	No formal threshold. Dictionary examples range from removing a name to comprehensively stripping identifiers – the test applied varies across examples.	**Probability test:** Whether re-identification is – *reasonably likely,* taking into account all means reasonably likely to be used (Recital 26). *Current conditions* determine the assessment.	Aligned with GDPR probability orientation, though expressed differently. CIOMS asks whether there is a *realistic possibility* of re-identification. Acknowledges that genomic data and data linkage present increasing re-identification risk regardless of precautions taken.	**Capacity test:** Whether re-identification is possible by *any reasonably foreseeable method* (section 1). The test does not require likelihood. Possible future capacity is sufficient to prevent a finding of de-identification. More stringent than GDPR.
4. Reversibility	Silent on reversibility. Examples range from what appear to be reversible acts (using initials, withholding a name) to irreversible ones (destruction of identifiers). No consistent standard.	Binary/irreversible in principle. Anonymisation is treated as a stable categorical state: once anonymised, data are no longer personal data and exits the instrument's regulatory scope. The classification should be stable.	Aligned broadly with GDPR irreversibility principle, but acknowledges practical instability: technological advances mean that PI considered anonymous today may become re-identifiable in future. Irreversibility treated as aspirational rather than guaranteed.	Conditional/spectrum orientated. De-identification is not treated as a permanent categorical state. Section 37 regulates re-identification as an ongoing risk, implying that de-identified status can be compromised and must be continuously maintained.
5. Scope consequence	No regulatory consequence attached. Dictionary definition carries no implication about what obligations attach or cease upon anonymisation.	**Exit from regulatory scope**	Reduced obligations rather than full exit. Anonymised data may allow waiver of certain requirements (e.g. individual informed consent), but do not remove research ethics oversight entirely. Ethics review remains required.	**Retention of regulatory interest**

(continued)

Table 1. (continued) Framework analysis matrix: Conceptualisation of personal identifier removal across four sources, including pseudonymisation

Analytical dimension	OED[14]	GDPR[2] (EU)	CIOMS guidelines[5] (international)	POPIA[3] (SA)
6. Treatment of re-identification	Not addressed. No dictionary definition of re-identification exists as a distinct concept in relation to data protection.	Re-identification is treated as a risk to be managed at the point of anonymisation: if re-identification is reasonably likely, the data are not anonymised. Re-identification after the fact is not separately regulated as an act.	Re-identification risk is a design consideration. Researchers are expected to assess and minimise re-identification risk prospectively. CIOMS does not regulate re-identification as a distinct prohibited act.	**Re-identification is explicitly prohibited as an act (section 37)**
7. Jurisdiction/ origin	English-language dictionary. No single jurisdiction. Reflects general Anglophone usage. Particularly relevant as a proxy for how non-specialist SA students are likely to interpret these terms before formal training.	EU. Grounded in the EU Charter of Fundamental Rights, Article 8: data protection as a fundamental right rooted in individual dignity and autonomy. Binary, exit-orientated model reflects this rights-inherent logic.	International (Geneva based). Developed primarily in GDPR-tradition regulatory language. Used in SA research ethics governance as if jurisdiction-neutral, but terminology carries GDPR-origin assumptions that may not map onto POPIA.	SA. Grounded in section 14 of the Constitution (right to privacy). Precautionary, capacity-based orientation reflects the historical context of state misuse of PI. Section 37's prohibition on re-identification reflects a regulatory logic of ongoing vigilance shaped by that history.

OED = Oxford English Dictionary; GDPR = General Data Protection Regulation; EU = European Union; CIOMS = Council for the International Organizations of Medical Sciences; POPIA = Protection of Personal Information Act; SA = South Africa; PI = personal information; WIL = work-integrated learning.

category for PI that is protected but not fully anonymised, and pseudonymisation fills that gap (Recital 28).[2] Pseudonymised PI remains PI under the GDPR but benefits from reduced obligations; it is a named, legally recognised risk management measure. POPIA has no statutory equivalent. Although pseudonymisation appeared in a proposed Code of Conduct for Research,[16] it was not approved by the Information Regulator, and rather became a voluntary framework without regulatory authority.[7] Pseudonymisation's absence is not an oversight, but a consequence of POPIA's different underlying logic. Where the GDPR manages the spectrum between full PI and fully anonymised PI through categorical distinctions, i.e. PI, pseudonymised PI, anonymous PI, POPIA manages it through the stringency of the de-identification test and the ongoing prohibition on re-identification. The practical consequence for students in WIL and research contexts is significant: PI that cannot be fully de-identified under POPIA's stringent capacity test has no recognised intermediate status. It is either de-identified or it is not, and if it is not, full POPIA obligations apply.

Finding 3: CIOMS and POPIA are more conceptually compatible than their terminology suggests, but the terminology creates false signals for SA students

The third finding is perhaps the most important for this article's practical argument. A reader of the Framework Method analysis matrix might conclude that the primary problem is the divergence between POPIA and GDPR, and that CIOMS simply reproduces the GDPR position. Such a reading is too simple. The CIOMS guidelines treat anonymisation in a more nuanced manner than a direct GDPR restatement. The CIOMS guidelines explicitly acknowledge that complete anonymisation is increasingly difficult to achieve in practice, given the capacity for cross-matching large datasets, genomic technologies, and the narrowness of many research populations.[5] This acknowledgement aligns CIOMS closer to POPIA's precautionary orientation than to the GDPR's binary model: both CIOMS and POPIA treat the achievability of full anonymisation or de-identification as something that must be continuously monitored and maintained, rather than something that is permanently achieved. In this respect, an SA researcher reading CIOMS carefully is receiving guidance that is more compatible with POPIA's underlying logic than the terminology alone would suggest.

The problem lies in the terminology. The CIOMS guidelines use anonymisation, which is a GDPR-tradition term that carries the GDPR's binary, exit-orientated associations to describe a concept that in practice functions more like POPIA's precautionary, spectrum-orientated approach. When an SA student reads the word 'anonymisation' in CIOMS guidance, they are likely to interpret it through the GDPR-tradition understanding of that word, an understanding reinforced by the dictionary, the literature, and most international research ethics training. They will apply a probability-based threshold without knowing it and draw the wrong conclusion about what POPIA requires. This is particularly unfortunate because CIOMS's own treatment of anonymisation is more cautious than the GDPR's. However, that nuance is lost when the terminology carries assumptions from a different jurisdictional tradition.

This finding has a direct implication for the educational and institutional governance recommendations that follow. The problem is not that CIOMS gives wrong guidance, it is that CIOMS guidance requires explicit domestic contextualisation before it can be applied correctly by an SA student. Making that contextualisation explicit is the responsibility of educators, supervisors, and the institutional governance structures of WIL programmes. A student

who understands that the word 'anonymisation' in CIOMS carries a different threshold and a different regulatory consequence to the word 'de-identification' in POPIA is equipped to apply both instruments correctly. A student who does not understand this is at risk of inadvertent non-compliance, with potentially serious consequences under a statute that imposes strict vicarious liability on institutions (POPIA, section 99).

The three findings above establish that terminological divergence between POPIA and internationally derived research ethics instruments is not a surface-level naming problem. Instead, it reflects materially different legal standards with real consequences for the students with whom this article is concerned.

Discussion

The terminological confusion this article addresses is well documented,[7-10] with findings including many researchers either using the terms interchangeably or not defining them in publications.[9] Such findings confirm that the barrier to understanding this article identifies is not confined to students or novice researchers: it is pervasive across the professional literature and has practical consequences for international research collaboration.

A recent comparative analysis across 12 African jurisdictions, with the GDPR and HIPAA as comparators, confirms the core finding of this article: POPIA's de-identification standard is more stringent than GDPR's anonymisation standard and does not operate as a categorical exit from regulatory scope.[7] Comparative analysis of GDPR and POPIA data protection principles more broadly reaches compatible conclusions.[17,25]

This article addresses an audience and a problem that the legal comparative literature[7,17,25] has not examined: the consequence for non-law students of being trained on internationally derived research ethics instruments without explicit domestic contextualisation. It also treats the CIOMS guidelines as a distinct normative governance layer alongside data protection legislation. This dimension is absent from purely legal comparative analyses. The finding that CIOMS is more conceptually compatible with POPIA than its GDPR-tradition terminology suggests, is only visible from this interdisciplinary vantage point.

This article has several limitations. The Framework Method analysis is confined to four sources, the OED, GDPR, CIOMS and POPIA, and does not include the HIPAA or data protection instruments from other African jurisdictions. The terminological landscape across the African continent is considerably more varied than a four-source analysis captures.[7] Future research should extend the comparison to include a wider range of jurisdictions and instruments. The analysis also lacks input from the primary stakeholders, e.g. students, WIL supervisors, and research ethics committee members whose lived experience of the comprehension barrier described here would considerably enrich the findings. Empirical research into how students and supervisors actually navigate these terminological differences in practice remains an important gap in the literature.

Implications for practice

Researchers

Researchers operating under POPIA should use legislated definitions, 'de-identified' and 'not de-identified', in their publications and protocols rather than defaulting to the GDPR-tradition terms 'anonymised' and 'pseudonymised'. Where internationally derived instruments such as the CIOMS guidelines are used to frame a study, researchers should explicitly note the terminological translation required for the SA context, and should not assume that compliance with CIOMS language constitutes compliance with POPIA.

Two examples of practical decisions about study design can be highlighted. In medically based research, the need to follow up is often foreseeable, and this precludes de-identification under POPIA's capacity-based threshold. Qualitative researchers face particular challenges, as thick descriptions of participants that are methodologically necessary can render de-identification impossible. This tension can be managed practically by describing participants collectively, and presenting contextual details in broader strokes.

Human research ethics committees

HRECs occupy a critical gatekeeping position in relation to the comprehension barrier this article identifies. Their documentation and researcher training should include explicit definitions of 'de-identified', 'not de-identified', and, where relevant, 'pseudonymised' as used in voluntary compliance frameworks, e.g. the Academy of Sciences of South Africa guidelines, alongside clear guidance on the processes each requires. HRECs should alert researchers to the terminological differences between POPIA and the instruments they are applying, and should not treat the use of CIOMS language as sufficient evidence of POPIA compliance.

Educators

Students in health sciences and education who process PI during WIL are not equipped by general data protection awareness training to navigate the terminological divergence this article identifies. Educators must explicitly teach the difference between POPIA's capacity-based de-identification threshold and the probability-based threshold that underpins internationally derived research ethics instruments. This is not a supplementary topic, but a core competency for any student whose programme includes WIL or research involving PI.

Students should be taught to pseudonymise as a routine practice: identifiers must be deliberately separated from other PI and stored separately, with access to the identifiers controlled by the WIL placement or research supervisor rather than retained on the student's own devices. Universities must ensure that learning requirements do not oblige students to store PI on personal electronic devices, which creates both a POPIA compliance risk and a practical barrier to the pseudonymisation practices students are being taught.

Institutional governance of WIL programmes

The responsibility for contextualising internationally derived instruments within the domestic legal framework rests with institutional governance structures. It is not a curriculum problem that can be solved by adding a module on data protection terminology to an already crowded curriculum. Contextualising internationally derived instruments within the domestic legal framework means that WIL placement agreements, student orientation materials, and supervisory frameworks should include explicit guidance on POPIA obligations and their relationship to the research ethics instruments students will use. The responsibility

for this contextualisation rests with institutions, not with students who cannot reasonably be expected to identify a terminological divergence they have not been told exists.

Conclusion

This article sets out to determine the extent to which the key PI protection terms in internationally derived research ethics instruments are conceptually equivalent to their counterparts in POPIA, and to identify the implications of any divergence for students processing PI during WIL and research. The Framework Method analysis of four sources, the OED, the GDPR, the CIOMS guidelines and POPIA, shows that the terms are not interchangeable at contextual or conceptual levels. Although anonymisation and de-identification share the same stated purpose, they apply different threshold tests, reflect different foundational orientations towards what data protection requires, and carry different regulatory consequences. The CIOMS guidelines use GDPR-tradition terminology that creates false signals of equivalence for SA students who have not been explicitly taught otherwise.

The broader significance of the findings extends beyond the SA context. Terminology in data protection law is never neutral: it carries the legal traditions, foundational rights philosophies, and historical experiences of the jurisdictions that produced it. When that terminology travels across jurisdictions through internationally derived research ethics instruments, it does not travel alone but carries its assumptions with it. Recognising this is a prerequisite for ethical practice in a globalised research environment. It is not, however, a recognition that can reasonably be expected of non-law students navigating these instruments without guidance. Making it explicit is the responsibility of the institutions, educators, and governance structures within which and with whom those students work.

Declaration. The research for this study was done in partial fulfilment of the requirements for MvN's PhD degree at the University of the Witwatersrand.

Acknowledgements. The author wishes to thank the anonymous reviewers of the article, whose input strengthened the article. Artificial intelligence was used to assist with creating the abstract and to summarise definitions for the purpose of the Framework Method analysis matrix.

Author contributions. Sole author.

Funding. None.

Conflicts of interest. None.

1. Marine Board of Investigation. CG115 interview Deep Sea Explorer redacted. Washington, DC, September 2025. https://media.defense.gov/2025/Sep/17/2003800984/-1/-1/0/CG-115_INTERVIEW-DEEP-SEA-EXPLORER_REDACTED.PDF (accessed 20 October 2025).
2. European Parliament, European Council. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1. 2016. https://gdpr-info.eu/ (accessed 16 September 2025).
3. South Africa. Protection of Personal Information Act 4 of 2013. https://www.gov.za/documents/protection-personal-information-act (accessed 16 September 2025).
4. South African Law Reform Commission. Project 124 – privacy and data protection report. 2009. https://www.justice.gov.za/salrc/reports/r_prj124_privacy%20and%20data%20protection2009.pdf (accessed 16 September 2025).
5. Council for the International Organizations of Medical Sciences (CIOMS). International ethical guidelines for health-related research involving humans. Geneva: CIOMS, 2016. https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf (accessed 5 August 2025).
6. World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Participants. 19 October 2024. https://www.wma.net/policies-post/wma-declaration-of-helsinki/ (accessed 27 February 2026).
7. Edgcumbe A, Botes M, Donnelly DL, Townsend B, Shachar C, Thaldar D. 'Potato potahto'? Disentangling de-identification, anonymisation, and pseudonymisation for health research in Africa. J Law Biosci 2025;12(1):lsae029. https://doi.org/10.1093/jlb/lsae029
8. Kushida CA, Nichols DA, Jadrnicek R, Miller R, Walsh JK, Griffin K. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. Med Care 2012;50(Suppl 1):S82-S101. https://doi.org/10.1097/MLR.0B013E3182585355
9. Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and understanding of anonymization and de-identification in the biomedical literature: Scoping review. J Med Internet Res 2019;21(5):e13484. https://doi.org/10.2196/13484
10. Swales L. The Protection of Personal Information Act and data de-identification. S Afr J Sci 2021;117(7/8). https://doi.org/10.17159/SAJS.2021/10808
11. Wallace SE. What does anonymization mean? DataSHIELD and the need for consensus on anonymization terminology. Biopreserv Biobank 2016;14(3):224-230. https://doi.org/10.1089/bio.2015.0119
12. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess RG, eds. Analyzing Qualitative Data. New York: Routledge, 2002:173-194.
13. Gale N, Heath G, Cameron E, Rashid S, Redwood S. Using the Framework Method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13:117. https://doi.org/10.1186/1471-2288-13-117
14. Oxford English Dictionary [Internet]. 2024. https://www.oed.com/?tl=true (accessed 26 September 2025).
15. De Stadler E, Luttig Hattingh I, Esselaar P, Boast J. Over-thinking the Protection of Personal Information Act: The Last POPIA Book You Will Ever Need. Cape Town: Juta, 2021.
16. Adams R, Adeleke F, Anderson D, et al. POPIA Code of Conduct for Research. S Afr J Sci 2021;117(5-6). https://doi.org/10.17159/SAJS.2021/10933
17. Roos A. Data protection principles under the GDPR and the POPI Act: A comparison. THRHR 2023;86(1):1-26.
18. Syed H, Genç Y. General data protection regulation: A transformative law. Balkan J Soc Sci 2020;9(17):209-216. https://dergipark.org.tr/en/download/article-file/1140029 (accessed 27 March 2026).
19. Legodi LF, Mukonza RM. The promise and peril of access to information as a human right: A critical analysis of South Africa's experience. J Public Adm 2024;59(4):708-726. https://doi.org/10.53973/jopa.2024.59.4.a4
20. Mahomed S. The evolution of privacy governance in healthcare in post-apartheid South Africa. In: Dove ES, ed. Confidentiality, Privacy, and Data Protection in Biomedicine: International Concepts and Issues. London: Routledge, 2024:150-170.
21. International Network of Civil Liberties Organizations. Surveillance and democracy: Chilling tales from around the world. 11 October 2016. https://inclo.net/publications/surveillance-and-democracy-chilling-tales-from-around-the-world/ (accessed 28 March 2026).
22. Metz T. *Ubuntu* as a moral theory and human rights in South Africa. Afr Hum Rights Law J 2011;11(2):532-559. https://hdl.handle.net/10520/EJC51951 (accessed 27 March 2026).
23. Radebe SB, Phooko MR. *Ubuntu* and the law in South Africa: Exploring and understanding the substantive content of *ubuntu*. S Afr J Philos 2017;36(3):239-251. https://doi.org/10.1080/02580136.2016.1222807
24. Mokgoro Y. *Ubuntu* and the law in South Africa. Potchefstroom Electron Law J 1998;1(1):17-27. https://perjournal.co.za/article/view/2897/2848 (accessed 28 March 2026).
25. Roos A. The European Union's General Data Protection Regulation (GDPR) and its implications for South African data privacy law: An evaluation of selected 'content principles'. Comp Int Law J South Afr 2020;53(3):7985. https://doi.org/10.25159/2522-3062/7985

Received 26 September 2025; accepted 2 April 2026.