# Privacy Risk in Organ-on-Chip: From Re-identification to Pseudo-Reidentification

Dmitry Prokhorenkov*,

Stelar Security Technology Law Research, Versmannstr. 4, 20457 Hamburg, Germany,
dp@stelar.de

**Abstract.** This study analyses the concept of pseudo-reidentification within the organ-on-chip (OoC) domain. A rapid Rodgers' concept analysis method was utilised to analyse academic studies published between 2016 and 2026, identified through a systematic database search. The findings support the development and identification of the context, definitions, *antecedents*, *attributes*, and *consequences* of the concept under study for OoC. The analysis of the selected studies provides enhanced conceptual clarity and introduces a conceptual model for pseudo-reidentification. The study also highlights the significance of privacy risks in the OoC domain by facilitating discussion of practical, operational risk-management insights specific to reidentification and pseudo-reidentification. The implications for practitioners, researchers, and policymakers are addressed, emphasising the potential to enhance data subjects' privacy and to ensure accurate interpretation of privacy risks. This analysis establishes a foundation for future research, policy development, and risk management, supporting enhanced data subject privacy and the effective use of OoC technologies in accordance with EU policy requirements.

**Keywords:** reidentification · pseudo-reidentification · Organ-on-Chip · privacy

## 1 Introduction

Organ-on-chip (OoC) has received considerable attention and offers a promising alternative to animal testing, with the goal of developing tailored, data-driven, personalised drugs and improving personalised drug development. The OoC domain demands significant attention from academics and practitioners; however, issues related to privacy and governance risk have received limited attention and remain understudied [46]. In particular, risks associated with the potential re-identification of coded or pseudonymized donor data, patient-specific cells, tissue samples, and biometric data, together with responsibilities of data controllers and processors, require further investigation. It is critical to identify referred risks early, at every stage of the data lifecycle, for organisations and third parties engaged in research and drug development. As OoC technologies advance and contribute to enhanced quality of life and more precise drug development, novel privacy risks continue to arise.

Despite the great attention to privacy risks in the biomedical domain [2], academic and industry efforts remain relatively sparse in the context of OoC compared to other domains. However, industry standards such as ISO/IEC 27559 and ISO/IEC 31000, together with regulatory frameworks, General Data Protection Regulation (GDPR) and the EU Artificial Intelligence Act (AIA), as well as data protection impact assessments (DPIA) and fundamental rights impact assessments (FRIA), do not provide comprehensive or harmonised methods for integrating legal and technical considerations. Moreover, the processes for meeting regulatory requirements and applying relevant assessments and industry standards could not benefit from an explicit, applicable, and unified taxonomy for privacy and related risks. Yet, ongoing debates on data transfers [33] and recommendations from data protection agencies [12] related to the OoC demonstrate the need for more explicit guidance and standardisation efforts to support tailored drug development and address privacy and ethical concerns [41].

As a result, organisations, third parties, and data subjects are left to navigate goal-based regulations under conditions of uncertainty. Nevertheless, efforts to improve the efficiency of the application of ISO standards are partly guided by a substantial body of research that systematises diverse privacy attacks and threat modelling frameworks, as well as the classification of privacy risks associated with the linkage process [27] and relevant domains. Consent, data transfer, and accountability are among the most challenging privacy risks to forecast and control [1]. However, these areas remain insufficiently examined in the OoC domain. In line with industry practice, such risks are primarily addressed through governance measures such as controlled-access repositories [19], while privacy-enhancing technologies (PETs) or other techniques are considered only as complementary layers for specific data flows and use cases [7]. Additionally, the growing adoption of AI-driven methods and the utilisation of complex, rich datasets further exacerbate this gap. Practitioners and researchers utilise various methods to address sensitive privacy issues in order to comply with current regulations [7]; nevertheless, the utility and social impact of these methods remain open.

Study [16] introduces and describes pseudo-reidentification[1] as the detection of unique structural, physiological, or behavioural patterns that enable a data subject to be singled out without attaching the results to an established identity. In the current work, pseudo-reidentification is applied to wearable devices, CGM, ECG, and retinal images as examples. We argue that such a definition closely resembles potential OoC based on study outcomes [37] data involving donor-derived cells, organs and other OoC-related outputs and given the definition of pseudo-reidentification leads to the detection of individual-specific patterns in identy-deatched data that enable a data subject or that data subject's records or data to be distinguished or linked without reliably establishing the person's

[1] **Definition**: The process by which AI, or analytical methods detect unique patterns in deidentified data that suggest individuality without directly linking to a specific person, unlike traditional reidentification, which requires external identifiers, or reference datasets

identity. Likewise, the evidence from the study [18] demonstrates the importance of considering pseudo-reidentification risk alongside reidentification risks arising from individual identification via single-cell RNA sequencing, underscoring the importance of the OoC domain. Furthermore, the term pseudo-reidentification is not defined or addressed in current regulations, including the GDPR, the AIA, and the Digital Omnibus proposal. Few publications explicitly investigate pseudo-reidentification in the OoC domain. The aim of this study is to expand the evidence base and clarify the definition of pseudo-reidentification. Additionally, this research seeks to raise awareness within the academic community and to mitigate potential privacy risks for data subjects within the EU.

To address this gap, we conducted a rapid Rodgers' evolutionary concept analysis [43] through a scoping review until studies no longer yielded new knowledge on the specified core elements, as described further. This approach aimed to identify key *attributes*, *antecedents*, *consequences*, related concepts, and forms of pseudo-reidentification, and was supplemented by legal analysis and technical evidence appraisal. We contend that this approach is appropriate for several reasons: (a) the concept of pseudo-reidentification is rarely addressed in academic, legal, and technical literature; (b) it is primarily an analytical concept rather than an autonomous legal category and is context-dependent; (c) it remains conceptually unstable; (d) it may differ from reidentification, singling out, linkability, and biometric identification; and (e) it can be applied across multiple disciplines. The primary objectives are to identify the *attributes*, *antecedents*, and *consequences* of pseudo-reidentification and, where relevant, to contrast them with related concepts. This study has certain limitations consistent with the rapid concept analysis methodology **§6**. The main research questions (RQs) are as follows: **RQ1**: *What are the defining attributes, antecedents, and consequences of pseudo-reidentification*? **RQ2**: *How can pseudo-reidentification manifest in human organ-on-chip data within the EU*?

**Contributions**: This study represents the first academic attempt to apply Rodgers' evolutionary concept analysis to analyse the pseudo-reidentification concept, which is relevant to privacy risk, rather than to apply the reidentification concept alongside established privacy-related definitions. It extends the body of knowledge regarding definitions of pseudo-reidentification within the EU and includes a critical analysis and comments on relevant EU regulations. The findings are intended to support future efforts to reduce privacy risks for data subjects and to inform legal, technical, policy, and standardisation initiatives.

## 2 Background

**OoC** Organ-on-chip systems encompass various complex systems and are defined as controlled microfluidic platforms that contain human cells to emulate specific human organs [46]. OoC is also referred to as a microphysiological system or a tissue chip. These platforms partially utilise human biological material, produce sensitive personal data, and can be stored and distributed via biobanks to enhance knowledge bases for drug toxicity screening, disease modelling, and

targeted drug discovery and development. OoC technology integrates diverse scientific and technical disciplines, including computer science, mechanobiology, disease modelling, drug development, diagnostics, and therapeutics [41]. Consequently, these systems and their outputs are subject to various regulatory frameworks within the EU. In addition to unresolved ethical challenges [46], it also poses emergent privacy implications related to reidentification which are not able to be addressed [42], for data involved not limited to, health[2], genetic[3], and other special category[4] data. To address these risks, practitioners employ a range of governance instruments, such as controlled access [19], data use agreements (DUAs), business associate agreements (BAAs), data processing agreements (DPAs), and other processor agreements. Furthermore, as expressed [29], the use of state-of-the-art privacy-preserving methods, such as FL, DP, TEE, and encryption in healthcare, is fragmented and depends on the specific application and context; nevertheless, the whole biomedical domain is subject to pseudo-reidentification implications and is recognised as an emerging attack vector.

Practitioners align their practices with relevant industry standards (ISO/IEC [5]). However, applying these standards is complex and demanding [26] [38] due to rapid technological advancement and limited knowledge transfer from academia to industry. Similarly, the impact of privacy risk and its context-dependent nature require early identification [28], [48] to respond to emerging risks adequately. Furthermore, due to the complexity and multi-case nature of the data involved in producing in the OoC domain, previously reidentification risk demands more specific evaluation against the recently introduced pseudo-reidentification notation [16], which highlights and refocuses aspects of privacy risk. Pseudo-reidentification is significant in the OoC domain, as it enables the identification of unique data patterns that, while not directly linked to individuals, could allow future linkage to external datasets. Unlike traditional reidentification, pseudo-reidentification enables the detection of unique patterns in anonymised datasets, often enhanced by AI, making it highly relevant to clinical data, the OoC domain, and personalised drug development.

**EU regulations** Given the complexity of the data generated, utilised, and transferred within the OoC domain, stakeholders are required to meet multiple legal, technical, and organisational obligations to ensure compliance with EU regulations. These obligations stem from binding EU legislation and are further clarified through regulatory opinions, guidelines, and technical recommendations, including [11], [13], and [10].

Current EU legislative frameworks, along with opinions and recommendations from data protection agencies [11], [13], [10], and the Digital Omnibus proposal, are intended to enhance transparency for data controllers and pro-

[2] GDPR, Article 4(15)
[3] GDPR, Article 4(13)
[4] GDPR, Article 9(1)
[5] 31000:2018, 27559

cessors and to clarify the relationship between the AIA and the GDPR. Relevant legislative proposals, such as the Digital Omnibus proposal draft, aim to enhance regulatory clarity for data controllers, processors, and third parties. These proposals focus on the interaction between the AIA and the GDPR, with particular emphasis on the definition of personal data. Within the regulatory context and potential changes to the definition of personal data, a central issue is whether data used or shared within the OoC domain can satisfy the notion of full anonymisation or remain personal data subject to the GDPR. Under the GDPR, data is considered anonymous if it does not relate to an identified or identifiable natural person [11]. The EDPB specifies three criteria for assessing anonymisation: *No Record Isolation*, *No Linkage*, and *No Inference*. These criteria are intended to "*assess the effectiveness of possible (re-)identification techniques*", which is likely practical for the OoC domain.

Biomedical data often contain or reveal personal and sensitive information, which can facilitate the identification of individuals [23]. The degree of identifiability depends on factors such as data granularity, sample size, the availability of auxiliary datasets, and the technical means reasonably available to potential recipients or adversaries. Furthermore, as reported in [9], maintaining the full utility of the original data while mitigating privacy risks remains an unresolved challenge. Similarly, as explained in [42], a standardised and generally applicable re-identification risk assessment framework for guiding data processors, controllers, and researchers, particularly for genetic data, is not yet been established. This limitation is especially significant for OoC research, where highly granular, multimodal, or longitudinal biomedical data may need to retain substantial scientific utility.

Even when such data are coded or pseudonymised, linking them with other datasets can enable the re-identification of individuals, including those represented in small samples, due to advances in data linkage, computational inference, genomic matching, and the increasing availability of auxiliary data [32]. Therefore, while pseudonymisation and coding can mitigate the risk of reidentification, they alone do not demonstrate that OoC data have been anonymised in accordance with the requirements of the GDPR.

**Related work** None of the studies reports or explores the impact of pseudoreindefication on the OoC domain. *Note: At the time, current research was conducted.* The OoC domain is characterised as interdisciplinary [44] and recognised for its implications in meeting legislative requirements, except when applied exclusively for research purposes. Several attempts are devoted mainly to studying ethical aspects and utility trade-offs, as well as anonymisation of data and ways to assess privacy risk through established assessments [45]. Further, we discuss recent academic contributions to provide an overview.

The primary body of knowledge within the OoC domain emphasises [46], [41] ethical considerations, such as informed consent and data ownership, and highlights the importance of data protection and privacy. Also, limited attention to the key challenges, as a) obtaining the personalised health data, which

includes ethical and consent issues, together with b) utilising personal data for the personalised drug development by merging our outputs with different labs or sharing datasets, results in a key challenge for the successful implementation of OoC technology. Several studies [49], [7] highlight the importance of anonymisation issues within the biomedical domain and in PETs, which pose a high risk of reidentification and entail utility trade-offs.

Subsequent studies [22], [20], [17], [15], [40], [30] related to the (pseudo)reidentification primarily investigate its implications within the biomedical domain. Studies use terms such as *identifiability*, *linkage*, *membership inference*, or *reidentification*, but they only partly lay the groundwork for emerging validation of whether the pseudonymised genomes or unique features remain distinguishable and subject to fall under the pseudo-reidentification concept. Likewise, a review [29] underscores the significance of privacy implications and identifies pseudo-reidentification as an emerging attack vector.

## 3 Methods

### 3.1 Concept analysis method

This study adopts Rodgers' evolutionary concept analysis [43], which builds on Wilson's approach to concept analysis [47] and emphasises the dynamic and context-dependent nature of concepts. Rodgers' method examines how a concept's meaning, *attributes*, *antecedents*, *consequences*, and related terms evolve across contexts and disciplines [*for example, comparing technical, legal, and policy interpretations of pseudo-reidentification in OoC*]. Although several similar approaches to concept analysis are available [?], including the most closely related approach [47], we consider Rodgers' method the most appropriate for three reasons. *First*, the literature across the technical, legal, and policy domains is limited and does not explicitly address pseudo-reidentification. *Second*, the terminology and associated concepts are likely to change and expand over time. *Third*, goal-based regulations increasingly require data processors and other entities handling sensitive data to assess and validate reidentification risks or comparable privacy-related concepts.

Within Rodgers' approach, the significance of the selected concept is a central consideration. In the OoC domain, pseudo-reidentification represents a purposeful analytical concept because it helps identify and mitigate privacy risks [*for example, inferring an individual's identity by combining contextual data across environments*]. It also provides an adequate and reliable characterisation of the phenomenon and its associated privacy implications. Defining the concept more precisely enables us to describe pseudo-reidentification in the OoC domain and identify the characteristics it assumes within different disciplines [*for example, technical identifiability thresholds may differ from legally recognised reidentification standards*].

The selected approach also enables the identification of interdisciplinary differences in how the concept is understood and applied, while incorporating

perspectives from the technical, legal, and policy domains. We therefore argue that Rodgers' evolutionary concept analysis is well-suited to examining pseudo-reidentification risks in the OoC domain, particularly because it requires insights from and coordination among these interconnected disciplines.

Rodgers' method is an iterative process consisting of six main steps that yield more contextual interpretation than a permanent definition, as follows. **1**. Specify a concept, purpose, surrogate terms, and related concepts. **2**. Determine and select the representative data sources. **3**. Collection of relevant data within the scope of concepts, attributes, and conceptual basis, including interdependent variations. **4**. analysis concept based on characteristics, attributes **5**. provisional example that satisfies the concept, if available. **6**. Formulate a contextual definition, hypotheses, and directions for the concept's future development.

### 3.2 Sources of data

The scope of domain studies includes topics related to the OoC, privacy risk, reidentification, and EU regulations. Primary search was conducted over the *ACM Digital Library*, *IEEE Explore Digital Library*, *SCOPUS*, *Web of Science*, *Scopus*, and *SpringerLink* using the search query, using the AND , OR operators as presented in **Fig. 1.** Guided by our inclusion criteria, we aim to include any full-length academic study that a) is written in English, b) covers topics related to privacy, reidentification, EU legislation and OOC, and c) utilises sensitive or personal data and is a primary study. Our primary exclusion criteria are studies that don't address the research questions, duplicate titles, articles published as position papers or posters, and those not written in English. To ensure reliability in the filtering and selection process, a literature management platform, such as Rayyan [24], was utilised. To assess relevance by reviewing the title and abstract, both researchers conducted a pilot study (blindly) comprising 10% of the original corpus (n=17). For this review, the agreement is predominantly plausible, as indicated by Cohen's kappa (0.57). Subsequently, one researcher reviewed the final subset of studies and consulted the second researcher only when eligibility was uncertain. A total of **169** studies were considered for eligibility, and **14** original studies were included in the final analysis.

### 3.3 Data analysis process

Selected studies were evaluated depending on their validity and quality, until the conducted literature search enabled to fullfil the properties of the concepts extracted from **14** studies. Data collection and analysis were conducted simultaneously, and given the topics involved, OoC-related studies were analysed first, followed by other domains, privacy, policies, and computer science, including other academic studies. The analysis results were then used to refine the definitions of the concepts. Consensus and disagreement regarding these concepts across disciplines were examined and updated. Recurring subjects identified during the analysis were organised into the categories of "*Attributes*," "*Antecedents*," and "*Consequences*".

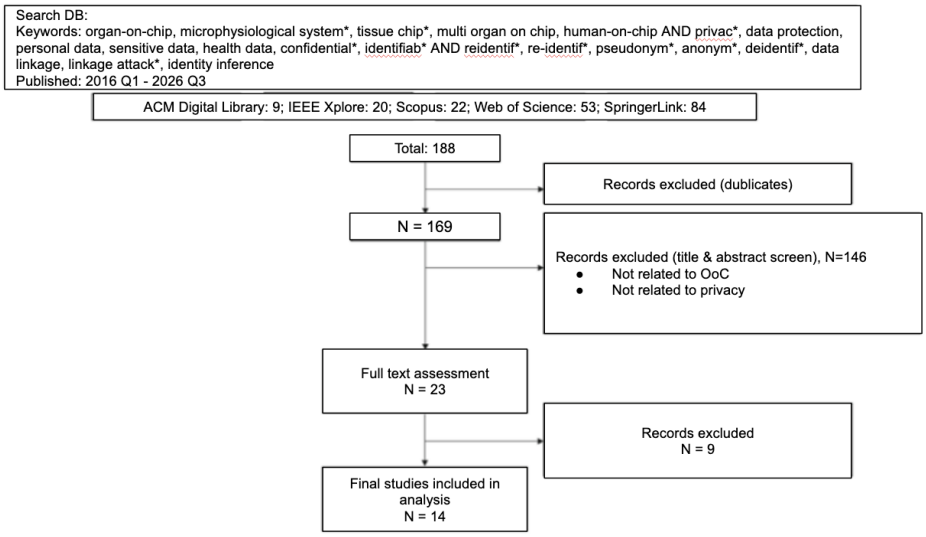


**Fig. 1.** Overview of the study selection process *Note: Search was adapted to the syntax and indexing requirements of each DB*

## 4 Results

This section demonstrates the results, with individual subsections addressing *Attributes*, *Antecedents*, and *Consequences* (**Table. 1**). The study's demographics are shown in **Fig. 3a**, while **Fig. 3b** illustrates the domain distributions of the included studies. Also, concepts are understood as dynamic and context-dependent within Rodgers' framework [43]; the specified *antecedents*, *attributes*, and *consequences* further describe provisional conditions associated with pseudo-reidentification across the OoC, privacy, technical, and regulatory domains (**Fig.2**).

### 4.1 OoC Antecedents

According to the Rodgers concept analysis [43], *antecedents* are a) events, b) conditions or c) circumstances that lead to the emergence of the concept. For pseudo-reidentification, it could be distributed across the technical, informational, organisational and contextual conditions that enable drawing conclusions, narrowing the search, or establishing a probabilistic connection for a specific data subject, likely without any confirmation of the data subject's identity. In the OoC domain, it could occur across various actions related to biological samples, OoC chips, experimental outputs, donor data, clinical records, outputs from AI models, datasets and contextual metadata, processed together or linkable across different systems or third parties. In other words, **antecedents** make pseudo-reidentification possible, whereas *attributes* demonstrate that pseudo-reidentification has occurred.

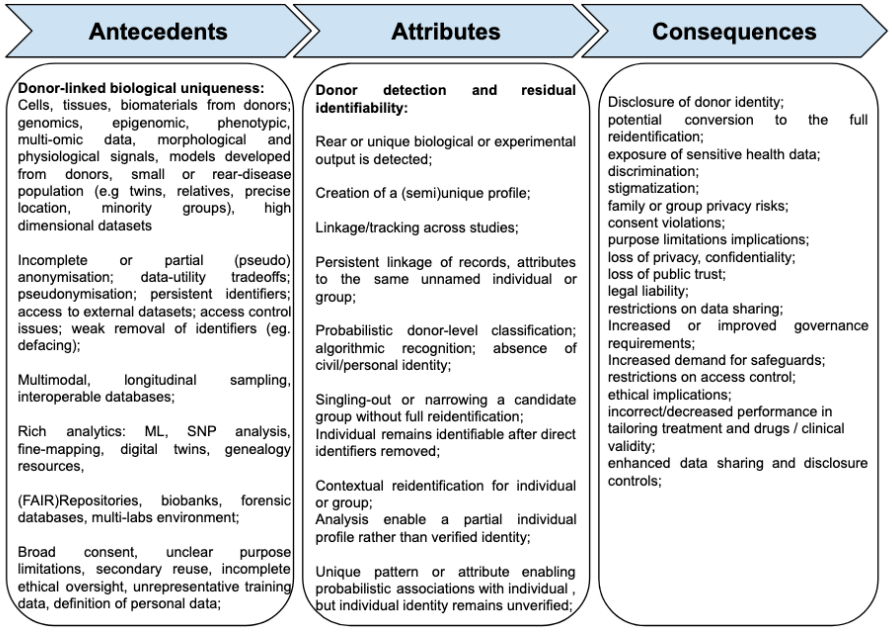


**Fig. 2.** Conceptual model of pseudo-reidentificaion in OoC *Note: Conceptual representation provides a preliminary model rather than a validated one. The principal contribution is to demonstrate that identifying donor identity or individuality without confirmed identity is both possible and practical*

Pseudo-reidentification in the OoC domain remained understudied and unknown, and reidentification has also attracted limited attention. Nevertheless, analysis of the relevant concepts indicates that the phenomenon under investigation is facilitated by donor-linked biological materials [P01, P05, P08-P10, P12-P14], high-dimensional genomic and multi-omics data [P01, P04, P05, P08, P09, P10, P13, P14], small, rare, or unique populations [P01, P03, P07, P08, P14], longitudinal and interoperable datasets [P01, P03, P07, P11, P14], insufficient de-identification [P01, P02, P03, P07, P08] and linkage to external resources [P03, P06, P08-P10, P12]. These conditions are further influenced by the utilisation of AI [P01, P07, P09, P11, P13, P14], (ii) SNP, fine-mapping [P04, P05,P06, P08–P10, P12], and digital-twin modelling [P01, P13, P14]. Notably, we have identified an ongoing EU project that directly addresses a close connection to the concept under study [P03].

### 4.2 Attributes

Following Rodgers' concept analysis [43], *attributes* are defined as the recurring characteristics that form a concept and distinguish it from related concepts. Thus, selecting key attributes is essential for clarifying and understanding the

**Table 1.** Mapping of evidence identified in the included studies

| Study ID | Ref. | Year | Mapping | Notes |
|---|---|---|---|---|
| P01 | [46] | 2025 | A, At, C | *?* |
| P02 | [36] | 2023 | A, C | *?* |
| P03 | [4] | 2026 | A, At, C | *?* |
| P04 | [21] | 2024 | A | *?* |
| P05 | [3] | 2026 | A, At | *?* |
| P06 | [34] | 2017 | At | *?* |
| P07 | [35] | 2022 | A, At, C | *?* |
| P08 | [39] | 2026 | A, At, C | *?* |
| P09 | [8] | 2026 | A, At, C | *?* |
| P10 | [25] | 2025 | A, At, C | *?* |
| P11 | [6] | 2025 | A, C | *?* |
| P12 | [5] | 2023 | A, C | *?* |
| P13 | [14] | 2026 | A, At, C | *?* |
| P14 | [31] | 2020 | A, At, C | *?* |

*Note:* A = antecedents; At = attributes; C = consequences;

concepts studied and distinguishing them from one another. In other words, antecedents make pseudo-reidentification possible, whereas **attributes** demonstrate that pseudo-reidentification has occurred. For the OoC domain, pseudo-reidentification attributes are: reidentification, inferential attribution, residual identifiability, information linkage, partial reidentification, contextual dependencies, and

As recodred, defining attributes are detection of a distinctive biological or experimental patterns [P05, P06, P09,P10, P11, P14]; result of a partial donor profile [P07, P09,P10, P12,P13,P14]; persistent cross-record or cross-study linkability [P01, P03, P07-P10, P12]; singling out or candidate-set narrowing [P06-P10, P12]; and probabilistic donor attribution while identity remains unverified [P06, P09, P10, P12]. Also, inference is context conditional [P07, P09, P13] and may be inaccurate, varying depending on supporting data [P06, P07, P09, P10, P12], analytical capabilities [P05, P06, P07, P09–P11, P13], and validation quality [P02, P06, P09–P11, P13].

### 4.3 Consequences

Consequences [43] are the events, situations, or outcomes that follow the occurrence of a concept. Consequences are not limited to direct harm and may include individual, social, organisational, legal, scientific, policy and governance-related effects. In the OoC domain, the consequences of pseudo-reidentification arise from the creation, aggregation, or utilisation of an unproven but potentially plausible connection between biological or experimental data, a dataset, model output, tissue chip output and a specific individual.

Consequences include loss of practical anonymity [P01, P07 - P10, P12, P13]; inference or disclosure of sensitive health data [P01, P07, P09, P10, P12, P13]; family or group exposure [P08, P09, P10, P12, P13]; discrimination/ stigmatisation [P01, P07- P10, P12- P13]; consent and purpose-limitation circumstances [P01, P03, P08, P09, P10, P12 - P14]; loss of trust [P01, P11, P13]; and likely escalation to confirmed reidentification [P01, P07 - P10, P12, P13]. These results may demand more robust governance and data-sharing controls [P01 - P03, P08 - P10, P12 - P14]. False attribution or privacy modifications that remove clinically relevant information may also undermine r&d/academic reproducibility or clinical validity [P02, P06, P09 - P11, P13].

## 5 Discussion

This study aimed to evaluate the applicability of the pseudo-reidentification concept within the OoC domain by identifying specific aspects and considering the analysis to provide a tailored provisional definition. The analysis further aims to demonstrate the potential relevance of the pseudo-reidentification concept in the OoC domain, stressing its role in enhancing donor privacy and facilitating the development of personalised drugs and treatments through the use of sensitive and granular data. A rapid Rodgers concept analysis, aligned with the SLR process, was utilised as the primary analytical approach. Utilising the chosen concept analysis indicates the practical value of analysing novel concepts in technology-driven domains lacking established taxonomies. Furthermore, the findings demonstrate the applicability of concept analysis within the OoC domain, which remains in clinical or research phases. These results were obtained with acknowledged limitations.

**Tailored definition**: *The process by which AI or added analytical methods/models detect and incorporate residual donor-specific biological, experimental, or contextual patterns in deidentified OoC data, enabling a donor to be singled out, the group of possible donors to be narrowed, or a probabilistic or partial attribution to be made without establishing the donor's confirmed identity.*

## 6 Limitations

Analysis limitations consist of a few aspects. The concept under study has limited academic representation across domains; the selected studies show limited engagement with pseudo-reidentification. Thus, various variables should be considered, sorted and excluded during the process. Search terms and database selection may impose limitations on the study conducted. Furthermore, the study design could be enhanced by incorporating additional forward and backward searches. Evaluating surrogate or potentially equivalent terms for pseudo-reidentification and reidentification may facilitate a more detailed analysis and a clearer definition of pseudo-reidentification.

## 7 Conclusion

This study is a systematic analysis of the pseudo-reidentification concept in the OoC domain and delivers an updated definition of the concept under study. The main goal of this study and analysis conducted is to supply a conceptual preliminary model for pseudo-reidenticiation, together with outlining the future research directions and providing advanced knowledge for the OoC application and further application of diverse technologies to provide insights from granular and sensitive data, preserving data subjects' privacy, consent and fulfilling current and future regulations. We argue that clarity and utilisation of the updated definition will facilitate the implementation of OoC technologies and, where necessary, personalised drug development, thereby enabling a transition from the research environment to subsequent stages. Furthermore, this analysis is intended to inform legal, technical, and policy experts in addressing emerging challenges related to enhanced privacy for donors and individuals. By identifying the antecedents, attributes, and consequences, this study seeks to enhance conceptual clarity and offer a supplemented model.

Advancing the integration of OoC technology into biomedical, clinical, and health systems requires substantial interdisciplinary collaboration. The conceptual evidence provided in this study is intended to inform future developments in both the OoC and privacy domains.

## 8 Acknowledgments


Suppressed Due to Excessive Length

AI-based tools and artifacts with language assistance was utilized exclusively to improve grammar, punctuation, and readability.

**Declaration of interests** The authors report there are no competing interests to declare.

## References

1. Alver, C.G., Drabbe, E., Ishahak, M., Agarwal, A.: Roadblocks confronting widespread dissemination and deployment of organs on chips. Nature communications **15**(1), 5118 (2024)
2. Backes, M., Berrang, P., Hecksteden, A., Humbert, M., Keller, A., Meyer, T.: Privacy in epigenetics: Temporal linkability of {MicroRNA} expression profiles. In: 25th USENIX security symposium (USENIX Security 16). pp. 1223–1240 (2016)
3. Bates, S.A., Budowle, B., Johnson, M., Ge, J., Mittelman, K., Mittelman, D.: Large-scale analysis of dna quantification metrics and snp sequencing performance in unidentified human remains. Forensic Science International: Genetics p. 103470 (2026)

4. Beyer, M., Kuznetsov, D., Kühn, S.: The german twin registry (gertrud): Towards harmonized, multisource twin research. Twin Research and Human Genetics pp. 1–11 (2026)
5. Budowle, B., Sajantila, A.: Revisiting informed consent in forensic genomics in light of current technologies and the times. International Journal of Legal Medicine **137**(2), 551–565 (2023)
6. Caminero, A., Tropini, C., Valles-Colomer, M., Shung, D.L., Gibbons, S.M., Surette, M.G., Sokol, H., Tomeo, N.J., Tarr, P.I., et al.: Credible inferences in microbiome research: ensuring rigour, reproducibility and relevance in the era of ai. Nature Reviews Gastroenterology & Hepatology **22**(11), 788–803 (2025)
7. Cho, H., Froelicher, D., Dokmai, N., Nandi, A., Sadhuka, S., Hong, M.M., Berger, B.: Privacy-enhancing technologies in biomedical data science. Annual review of biomedical data science **7**(1), 317–343 (2024)
8. El Khair, A.: Dna profiling in forensic genetics and anthropogenetics: Techniques, databases, and ethical considerations. Forensic Science, Medicine and Pathology pp. 1–13 (2026)
9. European Data Protection Board: EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, https://www.edpb.europa.eu/system/files/documents/files/file1/edpb_replyec_questionnaireresearch_final.pdf
10. European Data Protection Board: Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, https://www.edpb.europa.eu/system/files/documents/2024-12/edpb_opinion_202428_ai-models_en.pdf
11. European Data Protection Board: Guidelines 02/2026 on Anonymisation. Guidelines 02/2026, European Data Protection Board (2026), https://www.edpb.europa.eu/system/files/2026-07/edpb_guidelines_202602_anonymisation_v1_en_0.pdf
12. European Data Protection Board and European Data Protection Supervisor: Joint Opinion 3/2026 on the Proposal for a European Biotech Act. Joint Opinion 3/2026, European Data Protection Board and European Data Protection Supervisor (Mar 2026), https://www.edpb.europa.eu/system/files/2026-03/edpb_edps_jointopinion_202603_biotechact_en.pdf
13. European Union Agency for Cybersecurity: Deploying Pseudonymisation Techniques: The Case of the Health Sector, https://www.enisa.europa.eu/sites/default/files/publications/WP2021%20-%20O.2.3%20Pseudonymisation%20Healthcare%20.pdf
14. Gopukumar, S., Dwivedi, D., Rahamathulla, M., Ahmed, M.M., Soni, T.K., Natarajan, P.G., Ramkanth, S., Mitra, A., Das, U.: Multiscale predictive cellular modeling: integrating hypothesis grammars, digital twins, and multi-omics for in silico oncology and precision theranostics. Functional & Integrative Genomics **26**(1), 113 (2026)
15. Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y.: Identifying personal genomes by surname inference. Science **339**(6117), 321–324 (2013)
16. Hallaj, S., Heinke, A., Kalaw, F.G.P., Gim, N., Blazes, M., Owen, J., Dysinger, E., Benton, E.S., Cordier, B.A., Evans, N.G., et al.: Navigating open data sharing and privacy in the age of clinical ai research: from reidentification to pseudo-reidentification. EClinicalMedicine **91** (2026)

17. Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., Craig, D.W.: Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. PLoS genetics **4**(8), e1000167 (2008)
18. Kulhankova, L., Montiel González, D., Bindels, E., Kling, D., Kayser, M., Mulugeta, E.: Single-cell transcriptome sequencing allows genetic separation, characterization and identification of individuals in multi-person biological mixtures. Communications Biology **6**(1), 201 (2023)
19. Lin, D., McAuliffe, M., Pruitt, K.D., Gururaj, A., Melchior, C., Schmitt, C., Wright, S.N.: Biomedical data repository concepts and management principles. Scientific Data **11**(1), 622 (2024)
20. Lin, Z., Owen, A.B., Altman, R.B.: Genomic research and human subject privacy (2004)
21. Lincoln, M.R., Connally, N., Axisa, P.P., Gasperi, C., Mitrovic, M., van Heel, D., Wijmenga, C., Withoff, S., Jonkers, I.H., Padyukov, L., et al.: Genetic mapping across autoimmune diseases reveals shared associations and mechanisms. Nature genetics **56**(5), 838–845 (2024)
22. Malin, B., Sweeney, L.: How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. Journal of biomedical informatics **37**(3), 179–192 (2004)
23. Molnár-Gábor, F., Korbel, J.O.: Genomic data sharing in europe is stumbling—could a code of conduct prevent its fall? EMBO molecular medicine **12**(3), EMMM201911421 (2020)
24. Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A.: Rayyan—a web and mobile app for systematic reviews. Systematic reviews **5**(1), 210 (2016)
25. Pedroza Matute, S., Iyavoo, S.: Implementation of ngs and snp microarrays in routine forensic practice: opportunities and barriers. BMC genomics **26**(1), 541 (2025)
26. Polemi, N., Praça, I., Kioskli, K., Bécue, A.: Challenges and efforts in managing ai trustworthiness risks: a state of knowledge. Frontiers in big Data **7**, 1381163 (2024)
27. Powar, J., Beresford, A.R.: Sok: Managing risks of linkage attacks on data privacy. Proceedings on Privacy Enhancing Technologies (2023)
28. Prasser, F., Kohlmayer, F., Kuhn, K.A.: The importance of context: Risk-based de-identification of biomedical data. Methods of information in medicine **55**(04), 347–355 (2016)
29. Rahimian, M.A., Panny, B., Joshi, J.B.: Privacy at scale in networked healthcare. In: 2025 IEEE 7th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA). pp. 551–560. IEEE (2025)
30. Raisaro, J.L., Tramer, F., Ji, Z., Bu, D., Zhao, Y., Carey, K., Lloyd, D., Sofia, H., Baker, D., Flicek, P., et al.: Addressing beacon re-identification attacks: quantification and mitigation of privacy risks. Journal of the American Medical Informatics Association **24**(4), 799–805 (2017)
31. Rajewsky, N., Almouzni, G., Gorski, S.A., Aerts, S., Amit, I., Bertero, M.G., Bock, C., Bredenoord, A.L., Cavalli, G., Chiocca, S., et al.: Lifetime and improving european healthcare through cell-based interceptive medicine. Nature **587**(7834), 377–386 (2020)
32. Rocher, L., Hendrickx, J.M., De Montjoye, Y.A.: Estimating the success of re-identifications in incomplete datasets using generative models. Nature communications **10**(1), 3069 (2019)

33. Royo, R., Garcia, E., Schlünder, I., Hilmarsen, C., Thomsen, M.K., Cauët, E., Jensen, S.K., Giardina, S., Hebrant, A., Raicevic Toungouz, G., et al.: Genomic data sharing in research across europe: legal challenges and upcoming opportunities within the european health data space. European Journal of Public Health **35**(Supplement_3), iii25–iii31 (2025)
34. Ruegg, K.C., Anderson, E.C., Harrigan, R.J., Paxton, K.L., Kelly, J.F., Moore, F., Smith, T.B.: Genetic assignment with isotopes and habitat suitability (gaiah), a migratory bird case study. Methods in Ecology and Evolution **8**(10), 1241–1252 (2017)
35. Saglam, R.B., Nurse, J.R., Hodges, D.: Personal information: Perceptions, types and evolution. Journal of Information Security and Applications **66**, 103163 (2022)
36. Sahlsten, J., Wahid, K.A., Glerean, E., Jaskari, J., Naser, M.A., He, R., Kann, B.H., Mäkitie, A., Fuller, C.D., Kaski, K.: Segmentation stability of human head and neck cancer medical images for radiotherapy applications under de-identification conditions: Benchmarking data sharing and artificial intelligence use-cases. Frontiers in oncology **13**, 1120392 (2023)
37. Schiff, L., Migliori, B., Chen, Y., Carter, D., Bonilla, C., Hall, J., Fan, M., Tam, E., Ahadi, S., Fischbacher, B., et al.: Integrating deep learning and unbiased automated high-content screening to identify complex disease signatures in human fibroblasts. Nature communications **13**(1), 1590 (2022)
38. Schuett, J.: Risk management in the artificial intelligence act. European Journal of Risk Regulation **15**(2), 367–385 (2024)
39. Serena, A., Alessandra, D., Elena, C., Alessia, R., Paolo, G., Carlo, R.: Assessing ethical standards in forensic population genetics: A systematic review of forensic science international: Genetics and forensic science international: Reports publications. Forensic Science International: Genetics p. 103515 (2026)
40. Shringarpure, S.S., Bustamante, C.D.: Privacy risks from genomic data-sharing beacons. The American Journal of Human Genetics **97**(5), 631–646 (2015)
41. Thakar, R.G., Fenton, K.N.: Bioethical implications of organ-on-a-chip on modernizing drug development. Artificial organs **47**(10), 1553–1558 (2023)
42. Thomas, M., Mackes, N., Preuss-Dodhy, A., Wieland, T., Bundschus, M.: Assessing privacy vulnerabilities in genetic data sets: scoping review. JMIR Bioinformatics and Biotechnology **5**(1), e54332 (2024)
43. Tofthagen, R., Fagerstrøm, L.M.: Rodgers' evolutionary concept analysis–a valid method for developing knowledge in nursing science. Scandinavian journal of caring sciences **24**, 21–31 (2010)
44. Van Den Berg, A., Mummery, C.L., Passier, R., Van der Meer, A.D.: Personalised organs-on-chips: functional testing for precision medicine. Lab on a Chip **19**(2), 198–205 (2019)
45. Wairimu, S., Iwaya, L.H., Fritsch, L., Lindskog, S.: On the evaluation of privacy impact assessment and privacy risk assessment methodologies: A systematic literature review. IEEE Access **12**, 19625–19650 (2024)
46. Weidema, J., De Vries, M., Mummery, C., De Graeff, N.: The ethical aspects of human organ-on-chip models: a mapping review. Stem Cell Reports (2025)
47. Wilson, J.: Thinking with concepts. Cambridge University Press (1970)
48. Yuan, Z.S., Singh, M.M.: Privacy risks in health big data: a systematic literature review. arXiv preprint arXiv:2502.03811 (2025)
49. Zuo, Z., Watson, M., Budgen, D., Hall, R., Kennelly, C., Al Moubayed, N.: Data anonymization for pervasive health care: systematic literature mapping study. JMIR medical informatics **9**(10), e29871 (2021)