Saudi Journal of Medicine and Public Health

https://saudijmph.com/index.php/pub

https://doi.org/10.64483/202412467

A Scoping Review of the Ethical, Legal, and Technical Dimensions of Privacy in Big Data Health Research

Reem Menwer Owaid Alrashdi [(1)] , Reem Munawar Awad Al-Rashdi [(2)] , Salihah Abdullah Saeed Alghamdi [(3)] , Khuluod Ali Mohammed Rezgallah [(3)] , Abdulaziz Ali Abdulaziz Alghaythar [(4)] , Faisal Fahad Mohammed Alshammari [(5)] , Abdullah Jaber Eissa Faqihi [(6)] , Dhaifallah Mohammed Dhaifallah Moraya [(7)] , Ahlam Abdullah Ibrahim Aqeel [(8)] , Muath Mohammed Dhaifallah Moraya [(8)] , Khloud Masead Dhaif Allah Al-Mutairi [(9)] , Nasser Nashi Alshaibani [(10)] , Khaled Ibrahim Muhammad Mobaraki [(11)] , Mohammed Saleh Abdulkareem Al Juma, [(12)] , Sarah Ahmed Arif [(8)]

(1) Medical Records Technician, Saudi Health Center,Ministry of Health, Saudi Arabia,
(2) The First Health Cluster In Riyadh,Ministry of Health, Saudi Arabia,
(3) Al Imam Abdul Rahman Al Faisal Hospital,Ministry of Health, Saudi Arabia,
(4) Primary Health Care Center In Aldar Albedh2,Ministry of Health, Saudi Arabia,
(5) Eradah Complex And Mental Health-Hail,Ministry of Health, Saudi Arabia,
(6) Javan Eradh And Mental Health Hospital,Ministry of Health, Saudi Arabia,
(7) Jazan Mental Health Hospital , Ministry of Health, Saudi Arabia,
(8) Ministry Of Health, Saudi Arabia,
(9) Specialized Dental Center In Riyadh,Ministry of Health, Saudi Arabia,
(10) Sanitah General Hospital,Ministry of Health, Saudi Arabia,
(11) National Guard Health Affairs,Ministry of Health, Saudi Arabia,
(12) Mohammed Bin Abdulaziz Hospital,Ministry of Health, Saudi Arabia

Abstract

Background: The proliferation of big data in health research—encompassing genomic datasets, electronic health records (EHRs), wearables, and multi-omics—offers unprecedented potential for scientific discovery and personalized medicine. However, this data-driven paradigm poses profound and novel challenges to the privacy of individuals, demanding an integrated analysis of ethical, legal, and technical safeguards. **Aim:** This scoping review synthesizes contemporary literature (2015-2024) to map the ethical dilemmas, legal frameworks, and technical solutions concerning privacy in big data health research. **Methods:** A systematic search was conducted across PubMed, IEEE Xplore, Scopus, and Google Scholar. Literature was thematically analyzed to identify key themes, tensions, and emergent strategies across the three dimensions. **Results:** The review identifies a core tension between data utility for the public good and individual privacy rights. Ethically, key issues include re-identification risk, informed consent for future unspecified research, and algorithmic bias. Legally, a fragmented global landscape exists, with regulations like the GDPR providing strong protections but creating compliance complexity. Technically, privacy-enhancing technologies (PETs) such as federated learning, differential privacy, and homomorphic encryption offer promising, yet imperfect, solutions. **Conclusion:** Effective privacy preservation in big data health research requires a harmonized, interdisciplinary approach. A robust governance framework must interweave ethical principles, adaptable legal compliance, and state-of-the-art technical controls, foster public trust while enabling responsible innovation.

Keywords: Big Data; Health Research; Privacy; Ethics; Data Governance; Privacy-Enhancing Technologies

Introduction

The 21st-century healthcare landscape is undergoing a seismic shift driven by the advent of big data. This term refers to the massive volumes of high-velocity, complex, and variable digital information generated from diverse sources, including genomic sequencing, electronic health records (EHRs), insurance claims, wearable biosensors, and patient-generated health data (PGHD) (Raghupathi & Raghupathi, 2014). The convergence of these datasets—often termed "multi-omics" when combined with proteomic, metabolomic, and microbiomic data—promises to revolutionize biomedical research, enabling more powerful population health studies, the discovery of novel biomarkers, and the realization of truly personalized, predictive medicine (Dinov, 2016). However, this paradigm of data-intensive science brings with it a constellation of unprecedented privacy challenges that threaten to undermine public trust and stifle innovation if not adequately addressed (Vayena et al., 2017).

*Corresponding author e-mail: ralrashd@moh.gov.sa (Reem Menwer Owaid Alrashdi).
Receive Date: 19 December 2024, Revise Date: 30 December 2024, Accept Date: 31 December 2024

Privacy, in the context of health information, is a multifaceted concept encompassing the right of an individual to control the collection, use, and disclosure of their personal health data. In big data health research, this traditional conception is severely tested. Data is often aggregated, shared across institutional and national boundaries, and subjected to advanced analytics like machine learning, which can infer sensitive information not directly provided, such as disease risk, behavioral patterns, or even genetic predispositions of relatives (Rocher et al., 2019). The scale and complexity of these operations render traditional privacy models, such as simple de-identification, increasingly obsolete, as re-identification attacks leveraging auxiliary data sources become more sophisticated (Jiang et al., 2022).

The privacy challenge is not monolithic but exists at the intersection of three critical and interdependent dimensions: ethical, legal, and technical. Ethically, researchers grapple with foundational questions about autonomy, beneficence, and justice. How can meaningful informed consent be obtained for research whose future uses are unknown? Does the public good of research outweigh individual privacy risks, and who decides? How do we prevent big data analytics from perpetuating or exacerbating existing health disparities through algorithmic bias (Mittelstadt & Floridi, 2016)? Legally, a patchwork of regulations, from the European Union's General Data Protection Regulation (GDPR) to the United States' Health Insurance Portability and Accountability Act (HIPAA) and various national laws, creates a complex compliance landscape for international consortia. These laws often struggle to keep pace with technological change, creating ambiguities around the status of data types like genomic information or AI-derived inferences (Rezaeikhonakdar, 2023).

Technically, the field is responding with a suite of Privacy-Enhancing Technologies (PETs). These include cryptographic methods like homomorphic encryption (which allows computation on encrypted data), statistical techniques like differential privacy (which adds mathematical noise to query outputs), and architectural paradigms like federated learning (where algorithms are shared to the data, not data to a central server) (Xu et al., 2021). While promising, these tools often involve trade-offs between privacy, data utility, and computational cost, and their implementation requires significant expertise.

This scoping review aims to systematically map the contemporary discourse surrounding these ethical, legal, and technical dimensions of privacy in big data health research. By synthesizing literature across disciplines—bioethics, health law, computer science, and biomedical informatics—it seeks to elucidate the core tensions, evaluate current solutions, and identify gaps where interdisciplinary collaboration is urgently needed. The central argument is that siloed approaches are insufficient; a holistic governance framework that dynamically integrates ethical foresight, legal agility, and technical rigor is essential for sustaining the social license for big data health research.

Methodology

This study employed a scoping review methodology, which is ideally suited for mapping key concepts, evidence, and gaps in a complex, heterogeneous, and rapidly evolving field (Arksey & O'Malley, 2005). The objective was not to conduct a systematic meta-analysis of efficacy but to provide a broad, interdisciplinary synthesis of the landscape.

Search Strategy and Information Sources

To ensure a comprehensive and interdisciplinary capture of the literature, a systematic electronic search was conducted across four major databases selected for their complementary disciplinary strengths. PubMed was utilized to target the core biomedical and clinical research perspective. IEEE Xplore was searched to encompass the technical and engineering literature on privacy-enhancing technologies and security architectures. Scopus provided a broad, multidisciplinary coverage of journals across social sciences, law, and computer science. Finally, Google Scholar was included to capture influential grey literature, preprints, and seminal works that might not be indexed in the other databases. The search was temporally bounded from January 2015 to December 2024 to focus on the most current discourse, a period defined by the rapid ascent of deep learning applications in health and transformative regulatory shifts such as the implementation of the European Union's General Data Protection Regulation (GDPR). The search strategy employed Boolean logic to combine key terms within three conceptual clusters: (1) the **population/context** of interest, using terms such as "big data," "health data," "EHR," "genomic data," and "multi-omics"; (2) the core **concept** of "privacy" or "confidentiality"; and (3) the relevant **dimensions**, including "ethics," "legal," "law," "regulation," "GDPR," "HIPAA," "technical," "security," "encryption," and "de-identification." To further ensure saturation, the reference lists of all identified review articles and key primary studies were manually screened for additional pertinent sources not captured by the initial electronic search.

Eligibility Criteria and Study Selection

Inclusion criteria were broad to capture diverse perspectives. Studies were included if they: (1) primarily addressed privacy concerns; (2) focused on big data in a health or biomedical research context; and (3) discussed one or more of the ethical,

legal, or technical dimensions. All publication types were considered, including original research, reviews, commentaries, legal analyses, and technical reports. Exclusions were: articles not in English, those focusing solely on clinical care (non-research), privacy, and articles published before 2015. After deduplication, titles and abstracts were screened for relevance, followed by a full-text review of selected articles.

Data Extraction and Thematic Synthesis

Data were extracted using a standardized form capturing: author(s), year, article type, primary dimension(s) of focus (ethical, legal, technical), key arguments, proposed solutions, and identified gaps. Given the heterogeneous nature of the evidence, a formal quality appraisal was not conducted, consistent with scoping review guidelines (Tricco et al., 2018). An inductive thematic analysis was performed. Extracted data were coded and iteratively grouped into themes within and across the three pre-defined dimensions (ethics, law, technology). These themes were then analyzed to identify relationships, tensions, and overarching narratives, forming the structure of this review.

Autonomy, Justice, and Trust in a Data-Driven Era

The ethical challenges posed by big data health research are foundational, questioning long-standing principles of research ethics. The core tension lies between the collective benefit of scientific progress and the protection of individual rights.

The Illusion of De-identification and Informed Consent

Traditional ethical models rely heavily on de-identification and specific informed consent. Big data undermines both. Studies have repeatedly demonstrated that supposedly anonymized datasets can be re-identified by linking them with publicly available information, such as voter registries, genealogy databases, or even fitness app data (Rocher et al., 2019). Genomic data is particularly vulnerable, as an individual's DNA is a unique and immutable identifier that can reveal information about biological relatives (Erlich & Narayanan, 2014). This renders the promise of anonymity, a cornerstone of minimal-risk review by ethics boards, increasingly tenuous.

Consequently, the model of specific, study-by-study informed consent becomes impractical for research involving data repositories intended for countless future, unspecified questions. Alternatives have been proposed, including *broad consent* (consenting to a general area of research), *dynamic consent* (digital platforms allowing ongoing patient engagement and granular choice), and *meta-consent* (consenting to preferred consent models) (Steinsbekk et al., 2013; Kaye et al., 2015). Each has limitations. Broad consent may be too vague to be truly autonomous. Dynamic consent, while ethically robust, may be technologically and administratively burdensome and could exacerbate participation biases if only the digitally literate engage. The ethical debate continues whether a social license or a form of solidarity-based governance can supplement or replace individual consent for certain types of public-good research (Vayena & Blasimme, 2018).

Justice, Equity, and Algorithmic Bias

Big data analytics risks perpetuating systemic inequities. If training data for predictive algorithms over-represents privileged populations (e.g., those of European ancestry in genomic databases), the resulting models may be inaccurate or harmful for underrepresented groups, worsening health disparities (Obermeyer et al., 2019). This is a profound justice issue. Furthermore, privacy protections themselves can be inequitable. Populations already distrustful of medical research due to historical exploitation may be less likely to share data, leading to their further exclusion from research benefits—a "participation-divide" (Mittelstadt & Floridi, 2016). Ethicists argue that justice requires not only protecting privacy but also actively ensuring equitable representation in datasets and auditing algorithms for fairness.

Beneficence, Risk, and the Public Good

The principle of beneficence obliges researchers to maximize benefits and minimize harms. In big data research, the potential benefit (e.g., discovering a new drug target) is often vast but probabilistic and distant, while the privacy harm (e.g., discrimination, stigma, psychological distress) is more immediate and personal, though often low-probability (Ferretti et al., 2021). Weighing these incommensurate risks and benefits is ethically fraught. Ethicists increasingly call for a *precautionary-proportionality* approach: implementing strong privacy protections by default (precaution) but allowing them to be calibrated based on the sensitivity of the data and the purpose of the analysis (proportionality) (Shabani & Borry, 2018).

Navigating a Fragmented Regulatory Landscape

The legal framework governing health data privacy is complex, varying significantly by jurisdiction and often lagging behind technological innovation. Compliance is a major challenge for international research collaborations.

Key Regulatory Frameworks: GDPR, HIPAA, and Beyond

The GDPR, implemented in 2018, represents the most comprehensive and stringent data protection law globally. It applies to any entity processing the personal data of EU citizens, regardless of the entity's location. Key provisions for research include: the requirement for a lawful basis for processing (which could be explicit consent or "public interest" for research), the right to erasure (which may conflict with research integrity), and

strict rules on international data transfer (Rockwern et al., 2021). Its definition of personal data is broad, likely encompassing pseudonymized data and possibly even genomic data per se.

In the United States, HIPAA provides a more limited framework, primarily covering "covered entities" (healthcare providers, plans) and their "business associates." Its "de-identification safe harbor" (removing 18 specific identifiers) is a common standard, though its robustness in the big data age is questioned (El Emam et al., 2020). HIPAA does not generally apply to researchers not affiliated with a covered entity or to data collected from wearables or direct-to-consumer genetic tests. This creates significant gaps. Other important laws include the California Consumer Privacy Act (CCPA) and China's Personal Information Protection Law (PIPL), each with its own requirements.

Tensions and Ambiguities

Several legal tensions arise. First, consent vs. secondary use: Laws like GDPR emphasize purpose limitation, making the reuse of data for new research questions legally complex unless re-consented to or properly exempted. Second, data ownership and control: Who "owns" health data—the patient, the institution, or the researcher? Legal systems rarely grant true property rights in data, instead focusing on control rights via privacy law (Evans, 2020). Third, extraterritoriality and conflict of laws: A research project involving data from the EU, USA, and Asia must navigate potentially conflicting legal obligations, creating a compliance labyrinth.

Governance as a Legal Complement

In response to these challenges, there is a growing emphasis on institutional and project-specific *governance* as a complement to top-down regulation. Data Access Committees (DACs), controlled data enclaves (like the NIH dbGaP), and clearly defined Data Use Agreements (DUAs) are critical tools for implementing the principles of transparency, accountability, and security mandated by law (Staunton et al., 2019). Ethical, Legal, and Social Implications (ELSI) programs, often embedded in large genomics projects, serve as proactive governance structures.

Privacy-Enhancing Technologies (PETs) and Their Trade-offs

Technical safeguards are the operational backbone of privacy protection. A suite of advanced PETs has emerged, each designed to minimize data exposure while preserving analytical utility.

Federated Learning (FL)

FL is a distributed machine learning approach where the model is sent to multiple local datasets (e.g., at different hospitals), trained locally, and only the model updates (not the raw data) are shared and aggregated centrally (Xu et al., 2021). This minimizes the risk of bulk data breaches and can help comply with data localization laws. However, it is not a panacea; inference attacks on the shared model updates could potentially reveal information about the training data, requiring additional safeguards like secure aggregation (Li et al., 2020).

Differential Privacy (DP)

DP is a rigorous mathematical framework that guarantees the output of a query (e.g., the average blood pressure in a dataset) will be statistically indistinguishable whether any single individual's data is included or not. This is achieved by carefully calibrating random noise to the query (Nissim & Wood, 2021). DP is considered a gold standard for privacy in statistical releases (e.g., by the U.S. Census Bureau). In health research, it can be applied to aggregate statistics or to the training of machine learning models. The core trade-off is between the level of privacy guarantee (epsilon, ε) and the utility/accuracy of the output: stronger privacy requires more noise, reducing accuracy (Abadi et al., 2016).

Homomorphic Encryption (HE) and Secure Multi-Party Computation (SMPC)

HE allows computations to be performed directly on encrypted data, producing an encrypted result that, when decrypted, matches the result of operations on the plaintext. This enables, for example, a researcher to run an analysis on encrypted data in a cloud server without the server ever seeing the raw data (Acar et al., 2018). SMPC allows multiple parties to jointly compute a function over their inputs while keeping those inputs private. Both techniques offer very strong security guarantees but have historically been limited by massive computational overhead, making them impractical for complex analyses on large datasets, though performance is improving.

Synthetic Data

This involves generating artificial datasets that mimic the statistical properties and relationships of the original sensitive data without containing any real individual records (Jordon et al., 2022). Synthetic data can be useful for software testing, model development, and some exploratory analyses. The critical challenge is ensuring the synthetic data is both privacy-safe (cannot be linked to real individuals) and analytically valid (conclusions drawn from it hold for the real population) (Table 1 & Figure 1).

Table 1: Key Privacy-Enhancing Technologies (PETs): Mechanisms, Applications, and Limitations

Technology	Core Mechanism	Primary Application in Health Research	Key Limitations & Trade-offs
Federated Learning (FL)	Train algorithm across decentralized data silos; share only model parameters.	Multi-institutional model training (e.g., medical imaging AI) without centralizing patient data.	Vulnerable to inference attacks on shared gradients; requires homogeneous data formats; complex orchestration.
Differential Privacy (DP)	Add calibrated mathematical noise to query outputs or model training.	Releasing aggregate statistics (e.g., disease prevalence) or training privacy-preserving ML models.	Privacy-Utility trade-off: more noise increases privacy but reduces data accuracy/ model performance.
Homomorphic Encryption (HE)	Perform computations on encrypted data.	Secure outsourcing of data analysis to an untrusted cloud server.	Extremely high computational overhead, limiting scale and complexity of analyses.
Secure Multi-Party Computation (SMPC)	Cryptographic protocol for joint computation with private inputs.	Secure genomic association studies across multiple biobanks.	High communication complexity between parties; slower than plaintext computation.
Synthetic Data Generation	Create artificial datasets with similar statistical properties to real data.	Software development, methodology research, preliminary hypothesis testing.	Risk of statistical disclosure if generation model overfits; may not capture rare but critical outliers.

Figure 1: Intersections of Ethics, Law, and Technology in Big Data Health Privacy.

The Imperative for Interdisciplinary Governance

The preceding analysis underscores that the ethical, legal, and technical dimensions of privacy in big data health research are not isolated silos but are deeply and dynamically entangled (Vayena & Blasimme, 2018). This convergence presents both alignments and tensions. For instance, the legal mandate for data minimization under GDPR Article 5 resonates with the technical security principle of "least privilege" and the ethical principle of proportionality, collectively promoting the collection and use of only that data which is strictly necessary (Shabani & Borry, 2018; Hamza et al., 2022). Similarly, technical solutions like differential privacy are engineered to directly mitigate the ethical and legal quandary of re-identification, offering a quantifiable safeguard against the compromise of individual anonymity (Dwork et al., 2016; Rocher et al., 2019). However, significant and persistent tensions emerge at the intersection of these domains, revealing that progress in one area can create challenges or contradictions in another. Navigating these tensions is the central governance challenge of the big data era.

Three core, interrelated tensions dominate the landscape. First, the fundamental trade-off between data utility and privacy protection remains unresolved. Implementing robust technical safeguards like high-noise differential privacy or computation-intensive homomorphic encryption invariably degrades the analytical utility of the data or imposes prohibitive computational costs (Ghazi et al., 2021). Ethically, this creates a dilemma where the imperative to minimize harm through strong privacy can conflict with the principle of beneficence, as it may reduce the potential scientific or clinical value of the research. Legally, this tension can manifest as a conflict between privacy regulations and other mandates to foster scientific innovation and public health advancement (Cohen & Mello, 2018). Second, the drive for global research collaboration clashes with a fragmented patchwork of local and national laws.

Large-scale, statistically powerful studies on rare diseases or diverse populations require international data sharing, yet researchers face a labyrinth of conflicting regulations, such as the GDPR's restrictions on extra-EU data transfers juxtaposed with other nations' data sovereignty laws (Staunton et al., 2019). This legal heterogeneity stifles collaboration and can perpetuate inequities by walling off data resources. Third, the ethical ideal of meaningful informed consent conflicts with the scalable, agile nature of big data innovation. While models like dynamic consent offer granular, ongoing

autonomy, they are often administratively burdensome and may not be scalable to the volume and velocity of modern data research, inadvertently favoring digitally literate populations and creating a "participation divide" (Mascalzoni et al., 2022; Mittelstadt & Floridi, 2016). This pressure pushes systems toward broader consent models, which, while pragmatic, risk eroding the foundational ethical principle of individual autonomy.

These convergent challenges make it abundantly clear that solutions cannot emerge from ethics, law, or technology in isolation. They demand integrated, interdisciplinary governance frameworks that are adaptive, proactive, and operationalized at multiple levels (Ienca et al., 2018). Effective governance must function institutionally through modernized Institutional Review Boards (IRBs) and Data Access Committees (DACs) equipped with expertise spanning all three domains (Ohno-Machado et al., 2018). At the project level, it requires clear, comprehensive Data Management and Use Plans (DMUPs) that pre-define privacy protocols. Technically, it mandates a "privacy-by-design" architecture where PETs are not afterthoughts but foundational components of the research infrastructure (Xu et al., 2021). A leading model for such governance is a risk-proportionate, tiered access system. This model calibrates the level of control to the sensitivity of the data and the intended use, facilitating responsible sharing while enforcing necessary protections. It typically features open access for truly anonymized or synthetic data; registered, managed access via DAC review for sensitive datasets requiring basic safeguards; and highly secured, audited access within trusted research environments or data enclaves for the most sensitive individual-level data, such as genomic sequences (Ohno-Machado et al., 2018; Lu et al., 2023). The elements of such an integrated framework are detailed in Table 2 & Figure 2, which outlines the necessary components across ethical, legal, technical, and oversight layers.

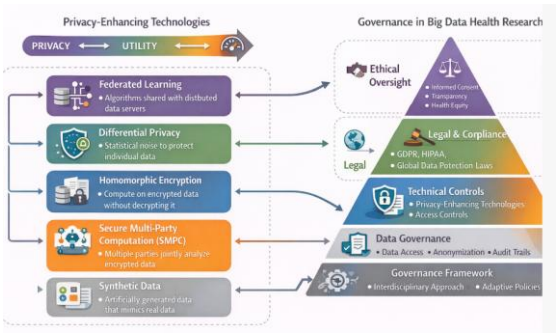

Figure 2: Privacy-Enhancing Technologies and Governance in Big Data Health Research

Table 2: Proposed Elements of an Integrated Privacy Governance Framework for Big Data Health Research

Governance Layer	Key Components	Interdisciplinary Actors
Ethical & Normative	- Adopt a precautionary-proportionality principle. - Implement fairness-aware algorithms and equity checks. - Develop culturally appropriate consent models (broad/dynamic/meta).	Bioethicists, Community Representatives, Social Scientists
Legal & Compliance	- Conduct Data Protection Impact Assessments (DPIAs). - Establish standardized Data Use Agreements (DUAs) for cross-border flows. - Clarify liability and accountability structures for data breaches.	Health Lawyers, Compliance Officers, Policy Experts
Technical & Operational	- Implement **Privacy-by-Design**: default use of PETs like DP or FL. - Create tiered data access systems with authentication/audit logs. - Develop benchmarks for evaluating PET performance (privacy/utility trade-off).	Data Scientists, Security Engineers, Clinical Informaticians
Oversight & Accountability	- Form multidisciplinary oversight boards with technical, ethical, legal expertise. - Ensure transparent public communication of data uses and benefits. - Establish clear channels for breach reporting and redress.	All stakeholders, including Patient Advocates

Conclusion and Future Directions

Big data holds immense promise for transforming health research, but this promise is contingent on resolving its profound privacy challenges. This scoping review demonstrates that privacy is not a single problem with a single solution but a complex landscape of interrelated ethical dilemmas, legal constraints, and technical possibilities. The traditional silos between these domains are a liability. Ethicists must understand the practical limits of PETs, lawyers must grasp the technical realities of de-identification, and computer scientists must appreciate the ethical weight of their design choices.

The path forward requires a commitment to interdisciplinary, adaptive governance. This involves moving beyond compliance-checking towards proactive, ethics-by-design and privacy-by-design approaches embedded in research projects from their inception. Key priorities for the future include: (1) Funding and developing more efficient PETs to improve the privacy-utility trade-off; (2) Harmonizing international regulations to facilitate global research while upholding strong protections, potentially through mutual recognition of adequacy or model contracts; (3) Investing in public engagement and literacy to build a socially informed consensus on acceptable data uses; and (4) Developing robust, real-time auditing tools for algorithmic bias and privacy violations.

Ultimately, safeguarding privacy in big data health research is not an obstacle to be circumvented but a cornerstone of ethical and sustainable science. By weaving together ethical principles, legal accountability, and technical ingenuity, the research community can build the trust necessary to unlock the full potential of big data for human health.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 308-318). https://doi.org/10.1145/2976749.2978318
2. Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys (Csur)*, *51*(4), 1-35. https://doi.org/10.1145/3214303
3. Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32. https://doi.org/10.1080/1364557032000119616
4. Cohen, I. G., & Mello, M. M. (2018). HIPAA and protecting health information in the 21st century. *Jama*, *320*(3), 231-232. doi:10.1001/jama.2018.5630
5. Dinov, I. D. (2016). Volume and value of big healthcare data. *Journal of medical statistics and informatics*, *4*, 3. https://doi.org/10.7243/2053-7662-4-3
6. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2016). Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality*, *7*(3), 17-51. https://doi.org/10.29012/jpc.v7i3.405
7. El Emam, K., Mosquera, L., & Bass, J. (2020). Evaluating identity disclosure risk in fully synthetic health data: model development and validation. *Journal of medical Internet research*, *22*(11), e23139. https://doi.org/10.2196/23139
8. Erlich, Y., & Narayanan, A. (2014). Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics*, *15*(6), 409-421. https://doi.org/10.1038/nrg3723
9. Evans, B. J. (2020). The streetlight effect: regulating genomics where the light is. *Journal of Law, Medicine & Ethics*, *48*(1), 105-118. doi:10.1177/1073110520916998
10. Ferretti, A., Ienca, M., Sheehan, M., Blasimme, A., Dove, E. S., Farsides, B., ... & Vayena, E. (2021). Ethics review of big data research: What should stay and what should be reformed?. *BMC medical ethics*, *22*(1), 51. https://doi.org/10.1186/s12910-021-00616-4
11. Ghazi, B., Golowich, N., Kumar, R., Manurangsi, P., & Zhang, C. (2021). Deep learning with label differential privacy. *Advances in neural information processing systems*, *34*, 27131-27145.
12. Hamza, R., Hassan, A., Ali, A., Bashir, M. B., Alqhtani, S. M., Tawfeeg, T. M., & Yousif, A. (2022). Towards secure big data analysis via fully homomorphic encryption algorithms. *Entropy*, *24*(4), 519. https://doi.org/10.3390/e24040519
13. Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C., & Vayena, E. (2018). Considerations for ethics review of big data health research: A scoping review. *PloS one*, *13*(10), e0204937. https://doi.org/10.1371/journal.pone.0204937
14. Jiang, Y., Mosquera, L., Jiang, B., Kong, L., & El Emam, K. (2022). Measuring re-identification risk using a synthetic estimator to enable data sharing. *PLoS One*, *17*(6), e0269097. https://doi.org/10.1371/journal.pone.0269097
15. Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., ... & Weller, A. (2022). Synthetic Data--what, why and how?. *arXiv preprint arXiv:2205.03257*. https://doi.org/10.48550/arXiv.2205.03257

16. Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European journal of human genetics*, *23*(2), 141-146. https://doi.org/10.1038/ejhg.2014.71
17. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, *37*(3), 50-60. https://doi.org/10.1109/MSP.2020.2975749
18. Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., & Wei, W. (2023). Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062*. https://doi.org/10.48550/arXiv.2302.04062
19. Mascalzoni, D., Melotti, R., Pattaro, C., Pramstaller, P. P., Gögele, M., De Grandi, A., & Biasiotto, R. (2022). Ten years of dynamic consent in the CHRIS study: informed consent as a dynamic process. *European Journal of Human Genetics*, *30*(12), 1391-1397. https://doi.org/10.1038/s41431-022-01160-4
20. Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *The ethics of biomedical big data*, 445-480. https://doi.org/10.1007/978-3-319-33525-4_19
21. Nissim, K., & Wood, A. (2021, December). Foundations for robust data protection: Co-designing law and computer science. In *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)* (pp. 235-242). IEEE. https://doi.org/10.1109/TPSISA52974.2021.00026
22. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342
23. Ohno-Machado, L., Kim, J., Gabriel, R. A., Kuo, G. M., & Hogarth, M. A. (2018). Genomics and electronic health record systems. *Human molecular genetics*, *27*(R1), R48-R55. https://doi.org/10.1093/hmg/ddy104
24. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*(1), 3. https://doi.org/10.1186/2047-2501-2-3
25. Rezaeikhonakdar, D. (2023). AI chatbots and challenges of HIPAA compliance for AI developers and vendors. *Journal of Law, Medicine & Ethics*, *51*(4), 988-995. doi:10.1017/jme.2024.15
26. Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, *10*(1), 3069. https://doi.org/10.1038/s41467-019-10933-3
27. Rockwern, B., Johnson, D., Snyder Sulmasy, L., & Medical Informatics Committee and Ethics, Professionalism and Human Rights Committee of the American College of Physicians. (2021). Health information privacy, protection, and use in the expanding digital health ecosystem: a position paper of the American College of Physicians. *Annals of internal medicine*, *174*(7), 994-998. https://doi.org/10.7326/M20-7639
28. Shabani, M., & Borry, P. (2018). Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics*, *26*(2), 149-156. https://doi.org/10.1038/s41431-017-0045-7
29. Staunton, C., Slokenberga, S., & Mascalzoni, D. (2019). The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks. *European Journal of Human Genetics*, *27*(8), 1159-1167. https://doi.org/10.1038/s41431-019-0386-5
30. Steinsbekk, K. S., Kåre Myskja, B., & Solberg, B. (2013). Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem?. *European Journal of Human Genetics*, *21*(9), 897-902. https://doi.org/10.1038/ejhg.2012.282
31. Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, *169*(7), 467-473. https://doi.org/10.7326/M18-0850
32. Vayena, E., & Blasimme, A. (2018). Health research with big data: time for systemic oversight. *The journal of law, medicine & ethics*, *46*(1), 119-129. https://doi.org/10.1177/1073110518766026
33. Vayena, E., Dzenowagis, J., Brownstein, J. S., & Sheikh, A. (2017). Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, *96*(1), 66. https://doi.org/10.2471/BLT.17.197426
34. Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated learning for healthcare informatics. *Journal of healthcare informatics research*, *5*(1), 1-19. https://doi.org/10.1007/s41666-020-00082-4.