American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Ethical Data Reuse in Multidisciplinary Health and Social Research

Dr. Claire Thompson

Professor, University of Toronto, Toronto, Canada

Abstrac

The growing availability of electronic health records, administrative databases, longitudinal surveys, social-care records, genomic repositories, digital-behavior traces, and population datasets has created substantial opportunities for multidisciplinary health and social research. Data reuse can reduce duplication in primary data collection, enable longitudinal and cross-sector analysis, improve statistical power, support research on comparatively rare conditions, and facilitate investigation of the social determinants of health. These scientific benefits are accompanied by ethical challenges because data collected for clinical care, public administration, social services, previous research, or another legitimate purpose may subsequently be used in contexts that individuals did not specifically anticipate. Contemporary research-ethics frameworks therefore require more than simple de-identification or institutional possession of a dataset.

Ethical reuse must consider consent and reasonable expectations, scientific and public value, proportionality, privacy, re-identification risk, data minimization, group harms, equitable benefit, secure access, accountability, and community interests. Current international developments reinforce this governance-oriented direction. The World Health Organization supports secondary research reuse under technical, ethical, and legal safeguards; NIH policy combines data-sharing expectations with privacy and controlled-access mechanisms; and the European Health Data Space establishes a developing framework for regulated secondary use through permits and secure processing environments.

The present study develops an Ethical Multidisciplinary Data Reuse Framework (EMDRF) and evaluates four hypothetical governance models through an explicitly simulation-based methodology. A synthetic corpus of 400 proposed data-reuse scenarios was modeled across six domains: purpose legitimacy, consent compatibility, privacy protection, secure access, social and community fairness, and accountability. The simulated Ethical Data-Reuse Readiness Score increases from 38 for minimally governed open reuse to 91 for interdisciplinary ethical reuse. The findings demonstrate that ethical data reuse should not be framed as a binary choice between unrestricted openness and complete data closure. Rather, responsible reuse requires proportional combinations of technical, organizational, ethical, legal, and participatory safeguards matched to the sensitivity of the data and the consequences of the proposed research.

Keywords: data reuse, secondary data research, health data, social research, research ethics, privacy, informed consent, data governance, multidisciplinary research

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

1. Introduction

Multidisciplinary health and social research increasingly depends on the ability to connect information produced in different institutional settings. A study of health inequality may require clinical records together with socioeconomic indicators, education histories, employment information, housing conditions, social-care records, environmental exposures, or longitudinal survey responses. Such integration can generate knowledge that would be difficult or prohibitively expensive to reproduce through new primary data collection. The FAIR principles have further strengthened expectations that scholarly data should be findable, accessible under appropriate conditions, interoperable, and reusable, thereby supporting reproducibility and cumulative scientific inquiry. Reusability, however, does not mean unrestricted accessibility. In sensitive health and social research, the ethical question is not simply whether data *can* technically be reused but whether a proposed reuse is sufficiently justified, proportionate, secure, and consistent with the interests and reasonable expectations of the people and communities represented in the data.

Secondary use creates a distinctive ethical problem because the original context of data collection and the later research context may differ substantially. Information collected during medical treatment may later be combined with administrative or socioeconomic records; survey data collected for one question may be reanalyzed to examine another; and datasets originally considered relatively low risk may become more identifying when linked with additional sources. McKeown and colleagues argue that data-intensive health-research platforms require renewed thinking about consent, reasonable expectations, trust, and social legitimacy because future research purposes cannot always be fully specified at the moment data are originally collected. The 2024 revision of the World Medical Association's Declaration of Helsinki similarly addresses foreseeable secondary use of identifiable or re-identifiable data and requires appropriate consent and ethics oversight, while recognizing that unforeseen secondary use may sometimes require ethics-committee consideration when obtaining renewed consent is impossible or impracticable.

The ethical complexity becomes greater in multidisciplinary research because different disciplines bring different assumptions about data ownership, confidentiality, participant relationships, and acceptable forms of inference. Biomedical researchers may emphasize individual privacy and clinical confidentiality, whereas social scientists may additionally confront contextual disclosure, stigmatization, structural inequality, group representation, and community-level harms. A technically anonymized dataset can still generate conclusions that adversely stereotype a small geographic, ethnic, occupational, or social population. Indigenous data-governance scholarship makes this point especially clearly. The CARE Principles complement technically oriented FAIR principles by emphasizing collective benefit, authority to control, responsibility, and ethics, thereby drawing attention to the people and communities represented through data. Ethical reuse must therefore extend beyond protection of isolated individual records to include relational and collective consequences.

Contemporary governance is consequently moving toward controlled and accountable reuse rather than a simplistic opposition between open and closed data. WHO's policy on sharing and reuse of health-related research data explicitly addresses technical, ethical, and legal considerations for secondary analysis. In the United States, the NIH Data Management and Sharing Policy requires planning for scientific-data management and sharing while providing mechanisms for justified limitations and

controlled access where sensitivity requires additional safeguards. NIH also provides specific resources concerning consent for secondary research with data and biospecimens. In Europe, Regulation (EU) 2025/327 establishing the European Health Data Space entered into force on March 26, 2025, although its secondary-use requirements are being introduced progressively, with major rules for most data categories scheduled to begin applying in March 2029. The framework includes health-data access bodies, permits, secure processing environments, restrictions on re-identification, and controls over permissible purposes.

2. Objectives of the Study

2.1 Development of an Ethical Data-Reuse Framework

The first objective is to construct a multidisciplinary framework capable of evaluating secondary use across clinical, population-health, social-science, administrative, and linked-data research. The framework moves beyond a single privacy criterion and incorporates scientific value, consent compatibility, proportionality, security, fairness, and accountability.

2.2 Comparison of Alternative Governance Models

The second objective is to compare open reuse with minimal governance, repository-based de-identified reuse, controlled-access governed reuse, and fully interdisciplinary ethical reuse. The comparison identifies the strengths and limitations of progressively stronger governance arrangements.

2.3 Identification of Conditions for Responsible Reuse

The third objective is to determine which safeguards are most important when previously collected information is repurposed across disciplinary or institutional boundaries. Particular attention is given to consent compatibility, re-identification risk, secure access, public interest, community interests, data minimization, accountability, and withdrawal or opt-out mechanisms where applicable.

3. Review of Literature

3.1 Consent, Reasonable Expectations, and Secondary Research

The traditional informed-consent model works most clearly when investigators know the purpose, procedures, risks, and anticipated uses of a study before participants enroll. Data-intensive research complicates this model because datasets may remain scientifically valuable long after the original project ends and future questions may not yet be foreseeable.

McKeown and colleagues examined this tension specifically in health-data platforms and argued that responsible reuse requires attention to reasonable expectations, public legitimacy, and trust rather than reliance on an oversimplified contractual understanding of consent. Their work highlights an important distinction between ethically justified future research and unrestricted reuse merely because a dataset exists.

Broad consent represents one response. Participants authorize a defined range of future uses rather than approving each individual project. The revised U.S. Common Rule includes provisions under which broad consent may support certain secondary uses of identifiable private information or identifiable biospecimens, although the regulatory structure contains specific requirements and

exemptions that must be interpreted in context. The Common Rule also recognizes circumstances in which secondary research involving existing identifiable information may qualify for specific exemptions, while consent waivers remain subject to criteria including practicability and protection of participants.

The WMA's current Declaration of Helsinki provides a complementary ethical standard. Its 2024 revision explicitly encompasses medical research involving identifiable human material or data. For stored biological material and identifiable or re-identifiable information, the framework addresses consent for foreseeable secondary use and ethics oversight for unanticipated reuse. The WMA Declaration of Taipei further provides ethical guidance for health databases and biobanks, including governance, privacy protection, access rules, consent, commercial use, transfer, and possible withdrawal implications.

Consent, however, should not become the only ethical criterion. A consent form cannot make scientifically weak, discriminatory, insecure, or exploitative research ethically acceptable. Likewise, data reuse without study-specific consent is not automatically unethical if a legitimate regulatory or ethical pathway exists, the research has substantial value, risks are minimized, and meaningful governance safeguards apply.

3.2 Privacy, Re-identification, Secure Access, and Data Stewardship

De-identification is one of the most frequently invoked safeguards in secondary research, but it should not be treated as a permanent guarantee of anonymity.

Risk depends on the characteristics of the dataset, availability of external information, data-linkage possibilities, geographic and demographic specificity, and technological developments. Research involving multiple linked datasets may create additional inferential opportunities even when obvious direct identifiers have been removed.

OHRP distinguishes between data that investigators can link to specific individuals and information that investigators cannot readily link through coding arrangements when assessing whether research involves identifiable private information. This regulatory distinction is important, but ethical risk assessment should remain sensitive to practical re-identification possibilities and contextual harms.
A proportionate approach uses multiple protective layers.

The Five Safes framework provides one established model. It evaluates safe projects, safe people, safe settings, safe data, and safe outputs rather than relying exclusively on modification of the dataset itself. The framework is widely used for controlled research access in government and social-data environments. Green and Ritchie note both its continuing relevance and the need to understand the framework as part of evolving principles-based data governance rather than a rigid checklist.

Controlled-access repositories and trusted research environments operationalize a similar principle. Highly sensitive information can remain more detailed when access occurs only for approved researchers, approved projects, in restricted computing environments, with disclosure review of outputs. Conversely, genuinely public datasets require much stronger protection in the data itself because researchers and subsequent uses cannot be controlled.

NIH currently encourages established repositories and recommends considering dataset sensitivity, complexity, and expected demand when choosing sharing mechanisms. It also provides controlled-access options as an additional privacy safeguard for human-participant data.

The emerging European Health Data Space adopts an especially explicit controlled-use architecture. Under the regulation's secondary-use framework, qualifying access will require authorization from a health data access body; processing is to occur within secure processing environments; pseudonymized information may be used when anonymized information is insufficient; downloading personal data from these environments is restricted; and re-identification attempts are prohibited. The major secondary-use provisions are scheduled to phase in from 2029 rather than being fully operational in 2021

These developments support a broader principle: ethical openness does not require identical access conditions for every dataset.

3.3 Fairness, Public Value, and Multidisciplinary Governance

Health and social research generates value partly because it reveals relationships that individual datasets cannot show. Linking clinical data with social determinants may identify preventable inequalities or improve public services. Yet the same analytical power can increase ethical risk.

Research conclusions may affect communities even when no person is re-identified. Analyses can label geographic areas as high risk, associate social groups with stigmatized outcomes, or generate predictive systems that subsequently influence access to insurance, employment, healthcare, or social services.

The Belmont Report remains relevant through its foundational principles of respect for persons, beneficence, and justice, which provide a basis for considering autonomy, risk, benefit, and equitable treatment in research. CIOMS broadened its 2016 guidelines from biomedical research to health-related research partly to encompass research involving health-related data and the increasingly overlapping boundaries among biomedical, epidemiological, behavioral, and social research.

The CARE Principles add a crucial collective dimension. Whereas FAIR focuses on the technical and stewardship conditions that make data reusable, CARE emphasizes whether reuse generates collective benefit, respects authority to control, fulfills researcher responsibilities, and remains ethically grounded. These principles are particularly important when datasets concern Indigenous peoples or other communities with historical experiences of extractive research.

4. Research Methodology

4.1 Research Design and Analytical Corpus

The study adopts a simulation-based methodological design.

No real patient records, social-service data, survey responses, administrative files, participant identifiers, or linked individual-level information were accessed.

equally across four governance configurations:

1. Open reuse with minimal governance;
2. De-identified repository reuse;
3. Controlled-access governed reuse; and

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

4. Interdisciplinary ethical reuse.

Each hypothetical proposal was evaluated across six dimensions representing ethical readiness.

Table 1: Ethical Multidisciplinary Data Reuse Framework

Ethical Dimension	Core Assessment Question	Illustrative Safeguards	Scale
Purpose Legitimacy	Is the proposed secondary use scientifically credible and socially defensible?	Scientific review, public-value justification, proportionality assessment	0–100
Consent & Expectation Compatibility	Is the reuse compatible with original consent, permissions, lawful authority, or reasonable participant expectations?	Consent review, ethics waiver where valid, opt-out/broad consent where applicable	0–100
Privacy & Data Minimization	Is only the information necessary for the research made available?	De-identification, pseudonymization, minimization, linkage controls	0–100
Secure Access	Are access conditions proportionate to data sensitivity?	Trusted research environments, authentication, logging, role-based access	0–100
Social & Community Fairness	Have group harms, discrimination, representation, and benefit distribution been considered?	Community engagement, equity assessment, CARE-informed governance	0–100
Accountability & Transparency	Can institutions demonstrate who uses data, for what purpose, under which controls, and with what outputs?	Data-use agreements, audits, access committees, publication/output review	0–100

Source: Author-developed methodological framework synthesizing WHO, NIH, WMA, CIOMS, Five Safes, FAIR/CARE, and contemporary health-data governance principles.

4.2 Construction of the Four Governance Models

The Open Reuse with Minimal Governance model represents circumstances in which a dataset is redistributed with limited project screening, weak monitoring of downstream use, and strong dependence on de-identification as the principal safeguard.

The De-identified Repository Reuse model adds institutional repository management, metadata, standardized data-use terms, removal or transformation of direct identifiers, and basic review of access requests.

The Controlled-Access Governed Reuse model adds project approval, researcher authentication, data-use agreements, controlled computing environments, logging, output review, and stronger restrictions on linkage and onward disclosure. This structure resembles several principles used in Five Safes–oriented access systems.

The Interdisciplinary Ethical Reuse model incorporates all of the above while adding explicit assessment of consent compatibility, research ethics, public value, community and group harms, equity, disciplinary context, downstream reuse, and continuing accountability.

.

4.3 Analytical Procedure and Reproducibility

A simulated Ethical Data-Reuse Readiness Score (EDRRS) was calculated on a 0–100 scale using the six framework dimensions.

Equal weighting was used for methodological simplicity. The equal weights should not be interpreted as empirically validated coefficients.

A future empirical validation could use a Delphi panel comprising:

- research-ethics specialists;
- health-data scientists;
- social scientists;
- privacy professionals;
- patient and public representatives;
- community representatives;
- data stewards;
- statisticians;
- cybersecurity specialists; and
- institutional data-access committee members.

Future validation should also test inter-rater agreement and examine whether weighting needs to vary according to data type.

For example, social-media research may require greater emphasis on contextual integrity, whereas genomic reuse may require stronger attention to familial and group implications.

5. Analysis

5.1 Simulated Ethical Performance Across Governance Models

The simulated results demonstrate a substantial increase in ethical readiness as reuse moves from de-identification-centered access toward interdisciplinary governance.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Table 2: Simulated Ethical Data-Reuse Readiness Scores

Governance Model	Purpose Legitimacy	Consent Compatibility	Privacy & Minimization	Secure Access	Social/Community Fairness	Accountability	Overall Score
Open Reuse with Minimal Governance	52	34	46	29	27	40	**38**
De-identified Repository Reuse	68	55	73	61	48	61	**61**
Controlled-Access Governed Reuse	82	71	86	91	65	79	**79**
Interdisciplinary Ethical Reuse	94	90	93	94	86	89	**91**

Source: Simulated analytical values. Scores are methodological illustrations and not empirical estimates.

The lowest score occurs in the minimally governed model because de-identification alone does not adequately address purpose legitimacy, downstream accountability, group consequences, or changing re-identification risk.

Repository-based reuse improves privacy management and documentation but remains weaker in consent compatibility and community fairness.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

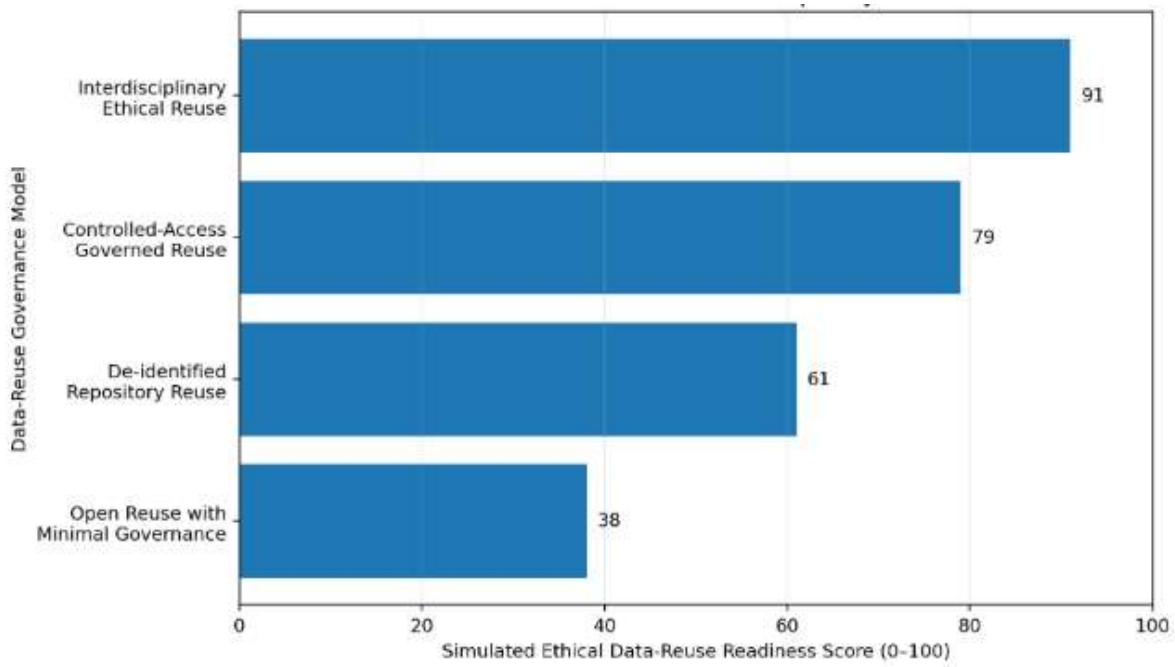
E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Controlled access produces a marked increase because researcher identity, project purpose, computing environment, and outputs can be governed.

The interdisciplinary model receives the highest simulated score because protection operates across the entire research lifecycle.

5.2 Ethical Readiness Across Data-Reuse Models

Graph 1: Simulated Ethical Data-Reuse Readiness



Graph 1 shows simulated scores of 38 for minimally governed open reuse, 61 for repository-based de-identified reuse, 79 for controlled-access governed reuse, and 91 for interdisciplinary ethical reuse.

The increase from 61 to 79 is particularly significant conceptually because it demonstrates the difference between altering the dataset and governing the research environment.

This logic is broadly consistent with the Five Safes framework while extending it through consent, equity, and community governance.

5.3 Ethical Reuse as a Decision Pathway

The analysis suggests that multidisciplinary data reuse should proceed through sequential questions.

- First, researchers should establish whether the reuse has sufficient scientific and public value.
- Second, the proposed purpose should be compared with the original collection context, participant information, consent terms, legal basis, and reasonable expectations.
- Third, the minimum dataset required to answer the question should be identified.
- Fourth, privacy and re-identification risks should be assessed after considering linkage with external information.
- Fifth, access conditions should be matched to sensitivity.

- Sixth, researchers should evaluate whether findings could harm identifiable communities even if individual participants remain anonymous.
- Seventh, an accountable institution should retain oversight throughout analysis and dissemination.

6. Discussion

6.1 Beyond the False Choice between Openness and Privacy

A major contribution of the proposed framework is its rejection of the binary assumption that research data must either be completely open or permanently inaccessible. In practice, data access can operate across a spectrum according to sensitivity, potential harm, and research value. Low-risk aggregate information may be openly disseminated, while moderately sensitive de-identified information may be shared through data-use agreements. Highly detailed individual-level datasets can be made available only within secure research environments and particularly sensitive or community-governed information may require additional authorization or remain inaccessible for certain purposes.

This proportional approach is reflected in contemporary health-data governance frameworks that combine expectations for scientific data sharing with justified restrictions and controlled-access mechanisms. The ethical objective should therefore be maximum responsible usefulness rather than maximum unrestricted exposure.

This distinction also provides a way to reconcile open-science objectives with privacy and confidentiality obligations. Principles such as FAIR—Findable, Accessible, Interoperable, and Reusable—do not necessarily require every dataset to be publicly downloadable without restrictions. Accessibility can instead be achieved through clearly defined, proportionate, and accountable conditions that allow legitimate researchers to obtain and use information while reducing unnecessary exposure. Such an approach recognizes that openness and privacy are not inherently opposing values.

6.2 Consent Is Essential but Governance Must Continue After Consent

A second implication concerns the tendency to treat informed consent as the final ethical authorization for all subsequent data processing and reuse. Consent remains fundamental whenever it is ethically and legally required, but it cannot always function as the sole safeguard for long-term multidisciplinary research. Participants may be unable to anticipate future research questions, emerging technologies, new forms of data linkage, or the involvement of future commercial and public-sector actors. Moreover, a dataset considered relatively low risk when consent was obtained may become substantially more revealing as analytical techniques and linkage capabilities develop.

Consequently, research institutions need continuing governance mechanisms even when participants have provided broad or specific consent. A stronger ethical model should therefore distinguish between the adequacy of the original authorization and the ethical acceptability of a proposed reuse at a later stage. Institutional oversight, data-access committees, privacy protections, purpose limitations, and continuing monitoring can provide safeguards when research circumstances change.

6.3 Multidisciplinary Reuse Requires Social and Community Accountability

The third implication is particularly important for research that combines health and social information. Health datasets often concern individual characteristics and clinical outcomes, whereas social datasets

may describe households, relationships, schools, workplaces, neighborhoods, socioeconomic circumstances, migration, and disability, ethnicity, or community membership. Combining these sources can produce valuable evidence about structural determinants of health and social inequality, but it can also create group-level risks that are not captured by conventional individual privacy protections. For example, research identifying an association between a particular community and a negative health outcome may unintentionally contribute to stigmatization even when individual records are anonym zed. Similarly, predictive models initially developed for population-level research may later be adapted for individual risk scoring without participants having anticipated such applications.

The CARE Principles provide an important corrective by emphasizing collective benefit, authority, responsibility, and ethics in data governance. Although community accountability is particularly important for Indigenous and historically marginalized populations, the underlying principle is broadly applicable to multidisciplinary social research because data can produce collective as well as individual consequences.

7. Conclusion

Ethical data reuse has become an increasingly important component of multidisciplinary health and social research because previously collected datasets can support longitudinal investigation, comparative analysis, reproducibility, policy evaluation, and research on complex relationships between health and social conditions. However, the availability of data does not itself justify unrestricted secondary use. The present study demonstrates that responsible reuse requires a coordinated governance model that considers scientific and public value, consent and participant expectations, privacy protection, data minimization, secure access, social and community fairness, and institutional accountability. The simulation-based comparison further illustrates that reliance on de-identification alone provides substantially weaker ethical protection than controlled-access and interdisciplinary governance models. Ethical reuse should therefore be understood as a continuing process of responsible stewardship rather than a one-time decision to release or share a dataset.

The findings also emphasize that consent, although important, cannot function as the sole ethical safeguard for secondary research. Future data uses may involve new technologies, additional datasets, different institutional actors, or analytical purposes that were not foreseeable when the information was originally collected. For this reason, continuing ethics review, proportional access controls, data-use agreements, secure research environments, and transparent institutional oversight remain necessary even where broad or specific consent exists. Similarly, data minimization and controlled access should be matched to the sensitivity of the information and the potential consequences of misuse rather than applied uniformly to every dataset. This proportional approach supports scientific reuse while reducing unnecessary exposure of individual-level information.

Multidisciplinary research also requires ethical attention beyond individual privacy. Linking clinical, administrative, economic, demographic, geographic, and social information can generate findings about identifiable populations even when individual records remain protected. Such analyses may unintentionally contribute to stigmatization, discriminatory profiling, or unequal distribution of research benefits. Ethical governance should therefore assess collective and community-level consequences alongside individual confidentiality. Researchers should consider who is represented in

the data, whether particular populations face disproportionate risks, how findings are communicated, and whether communities contributing valuable information are reasonably positioned to benefit from the resulting knowledge.

The study ultimately concludes that ethical data reuse should not be framed as a choice between complete data closure and unrestricted openness. Responsible research can operate through a spectrum of access arrangements ranging from publicly available low-risk aggregate data to highly controlled environments for sensitive individual-level information. The appropriate governance model should depend on the purpose of the research, sensitivity of the dataset, potential for re-identification, consent conditions, expected social value, and possible consequences for individuals and communities. In this sense, the most defensible model of multidisciplinary data reuse is one that maximizes legitimate scientific and public benefit while maintaining proportionate privacy, security, fairness, transparency, and accountability throughout the entire research lifecycle.

8. References

1. Zhang, H. H. (2019). Ethical review and related considerations for research involving secondary use of health data. *Medicine & Philosophy, 40*(3), 26–28.
2. Peloquin, D., DiMaio, M., Bierer, B., & Barnes, M. (2020). Disruptive and avoidable: GDPR challenges to secondary research uses of data. *European Journal of Human Genetics, 28*, 697–705.
3. Arora, C. (2019). Digital health fiduciaries: Protecting user privacy when sharing health data. *Ethics and Information Technology, 21*, 213–222.
4. Stockdale, J., Cassell, J., & Ford, E. (2019). "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Research, 4*, 6.
5. Global Forum on Bioethics in Research. (2018). *The ethics of data sharing and biobanking in health research*. *Developing World Bioethics*.
6. World Health Organization. (2018). *Big data and artificial intelligence for achieving universal health coverage: An international consultation on ethics*. World Health Organization.
7. Shah, S. M., & Khan, R. A. (2020). Secondary use of electronic health record: Opportunities and challenges. *arXiv preprint arXiv:2001.09479*.
8. Scheibner, J., Raisaro, J. L., Troncoso-Pastoriza, J. R., Ienca, M., Fellay, J., Vayena, E., & Hubaux, J.-P. (2020). Revolutionizing medical data sharing using advanced privacy-enhancing technologies: Technical, legal and ethical synthesis. *arXiv preprint arXiv:2010.14445*.
9. Richter, G., Borzikowsky, C., Lesch, W., Semler, S. C., Bunnik, E. M., Buyx, A., et al. (2020). Secondary research use of personal medical data: Attitudes from patient and population surveys in The Netherlands and Germany. *European Journal of Human Genetics*.
10. Vayena, E., Blasimme, A., & Sugarman, J. (2018). Policy and ethical implications of big data in translational research. *Journal of General Internal Medicine, 33*, 1513–1519.
11. Murtagh, M. J., Blell, M. T., Butters, O. W., & et al. (2012). Better together? Interdisciplinary approaches to health data governance and research. *Journal of Medical Ethics*.
12. Dove, E. S. (2015). Biobanks, data and the governance of health research. *Journal of Law and the Biosciences, 2*(3), 620–625.

13. Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review, 154*(3), 477–560.
14. Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*(1), 119–158.
15. Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*, 1701–1777.
16. Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *2008 IEEE Symposium on Security and Privacy*, 111–125.
17. Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science, 339*(6117), 321–324.
18. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health and Human Services.
19. Council for International Organizations of Medical Sciences. (2016). *International Ethical Guidelines for Health-Related Research Involving Humans*. CIOMS.
20. World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA, 310*(20), 2191–2194.