

Viewpoint

The Continuity Trap in Data Science Health Research

Clement Adebamowo^{1,2}, BMChB Hons, SCD; Sally Nneoma Adebamowo^{1,2}, MBBS, MSc, SCD; Adeola Akintola³, MSc; Peter Ikhane^{1,3,4}, MSc, PhD; Simisola Akintola^{4,5}, BL, LLB, LLM, PhD; Temidayo Ogundiran^{3,4,6}, MBBS, MHSc; Ayodele Jegede^{4,7}, PhD; Olusegun Adeyemo³, MSc; Shawneequa Callier⁸, MA, JD; Muhammad Imam-Tamim⁹, PhD; Ibrahim Uthman¹⁰, PhD; BridgeELSI Project as part of the DS-I Africa Consortium³

¹Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States

²University of Maryland Marlene and Stewart Greenebaum Comprehensive Cancer Center, Baltimore, MD, United States

³Department of Research, Center for Bioethics and Research, Ibadan, Oyo, Nigeria

⁴Department of Bioethics and Medical Humanities, Faculty of Multidisciplinary Studies, University of Ibadan, Ibadan, Oyo State, Nigeria

⁵Faculty of Law, University of Ibadan, Oyo, Oyo State, Nigeria

⁶Department of Surgery, College of Medicine, University of Ibadan, Oyo, Oyo State, Nigeria

⁷Department of Sociology, Faculty of Social Sciences, University of Ibadan, Ibadan, Oyo State, Nigeria

⁸Department of Clinical Research and Leadership, School of Medicine and Health Sciences, George Washington University, Washington DC, DC, United States

⁹Department of Islamic Law, Faculty of Law, University of Ilorin, Ilorin, Kwara State, Nigeria

¹⁰Department of Arabic and Islamic Studies, Faculty of Islamic Studies, University of Ibadan, Ibadan, Nigeria

Corresponding Author:

Clement Adebamowo, BMChB Hons, SCD

University of Maryland Marlene and Stewart Greenebaum Comprehensive Cancer Center

660 W Redwood St

Baltimore, MD, 21201

United States

Phone: 1 4107066116

Email: cadebamowo@som.umaryland.edu

Abstract

Secondary use is now the ordinary condition of data science health research rather than an exception to it. Electronic health records collected for clinical care become prediction tools and inputs for generative AI; imaging archives become foundation-model corpora; genomic datasets become resources for polygenic risk scores; and legacy biospecimens become renewable, indefinitely distributable cell lines. Governance has responded by emphasizing verifiable instruments such as provenance logs, repository approvals, broad-consent forms, data-use agreements, model cards, records of processing, and locality-preserving architectures. These instruments are necessary, and they answer real questions about lineage, privacy, institutional responsibility, and accountability, but they are not sufficient to establish that a present use remains ethically justified. We define ethical continuity as the persistence of normatively relevant relationships between the original conditions of data generation or material collection and subsequent downstream uses, such that current uses remain justifiable in light of the expectations, permissions, meanings, and relational obligations present at entrustment. We then define the Continuity Trap as a review-stage governance error in which a salient signal of continuity in one domain is treated as sufficient evidence of ethical continuity overall, causing inquiry into the remaining domains to close prematurely. The trap is not ordinary noncompliance, ethics creep, or a demand for universal rereview; it is a cross-domain inference error that can arise even in careful, good-faith review. We distinguish it from proxy closure, of which it is a continuity-specific subtype, and from Goodhart's and Campbell's laws, which describe how measures degrade once they become targets. We operationalize ethical continuity across 4 domains: provenance, semantics, authorization, and relational standing, developed in our Representational Veracity framework, and we show that these domains can diverge as data are linked, transformed, modeled, and redeployed. We identify the institutional mechanisms—provenance privilege, descriptor sedimentation, authorization fossilization, and community effacement—that cause auditable signals to be overread, and we examine how the US Health Insurance Portability and Accountability Act (HIPAA) of 1996, the General Data Protection Regulation, the European Health Data Space, US Food and Drug Administration guidance, the US National Institute of Standards and Technology (NIST) AI Risk Management Framework, and federated-learning governance can reduce risk while still inducing continuity traps. We apply the framework to consent and nonconsent settings, including public health, immunization, syndromic, and wastewater

surveillance, polygenic risk scores, induced pluripotent stem cells, federated learning, and health-related large language models. The policy implication is trigger-based continuity review: rather than rereviewing every reuse, investigators and reviewers should identify the weakest continuity domain at the present data stage and impose a domain-matched safeguard, recorded in a short continuity statement. This reframing is intended for the committees, repositories, funders, and governance bodies that decide whether reuse may proceed, and it matters most in cross-border and low-resource settings. Provenance should begin ethical review; it should not end it.

(*J Med Internet Res* 2026;28:e98699) doi: [10.2196/98699](https://doi.org/10.2196/98699)

KEYWORDS

concept drift; data life cycle; dataset shift; ethical governance continuity; machine learning governance; management; model accountability

Introduction

Data science health research increasingly depends on data and materials generated for other purposes. Clinical records become prediction models; images become foundation-model corpora; genomic datasets become inputs for polygenic risk scores; and stored biospecimens become cell lines or other derivatives. Secondary use is therefore not an exception to the field but one of its ordinary conditions, driven by the falling cost of storage, the reach of linkage and deidentification methods, and the appetite of contemporary machine-learning systems for large and heterogeneous training corpora. Recent discussions of consent for secondary use in AI models, clinical machine-learning ethics, health-equity harms, privacy in the age of medical big data, the ethics of sharing clinical imaging data, and dataset and model documentation all converge on the same point: legal access and technical documentation are indispensable, but they do not exhaust the ethical inquiry [1-11]. Governance frameworks grew up around these auditable artifacts because they can be produced, filed, and inspected, which is also why they come to stand in for the harder judgments they cannot make.

The ethical governance difficulty that arises from this is that the most visible forms of continuity are often the easiest to overread. An institutional review board, data-access committee, repository, material-transfer office, or AI governance body can verify a consent form, access approval, chain-of-custody record, data-use agreement, source-data statement, model card, or locality-preserving architecture. These instruments answer important questions about lineage, privacy, institutional responsibility, and accountability, and they are frequently the only artifacts a reviewer can inspect within the time and authority available to a single review episode. They do not, by themselves, show that the present use remains ethically continuous with the conditions under which persons, communities, clinicians, public institutions, and source organizations entrusted data or material to health research. Careful reviewers acknowledge as much, yet the structure of review tends to reward the signals that can be filed and audited over those that must be interpreted [5-7,12,13]. The asymmetry is structural rather than a matter of individual diligence: filed artifacts survive audit, while interpretive judgments about meaning and standing rarely leave a trace that a later inquiry can reward.

Ethical continuity refers to the persistence of normatively relevant relationships between the original conditions of data generation or material collection and subsequent downstream uses, such that current uses remain justifiable in light of the expectations, permissions, meanings, and relational obligations under which the original collection occurred. This definition is deliberately broader than consent. It includes consent where consent is ethically and legally relevant, but it also includes public justification, statutory authority, proportionality, semantic validity, community expectations, and relations of trust. Framing the question as one of continuity rather than compliance keeps attention on the substantive relationship that governance is meant to protect, and it aligns with recent arguments that ethics frameworks for big data and for AI must attend to more than the presence of principles or paperwork [14-16]. *Continuity* is therefore a property of a relationship over time, not of a document, and it can lapse even where every document remains valid.

The aim of this viewpoint is to define and operationalize the Continuity Trap in data science health research. The intended audience is those who decide whether secondary use may proceed. They include research ethics committees, data-access committees, repositories, health-system AI governance bodies, funders, journals, regulators, investigators, and developers. Our central claim is that ethical continuity is domain-specific. It must be assessed across provenance, semantics, authorization, and relational standing domains. The operational goal is not to require full rereview for every request for secondary data use. It is to prevent one visible continuity signal from closing inquiry into the other domains too early. We proceed by situating the trap among adjacent governance critiques, defining it precisely, specifying its 4 domains and the institutional mechanisms that produce it, testing it against regulatory frameworks and a set of consent and nonconsent vignettes, and setting out a proportionate policy response (Textbox 1). This formulation builds on the Representational Veracity framework, institutional sociology, audit studies, contextual integrity, and the AI and data-ethics literature on life cycle harms, documentation, fairness, and governance [8,13,14,17-19]. The structure of the argument is cumulative: each later section presupposes the domain distinctions introduced earlier, and the vignettes are meant to test the framework rather than merely illustrate it.

Textbox 1. Key terms and concepts.**Ethical continuity**

- Persistence of normatively relevant relationships between original collection conditions and later uses, such that current uses remain justifiable in light of expectations, permissions, meanings, and relational obligations present at entrustment.

Continuity Trap

- A review-stage governance error in which a salient signal of continuity in one domain is treated as sufficient evidence of ethical continuity overall, causing inquiry into other continuity domains to close prematurely.

Continuity signal

- A visible artifact or feature, such as provenance, broad consent, repository approval, locality, or model documentation, that suggests continuity in one domain but may not establish ethical continuity overall.

Cross-domain closure

- An inference error in which continuity in one domain closes inquiry into other domains, including semantics, authorization, relational standing, or provenance itself.

Provenance continuity

- Continuity of lineage, handling, access history, transformation, and traceability between earlier and present data-stages.

Semantic continuity

- Continuity between labels, descriptors, ontologies, variables, and the constructs or populations they are taken to represent.

Authorization continuity

- Continuity between the current use and the permissions, expectations, legal authorities, and justificatory scope under which data or material were originally entrusted or collected.

Relational standing

- The continuing capacity of affected persons, communities, or source institutions to contest, shape, govern, or share in value when downstream uses create group-level claims, risks, or benefits.

Data-stage

- The specific state of a dataset, biospecimen, model, derivative, embedding, cell line, or deployment at the point of review.

Domain-matched safeguard

- A safeguard chosen because it addresses the weakest continuity domain, rather than adding generic paperwork.

Existing Governance Problems

The Continuity Trap is related to established critiques of governance, but it does not collapse into any of them (Table 1). Proxy closure is the broader genus: a partial and administratively convenient proxy is treated as if it exhausts a more complex moral or institutional object. In health data science, a provenance log can stand in for ethical legitimacy, a consent document can stand in for present authorization, and a model card can stand

in for accountability. The proxy may be accurate as far as it goes, but the error lies in treating it as complete. What distinguishes the Continuity Trap from proxy closure in general is its object and its timing. The proxied object is specifically the continuity of an entrustment relationship over time, and the closure occurs at an identifiable review episode where a decision to proceed is made [5,6,17,20]. A repository approval, for instance, records that access was granted under stated terms, but it says nothing about whether those terms still fit a use that has since changed in purpose or population.

Table 1. The Continuity Trap compared with adjacent governance concepts.

Concept	What is overread or optimized?	Unit of failure	Why it matters for secondary use
Proxy closure	A partial proxy is treated as if it exhausts a complex value.	Any governance or evaluative setting.	Broader genus: legibility substitutes for fidelity.
Continuity Trap	A continuity signal in one domain is treated as sufficient evidence of ethical continuity overall.	Secondary-use review episode.	Continuity-specific subtype: provenance, broad consent, locality, or documentation closes inquiry into semantics, authorization, and relational standing.
Goodhart/Campbell dynamics	A measure becomes a target, and optimization degrades its relation to the underlying goal.	Incentive or compliance regime.	Dynamic intensifier once continuity signals become targets.
Ethics creep	Ethics review expands to more activities and contexts.	Institutional jurisdiction.	Important background risk, but the trap concerns premature closure within review.
Organizational decoupling	Formal compliance structures persist while substantive practice is weakly connected to them.	Organizational practice.	Explains how continuity review can appear complete while the relevant relation remains underexamined.

Goodhart’s and Campbell’s laws describe a related dynamic in which, once a measure becomes a target, actors optimize toward the measure and weaken its relation to the value it was supposed to track. A health system rewarded for complete documentation may produce excellent model cards while leaving unresolved whether descriptors are meaningful, whether public-health data have been repurposed beyond their justification, or whether source communities have recourse. These dynamics intensify the Continuity Trap once provenance completeness, locality preservation, or documentation becomes the compliance target, because the very thoroughness of the auditable record can make the unexamined domains harder to see. The 2 ideas are nonetheless distinct. Goodhart and Campbell describe the gradual corruption of a measure under incentive pressure, whereas the Continuity Trap can occur in a single, good-faith review in which no measure has yet been gamed [21,22]. In continuity terms, the danger is that a target once adopted for good reasons quietly redefines an ethical question as a documentary one.

Ethics creep and organizational decoupling explain the institutional background. Ethics creep concerns the expansion of review into more activities and contexts; organizational decoupling describes the separation between formal compliance structures and actual practice, in which organizations maintain the appearance of conformity while substantive practice drifts. The Continuity Trap is a narrower concept than either of these. It names an inference failure inside a review episode, where the review closes because one continuity signal is asked to do governance work that properly belongs to other domains. It is compatible with a review system that is neither overextended nor decoupled, which is why naming it adds something the existing critiques do not already supply [18,19,23,24]. Keeping the concept narrow is deliberate, because a diagnosis that applies everywhere offers reviewers no guidance about where to look first.

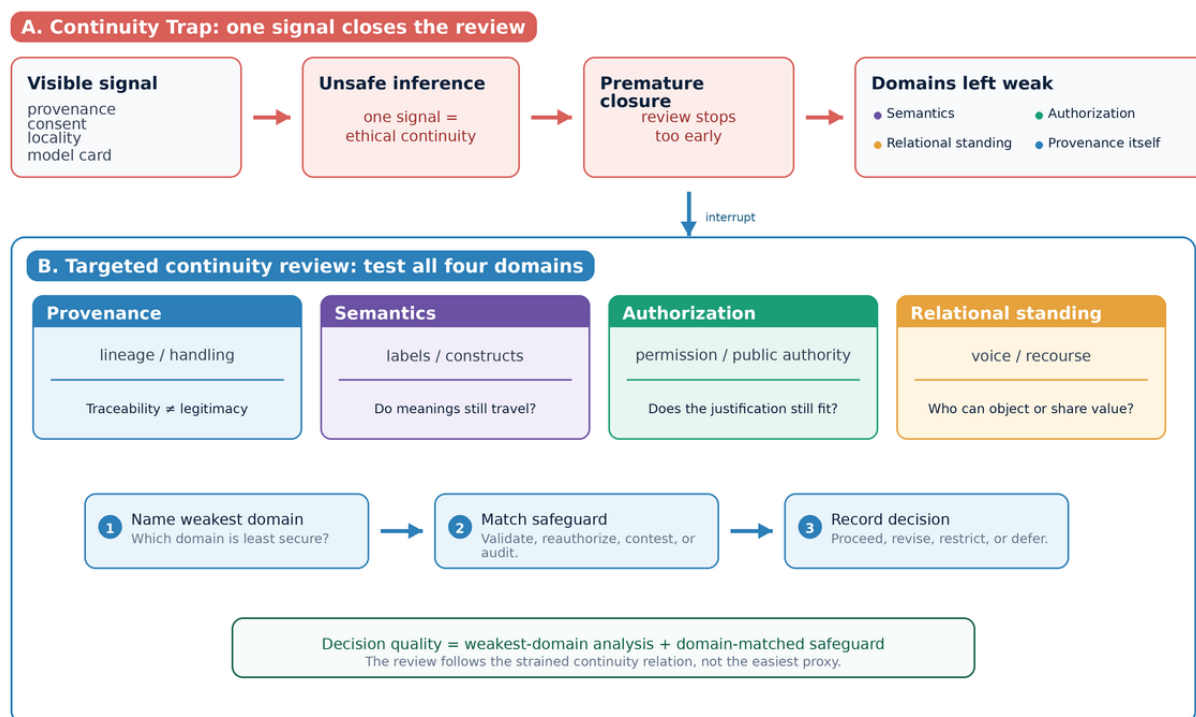
Defining the Continuity Trap

The Continuity Trap is a continuity-specific form of proxy closure, and it has 3 features. First, the proxied object is ethical

continuity in secondary data use, and not quality, fairness, privacy, or compliance in the abstract. Second, the unit of failure is the review episode: an access decision, repository approval, derivation decision, transfer authorization, model-release decision, or deployment clearance. Third, the error is cross-domain closure. Continuity in one domain, most commonly the provenance domain, but sometimes consent, repository approval, locality, or model documentation, is treated as sufficient reason to stop inquiry into the other domains. Understood this way, the trap is a systems-level failure of inference rather than a failure of any single instrument, which is why remedies aimed only at improving the instrument tend to leave it intact [20,25]. Because the failure is one of inference rather than of any single record, adding another record rarely dislodges it; what changes the outcome is a reviewer asking which domain the visible signal has been allowed to speak for.

This definition avoids 2 unhelpful responses. The first is to assume that any secondary use requires full rereview. That would be impractical, would slow legitimate research, and would disadvantage institutions with fewer administrative resources. The second is to assume that legal access or adequate documentation settles the ethical question. That approach is administratively attractive but ethically thin. The Continuity Trap occupies the space between these positions. Reviewers should identify when the present data-stage has changed enough that continuity must be reexamined, and should specify which domain is weak, rather than defaulting either to blanket rereview or to documentary sufficiency [26,27]. The middle position asks more of reviewers than a checklist and less than a full rereview, which is what makes it usable under real institutional constraints.

Figure 1 summarizes the inferential structure. A visible continuity signal may be accurate in its own domain, but the inference from that signal to ethical continuity overall is unsafe unless the other domains have also been assessed. The corrective action is a targeted one. Identify the weakest domain and require a safeguard that addresses that domain.

Figure 1. Inferential structure of the Continuity Trap.

The 4 Domains of Ethical Continuity

Ethical continuity is not monolithic. It is relation-specific across 4 domains: provenance, semantics, authorization, and relational standing, developed in the Representational Veracity framework [14]. The data-stage may be a linked dataset, a deidentified file, a trained model, an aggregate embedding space, a derived cell line, a commercial derivative, or a deployment. The governance question is whether the relationships that made earlier collection or use justifiable remain strong enough at the later data-stage to support the decision being made now. Separating the domains is what allows a reviewer to say precisely where continuity has weakened, instead of reaching a single undifferentiated judgment that the use is or is not acceptable [14]. A dataset that was semantically coherent when collected for one clinical purpose can become semantically unstable once it is linked, harmonized across sites, or repurposed as a model input, even though its provenance record remains unbroken.

The domains can diverge, and it is their divergence that the trap exploits. A model may be traceable yet semantically unstable. A dataset may be legally accessible yet relationally extractive. A deidentified file may satisfy a privacy standard yet remain vulnerable to reidentification through linkage or generative inference, so that provenance and privacy assurances mislead about residual risk. A public-health registry may be authorized for surveillance but unauthorized for immigration enforcement. A federated architecture may preserve locality while transforming clinical records into training inputs for a commercial model. A model card is relevant evidence for documentation and accountability, but it is not relational standing; a consent form is relevant to authorization, but it is not authorization continuity itself [13,28,29]. The point is not that any one signal is untrustworthy, but that a signal true in its own domain licenses no inference about the others without independent examination (Table 2).

Table 2. Domains of ethical continuity and proxies that commonly mislead review.

Domain	What it tracks	Visible proxy that can mislead	Load-bearing review question
Provenance	Lineage, handling, access history, and transformations.	Audit trail, repository approval, data-use agreement, ingestion log.	What does traceability establish, and what does it not establish, about meaning, authorization, and accountability?
Semantics	Descriptors, ontologies, labels, variables, and inferred constructs.	Stable variable names, harmonized labels, ancestry labels, model-feature names.	Do current categories still map onto the same constructs, persons, populations, or groups?
Authorization	Permissions, expectations, statutory authority, public justification, proportionality, and justificatory scope.	Broad consent, legacy IRB ^a approval, public-health authority, license terms, emergency exemption.	Is the present use still within the justified scope of entrustment or public authority?
Relational standing	Voice, recourse, contestability, governance participation, and claims to benefit when group-level claims or harms arise.	Naming a source group, token consultation, diversity language, community acknowledgment.	Who can object, shape conditions, participate in oversight, or share in value?

^aIRB: institutional review board.

Operational Conditions for Relational Standing

Relational standing should not be invoked whenever individual consent is thin. It becomes governance-relevant when 4 conditions are met: group-level claims are foreseeable; an affected collectivity is reasonably identifiable; likely harms or benefits are nontrivial; and a plausible route to representation exists or can be constructed. The fourth condition requires care. Communities with weak organizational infrastructure should not lose standing because they lack formal representatives, since that would penalize precisely the groups most exposed to extractive reuse. When the first 3 conditions are present but no representational route exists, that absence should itself be documented as a governance finding rather than used to bypass review [26,27,30]. Constructing a route to representation, where none yet exists, is itself a legitimate governance task rather than a precondition that affected groups must satisfy on their own.

This approach is especially important in cross-border and low-resource settings, where data and materials often move from communities with little bargaining power to institutions with substantial analytic and commercial capacity. Communitarian moral philosophy gives the relational domain a normative foundation by emphasizing identification, solidarity, reciprocal recognition, and shared benefit. On that view, extracting value from community-generated data without practical voice, benefit, or recourse is not only an autonomy problem. It also weakens the relationships through which data, identity, and social meaning were constituted, and it corrodes the trust on which future participation depends. Accounts of solidarity in biomedicine make the same point from a different direction, treating willingness to share in one another’s costs as a resource that governance can either sustain or deplete [31–35]. Where benefit and voice are absent, what appears to be permissible reuse can amount to a transfer of value away from the very people whose lives the data describe.

Institutional Mechanisms That Produce the Continuity Trap

The Continuity Trap is produced by recurrent institutional mechanisms that give disproportionate weight to what is most auditable. Provenance privilege arises because lineage is legible, serializable, and defensible in post hoc accountability inquiries. Descriptor sedimentation occurs when old labels become embedded in forms, repositories, metadata schemas, and analytic pipelines, so that categories persist long after the constructs they name have shifted. Authorization fossilization occurs when formal permissions outlive the ethical coherence of the uses they are asked to justify. Community effacement occurs when groups remain visible as sources of diversity, legitimacy, biological material, or epidemiological signal but disappear as participants in downstream governance. Each mechanism is a rational response to institutional incentives, which is why exhortation alone does not remove them [17,18,26,33,34]. None of these mechanisms requires bad faith; each is the predictable result of asking reviewers to account for their decisions using the artifacts an institution is equipped to store and defend.

These mechanisms reinforce one another. What can be documented receives more institutional weight than what must be interpreted, negotiated, validated locally, or sustained over time, and the accumulation of auditable records can itself become the evidence that further inquiry is unnecessary. The Continuity Trap is the point at which that preference hardens from a habit into a decision rule, so that the presence of a strong record in one domain is taken, without further argument, to license the decision to proceed [17,18]. At that point the record has stopped serving inquiry and has begun to substitute for it.

Regulatory and Governance Context

Regulatory and governance frameworks can reduce risk and still induce continuity traps. US Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation, the European Health Data Space, US Food and Drug Administration guidance for AI-enabled device software, the US National Institute of Standards and Technology (NIST) AI Risk Management Framework, the National Academy of

Medicine Artificial Intelligence Code of Conduct, World Health Organization guidance, institutional health-AI governance, and federated-learning arrangements all emphasize accountability, security, traceability, life cycle management, or responsible use in different ways [36-46]. These are real strengths, and nothing in the following argument is a case for weakening them. The risk is that institutions optimize toward what regulators and auditors make most explicit, and treat satisfaction of an explicit requirement as though it answered questions the requirement was never designed to reach. The same explicitness that makes a regime auditable can also narrow the questions a reviewer feels authorized to ask, so that what falls outside the checklist falls outside the review.

This regulatory critique is deliberately modest. HIPAA is not a theory of semantic validity; Food and Drug Administration device guidance is not a community-governance code; the NIST

framework is not a substitute for human-participant review; and World Health Organization surveillance guidance is not a license for unrelated enforcement use. The problem arises when compliance with one regime is treated as if it settled continuity relations that the regime was not designed to evaluate. A complete processing record, a valid waiver, a secure-aggregation protocol, or a model-governance template may be accurate and still leave the central ethical question unresolved. Work on regulating adaptive and continually learning systems, and on taking a system rather than a product view of medical-device software, makes a parallel point: governance has to track the object as it changes, not only certify it at a moment of approval [25,47,48]. Governance that certifies an object once, and treats the certificate as durable, is poorly matched to systems whose purpose, population, and behavior continue to change after approval (Table 3).

Table 3. Examining governance regimes through a continuity lens.

Regime or framework	Primary governance strength	Continuity risk	Supplement needed
HIPAA ^a /public health authority	Permitted disclosures, authorization, waiver, deidentification, data-use agreements, and public-health disclosure rules.	Legal permission may be mistaken for semantic or relational continuity.	Review purpose shift, descriptor validity, and downstream accountability beyond disclosure legality.
GDPR ^b /EHDS ^c	Lawful basis, accountability, records of processing, rights, access, and regulated secondary use of electronic health data.	Processing records may document activity without resolving group-level meaning or standing.	Embed continuity questions into records, impact assessments, access decisions, and transfers.
FDA ^d AI-enabled device guidance	Life cycle management, planned modifications, validation, and safety/effectiveness assessment.	Technical monitoring may not identify authorization drift or descriptor sedimentation.	Continuity review at modification and deployment-change points.
NIST ^e , NAM ^f , and WHO ^g AI guidance	Risk management, responsible and equitable health AI, impacted-person engagement, and LMM-specific governance.	High-level principles may not expose review-stage cross-domain closure.	Domain-specific continuity prompts within AI governance workflows.
Federated-learning governance	Locality preservation, privacy-preserving computation, aggregation controls, and distributed infrastructure.	“Data stayed local” may be mistaken for authorization, semantic, and relational adequacy.	Site-level semantic validation, purpose review, benefit-sharing, and contestability.

^aHIPAA: US Health Insurance Portability and Accountability Act of 1996.

^bGDPR: General Data Protection Regulation.

^cEHDS: European Health Data Space.

^dFDA: US Food and Drug Administration.

^eNIST: US National Institute of Standards and Technology.

^fNAM: US National Academy of Medicine.

^gWHO: World Health Organization.

Authorization Continuity Beyond Consent-Based Frameworks

Authorization continuity is often read too narrowly as consent continuity. Public health surveillance shows why that reading is incomplete. Surveillance, disease notification, immunization registries, syndromic systems, wastewater monitoring, occupational surveillance, and emergency exemptions may be ethically authorized without individual informed consent when they are necessary for a legitimate public-health purpose, proportionate to that purpose, governed by safeguards against misuse, and oriented to the common good. These are long-standing conditions in public-health ethics, and they locate the authority for nonconsensual collection in public legitimacy

rather than in the absence of objection [42,49-54]. Cancer registries, notifiable-disease reporting, and newborn screening rest on the same logic, and each carries a purpose limitation that the initial authorization was designed to serve.

The absence of consent does not mean the absence of authorization. What it means is that authorization rests on public legitimacy, statutory mandate, proportionality, transparency where feasible, safeguards against misuse, and the maintenance of trust. The continuity question is whether that structure still supports the present use. Data ethically collected for immunization tracking, outbreak detection, or communicable-disease control can lose authorization continuity when they are later used for a materially different purpose, because the justification that licensed collection without consent was tied to the original public-health aim and does not travel

automatically to new aims [49,50]. The relevant test is not whether a legal basis can still be cited, but whether the public justification that licensed collection extends to the use now proposed.

Consider an immunization information system. The original system may be defensible without individual consent because immunization surveillance protects population health and supports continuity of care. Repurposing the same data for immigration enforcement changes the justificatory structure. The problem is not merely that consent was absent; it is that the new use no longer fits the expectations, proportionality, or trust relationship that made nonconsensual public-health collection legitimate, and that visible reversals of this kind can deter participation in the surveillance systems on which population health depends. Contextual integrity explains this type of failure. Privacy and authorization depend on context, actors, purposes, and transmission norms, and not only on the type of data at issue, a lesson reinforced by debates over the responsible use of digital data during public-health emergencies [13,42,49,55]. Investigative genetic genealogy, in which forensic actors query consumer or research genomic databases, is a further instance in which lawful technical access coexists with a serious break in authorization and relational continuity.

Case Vignettes

Overview

The following vignettes show the same inferential pattern across distinct data-stages, in each of which one continuity signal is visible and true while other domains are weak. They are chosen to span material, computational, organizational, and generative transformations, so that the pattern is not mistaken for a feature of any single technology. In each case the corrective action is the same in form even though the endangered domain differs: identify the signal doing the justificatory work, and test whether the other domains actually support the decision [8]. Their order moves from the most material transformation to the most abstract, so that the reader can watch the same inference recur as the object of governance becomes progressively less tangible.

Polygenic Risk Scores

Polygenic risk scores show why the Continuity Trap must be distinguished from technical transportability. Linkage disequilibrium, allele-frequency differences, effect-size heterogeneity, and gene-environment interaction help explain why many scores perform differently across populations, and why scores developed largely in populations of European ancestry can transfer poorly to others. Those are technical and epidemiological problems. The Continuity Trap appears when a well-documented genome-wide association pipeline, ancestry adjustment, or validation workflow is treated as if it were sufficient governance warrant to move a score into new populations or clinical settings without reopening semantic and relational questions [56-58]. Treating a validated score as portable because its construction was rigorous confuses the reliability of the pipeline with the appropriateness of the categories and permissions it carries into a new setting.

Computational continuity is the persistence of technical reproducibility across analytic stages: the same pipelines, quality-control steps, ancestry adjustments, and validation procedures can be consistently applied. Governance continuity is different. It concerns whether categories, permissions, social assumptions, and relationships remain appropriate in new populations or deployment settings. Population descriptors such as African, European, Black, Yoruba, ethnicity, race, and ancestry do not do identical work across contexts, and treating them as stable technical variables can import social categories into clinical inference without warrant. Their responsible use requires explicit descriptor justification, limits on population claims, and documentation of context and uncertainty, so that a reproducible score is not mistaken for an interpretable or transferable one [59-62]. A descriptor that was a reasonable proxy in one population may encode a different construct in another, so semantic stability has to be argued rather than assumed.

Legacy Biospecimens and Induced Pluripotent Stem Cell Derivation

Legacy biospecimens transformed into induced pluripotent stem cell lines reveal the trap in a limiting case. Here the material continuity is exceptionally strong: the derivative is biologically traceable to the donor and retains genomic continuity. That strength can mislead review bodies into treating the cell line as an ethically continuous extension of the original specimen. Yet reprogramming changes the horizon of possible use, extending it to organoids, gene-edited derivatives, disease models, screening platforms, repository distribution, and commercial applications that the original donors are unlikely to have contemplated [63-66]. The consent that governed a stored specimen was given for a bounded set of uses, and biological traceability does not extend that consent to uses the donor could not have anticipated.

After derivation and distribution, induced pluripotent stem cell lines may become indefinitely self-renewing and difficult to withdraw from global circulation, which weakens the practical meaning of consent-based withdrawal rights once lines have propagated across laboratories. The continuity issue is not that such research is impermissible. It is that strong biological continuity can conceal weak authorization and relational continuity. Sustained engagement, local scientific participation, benefit-oriented governance, clear derivative-use limits, and attention to community expectations can interrupt the trap, and dynamic-consent and dynamic-governance models have been proposed for exactly this reason [64-66]. Governance that engages the source community early is also more robust, because it does not depend on the assumption that a widely distributed line can later be recalled.

Federated Learning

Federated learning privileges locality. The statement that raw data remain at the local institution is often true and privacy-relevant, and it has made multi-institutional collaboration feasible where data transfer would not be. It does not answer whether clinical records generated for care may legitimately function as distributed training inputs for another purpose, whether diagnostic categories are commensurable

across sites, whether lower-resource institutions bear disproportionate burdens in producing a global model, or whether patients and contributing sites have recourse when downstream products are commercialized or deployed asymmetrically [67-70]. Locality answers a transfer question; it does not answer a purpose question, and the two are easy to conflate when the architecture is presented as privacy-preserving by design.

Differential privacy, secure aggregation, gradient auditing, and bring-code-to-data arrangements reduce confidentiality and transfer risks, and privacy-preserving federated methods have advanced quickly in medical imaging. Their limitation is domain specificity. They reduce risks in the provenance and confidentiality space; they cannot decide whether the new purpose is within clinical trust, whether labels mean the same thing at each site, or whether source institutions have relational standing. The assurance that the data never moved can therefore become a proxy for the quite different claim that the governance problem is solved [68,71]. Confidentiality and legitimacy are different goods, and a method can secure the first while leaving the second untouched.

Health-Related Large Language Models

Health-related large language models intensify the Continuity Trap because the distance between input documentation and downstream influence is difficult to govern. A model trained on a governed, research-consented clinical dataset presents a different problem from one trained retrospectively on a large, opaque, cross-border body of data that happens to contain health-relevant information. The literature on stochastic parrots, World Health Organization guidance on large multimodal models, reviews of privacy and ethics in health large language models, and recent assessments of their clinical promise all stress that documentation, safety testing, and removability are necessary but incomplete governance instruments [43,72-75]. The governance difficulty is that influence in these systems is diffuse: no single training document determines an output, so

tracing a downstream harm back to the conditions of its source is far harder than for a conventional dataset.

Developers may document broad categories of source material without showing how particular persons or communities influenced model behavior, how downstream harms can be traced to source conditions, or what route exists for contesting a use. Machine unlearning may improve technical removability, but technical removability is not governance adequacy, and the ability to delete a training record does not restore a person's standing to object to how a deployed model behaves. The Continuity Trap lies in treating input documentation as if it were a sufficient account of authorization, relational standing, and accountability, which is why calls for regulatory oversight of these models emphasize postdeployment obligations rather than documentation alone [5,6,47,76,77]. Oversight therefore has to attach to the deployed system and its effects, not only to the provenance of its training corpus.

Toward Trigger-Based Continuity Review

The response to the Continuity Trap should be rigorous but parsimonious. Continuity review should be triggered when the data-stage undergoes a qualitative transformation in purpose, scope, inferential reach, or derivative status (Textbox 2). Routine reuse within the scope contemplated by the original data-access agreement should not ordinarily trigger review. A replication study using the same variables and population descriptors, a new analysis within the same research program, or internal quality-improvement work that does not generate group-level claims or commercial outputs would usually fall outside the trigger set. The distinction is between quantitative extension and qualitative transformation, and it is meant to concentrate scarce review capacity where the relationship of entrustment is actually at risk [25,44]. Framing review around triggers also gives investigators advance notice of what will occasion scrutiny, which is fairer and more predictable than review that appears to depend on the disposition of an individual committee.

Textbox 2. Trigger conditions and continuity assessment questions.

Trigger conditions

- Hard-to-reverse transformation of source data or material.
- Repurposing of descriptors across populations, sites, ontologies, or historical periods.
- Expanded uses, including commercialization, cross-border transfer, public-to-private transfer, or law-enforcement use.
- Outputs likely to generate group-level claims, sensitive inferences, deployment asymmetries, or reputational harms.
- Absence of a credible pathway for accountability, contestation, local validation, benefit-sharing, or meaningful withdrawal.

Continuity assessment questions

- What is the present data-stage?
- Which visible continuity signal is doing most of the justificatory work?
- What evidence supports provenance, semantic, authorization, and relational continuity?
- Which domain is weakest or most uncertain?
- What feasible safeguard is matched to that domain? What unresolved tension remains, and who will monitor it?

The Continuity Statement

For triggered cases, investigators or developers should submit a short continuity statement together with existing ethics, data-access, material-transfer, repository, or AI governance documentation. The statement is not a self-certification. The secondary data users draft it; reviewers evaluate it and may disagree with the domain assessment. If a trigger condition is present and the investigator asserts that all domains are secure, reviewers should treat the mismatch itself as a reason for scrutiny, in the same spirit as documentation instruments that ask producers to state the intended and unintended uses of a dataset or model [5,6,11]. A continuity statement of a page or two, attached to what already exists, is enough to make the reviewer's task tractable and the investigator's reasoning legible.

The safeguard must match the endangered domain. If semantic continuity is weak, the response should be descriptor

justification, local validation, uncertainty disclosure, subgroup performance evaluation, or limits on population claims. If authorization continuity is weak, the response may be renewed notice, reengagement, restriction on derivative applications, or explicit limitation on use. If relational standing is weak, appropriate responses include advisory review, joint oversight, benefit-sharing provisions, contestability pathways, local deployment conditions, or commercialization limits. If provenance itself is weak, lineage reconstruction and transformation auditing are the relevant safeguards. Adding a further generic form or a further layer of documentation, when the weak domain is semantic or relational, is precisely the failure the framework is designed to avoid [26,27,59-61]. The governing question at each step is not whether more can be documented, but whether the specific relationship that has weakened is the one the safeguard actually repairs. [Textbox 3](#) presents a worked example of a continuity statement.

Textbox 3. A worked example of a continuity statement. Federated learning sepsis model.

Present data-stage - A federated model trained across 5 hospitals using local electronic health records to predict sepsis deterioration, with raw data remaining at each site and a global model shared across the network. Strongest domain - Provenance/locality. Each site can document local data sources, extraction scripts, access logs, aggregation governance, and secure aggregation procedures. Weakest domain - Semantic continuity. "Sepsis," "suspected infection," "organ dysfunction," antibiotic timing, laboratory availability, and intensive care unit thresholds differ across hospitals. Authorization concern - Clinical data were generated for care. Network-level model training and possible licensing require explicit institutional justification and limits on secondary commercialization. Relational-standing concern - Smaller hospitals contribute data and bear implementation burdens but may have little influence over model updates, deployment thresholds, revenue sharing, or public communication of performance. Domain-matched safeguards - Require site-level label validation; report performance and calibration by site; restrict claims that the model is "network validated" until semantic heterogeneity is resolved; require network governance representation from all contributing sites; prohibit external commercialization without renewed authorization review. Review decision - Proceed with staged development, not full deployment. Approval is conditional on semantic validation and relational-governance safeguards, not additional provenance documentation alone.

Institutional Roles and Burden

Continuity-sensitive governance should be distributed across the institutions that actually control access, deployment, and value capture. Investigators and developers should characterize the data-stage, identify the continuity signal on which they rely, justify their descriptors, map downstream uses, and propose domain-matched safeguards. Research ethics committees, data-access committees, repositories, material-transfer

authorities, and AI governance bodies should evaluate whether the weakest domain has been correctly identified. Funders and journals can require disclosure of continuity triggers and safeguards. Contracts can carry derivative-use restrictions, downstream reporting obligations, joint-governance provisions, and benefit-sharing conditions [27,44-46,78]. Distributing the work in this way also keeps any single body from being held to account for a life course it never controlled, which is often what drives defensive overdocumentation.

The continuity statement should be concise and should reuse existing forms wherever possible, so that the framework lowers rather than raises the total burden of review by redirecting effort from generic paperwork to the domain under strain. No single committee can govern the entire life course of a dataset, model, or biospecimen. The framework therefore assigns tasks to the institution with practical control over the relevant data-stage: the repository for access, the review board for human-participant and authorization issues, the material-transfer office for derivative material, the AI governance body for deployment, and the contracting parties for downstream restrictions. Low-resource settings should not be asked to perform full sociological reconstruction for every request; triggered review should rely on consortium-level descriptor metadata, standing advisory structures, shared templates, and distributed review responsibilities, consistent with stewardship principles that make data findable, accessible, interoperable, and reusable without treating reuse as automatically authorized [25,78]. The aim is to relocate effort, not to add a stage, so that a triggered review spends its attention on the one or two domains that a routine filing would otherwise leave unexamined.

Conclusion

The Continuity Trap names a recurrent failure in the ethical governance of data science health research: institutions infer

too much from the continuity signal they can most easily verify. Provenance logs, repository approvals, consent documents, public-health authorities, model cards, and locality-preserving architectures all matter, and each answers a genuine governance question. None can certify ethical continuity alone. Ethical continuity is distributed across provenance, semantics, authorization, and relational standing, and these domains can diverge as data, materials, descriptors, models, and institutions change [10]. Recognizing that the domains are distinct is what keeps a single reassuring signal from being mistaken for the whole of ethical continuity.

The value of naming the trap is practical. It allows reviewers to distinguish routine secondary use from qualitative transformation; to identify which continuity domain is carrying too much justificatory weight; and to impose safeguards that match the domain under strain. That response is more precise than universal rereview and stronger than generic compliance documentation. Data science health research is often most vulnerable not when documentation is absent, but when documentation is so reassuring that institutions stop asking harder questions. Provenance should begin ethical review; it should not end it [12].

Acknowledgments

The authors declare the use of generative AI (GenAI) tools in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: evaluation of novelty of the concept and if something similar has been previously published, word count, copyediting, reference checking, and development of visual and tabular elements. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

The authors would like to thank the members of the BridgeELSI Project: Peter A. Ikhane PhD, MSc; Charlissee Caga-Anan JD, MA; Adeola Akintola MSc; Tobiloba Oyediran MSc; Oluchi C. Maduka BL, LLB, PhD; Temilola Yusuf MPH; Sally N. Adebamowo MBBS, MSc, ScD; Simisola Akintola BL, LLM, PhD; Temidayo Ogundiran MBBS, MHSc; Ayodele Jegede PhD; Olusegun Adeyemo MSc; Shawneequa Callier JD, MA; Clement Adebamowo BM ChB Hons, ScD; Muhammad K. Imam-Tamim PhD; Ibrahim Uthman PhD.

Data Availability

No primary data were generated or analyzed in this study. All sources cited are publicly available.

Funding

This project is supported by the Bridging Gaps in the Ethical, Legal, and Social Implications of Data Science Health Research in Nigeria (BridgeELSI) grant (NIH/NIMH U01MH127693). Additional support was received from the Maryland Department of Health's Cigarette Restitution Fund Program (CH-649-CRF); the Comprehensive Polygenic Risk Profiling Across Multiple Health Outcomes (CARDINAL, NIH/NHGRI 01HG011717); the Knowledge, Attitude and Recommendations of Community members, Researchers, and Bioethicists on creation of induced pluripotent stem cell from YRI HapMap samples (ENRICH Project; grant R25TW011811-S1); and the University of Maryland Greenebaum Comprehensive Cancer Center Support Grant (NIH/NCI P30CA134274). The funding agencies did not play any role in the publication.

Authors' Contributions

CA conceived the framework and led manuscript development. SNA, AA, PI, SA, TO, AJ, OA, SC, MI-T, and IU contributed to conceptual development and critical revision. All authors approved the final version. CA is the guarantor.

Conflicts of Interest

None declared.

References

1. Moulaei K, Akhlaghpour S, Fatehi F. Patient consent for the secondary use of health data in artificial intelligence (AI) models: a scoping review. *Int J Med Inform.* 2025;198:105872. [FREE Full text] [doi: [10.1016/j.ijmedinf.2025.105872](https://doi.org/10.1016/j.ijmedinf.2025.105872)] [Medline: [40107041](https://pubmed.ncbi.nlm.nih.gov/40107041/)]
2. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. *N Engl J Med.* 2018;378(11):981-983. [FREE Full text] [doi: [10.1056/NEJMp1714229](https://doi.org/10.1056/NEJMp1714229)] [Medline: [29539284](https://pubmed.ncbi.nlm.nih.gov/29539284/)]
3. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med.* 2018;169(12):866-872. [FREE Full text] [doi: [10.7326/M18-1990](https://doi.org/10.7326/M18-1990)] [Medline: [30508424](https://pubmed.ncbi.nlm.nih.gov/30508424/)]
4. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. [FREE Full text] [doi: [10.1126/science.aax2342](https://doi.org/10.1126/science.aax2342)] [Medline: [31649194](https://pubmed.ncbi.nlm.nih.gov/31649194/)]
5. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model Cards for Model Reporting. Atlanta GA USA.; 2019. Presented at: FAT* '19: Conference on Fairness, Accountability, and Transparency; 2019 January 29 - 31:220-229; Atlanta GA USA. URL: <https://dl.acm.org/doi/10.1145/3287560.3287596> [doi: [10.1145/3287560.3287596](https://doi.org/10.1145/3287560.3287596)]
6. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, III HD, et al. Datasheets for datasets. *Commun ACM.* 2021;64(12):86-92. [doi: [10.1145/3458723](https://doi.org/10.1145/3458723)]
7. Bender EM, Friedman B. Data statements for natural language processing: toward mitigating system bias and enabling better science. *Trans Assoc Comput Linguist.* 2018;6:587-604. [FREE Full text]
8. Suresh H, Gutttag J. A framework for understanding sources of harm throughout the machine learning life cycle. 2021. Presented at: EAAMO '21: Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization; 2021 October 5-9:1-9; NY USA. URL: <https://dl.acm.org/doi/10.1145/3465416.3483305>
9. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med.* 2018;15(11):e1002689. [FREE Full text] [doi: [10.1371/journal.pmed.1002689](https://doi.org/10.1371/journal.pmed.1002689)] [Medline: [30399149](https://pubmed.ncbi.nlm.nih.gov/30399149/)]
10. Price WN, Cohen IG. Privacy in the age of medical big data. *Nat Med.* 2019;25(1):37-43. [FREE Full text] [doi: [10.1038/s41591-018-0272-7](https://doi.org/10.1038/s41591-018-0272-7)] [Medline: [30617331](https://pubmed.ncbi.nlm.nih.gov/30617331/)]
11. Larson DB, Magnus DC, Lungren MP, Shah NH, Langlotz CP. Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology.* 2020;295(3):675-682. [doi: [10.1148/radiol.2020192536](https://doi.org/10.1148/radiol.2020192536)] [Medline: [32208097](https://pubmed.ncbi.nlm.nih.gov/32208097/)]
12. Char DS, Abramoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. *Am J Bioeth.* 2020;20(11):7-17. [FREE Full text] [doi: [10.1080/15265161.2020.1819469](https://doi.org/10.1080/15265161.2020.1819469)] [Medline: [33103967](https://pubmed.ncbi.nlm.nih.gov/33103967/)]
13. Nissenbaum H. Privacy as contextual integrity. *Wash Law Rev.* 2004;79(1):119-158. [FREE Full text]
14. Adebamowo C, Adebamowo S, Akintola A, Ikhane P, Akintola S, Ogundiran T, et al. A Persistence-Based Ethical Governance Framework for Data Science Health Research. SSRN. Apr 03, 2026. [FREE Full text] [doi: [10.2139/ssrn.6515498](https://doi.org/10.2139/ssrn.6515498)]
15. Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY, et al. An ethics framework for big data in health and research. *Asian Bioeth Rev.* 2019;11(3):227-254. [FREE Full text] [doi: [10.1007/s41649-019-00099-x](https://doi.org/10.1007/s41649-019-00099-x)] [Medline: [33717314](https://pubmed.ncbi.nlm.nih.gov/33717314/)]
16. Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nat Mach Intell.* 2019;1(11):501-507. [doi: [10.1038/s42256-019-0114-4](https://doi.org/10.1038/s42256-019-0114-4)]
17. Strathern M. Improving ratings: audit in the British university system. *Eur Rev.* 1997;5(3):305-321. [FREE Full text]
18. Meyer JW, Rowan B. Institutionalized organizations: formal structure as myth and ceremony. *Am J Sociol.* 1977;83(2):340-363. [doi: [10.1086/226550](https://doi.org/10.1086/226550)]
19. Brunsson N. The Organization of Hypocrisy: Talk, Decisions and Actions in Organizations. Chichester, UK. Wiley; 1989.
20. John Y, Caldwell L, McCoy DE, Braganza O. Dead rats, dopamine, performance metrics, and peacock tails: proxy failure is an inherent risk in goal-oriented systems. *Behav Brain Sci.* 2024;47:e67. [doi: [10.1017/S0140525X23002753](https://doi.org/10.1017/S0140525X23002753)]
21. Mattson C, Bushardt RL, Artino AR. "When a measure becomes a target, it ceases to be a good measure". *J Grad Med Educ.* 2021;13(1):2-5. [FREE Full text] [doi: [10.4300/JGME-D-20-01492.1](https://doi.org/10.4300/JGME-D-20-01492.1)] [Medline: [33680290](https://pubmed.ncbi.nlm.nih.gov/33680290/)]
22. Poku M. Campbell's law: implications for health care. *J Health Serv Res Policy.* 2016;21(2):137-139. [doi: [10.1177/1355819615593772](https://doi.org/10.1177/1355819615593772)] [Medline: [26113538](https://pubmed.ncbi.nlm.nih.gov/26113538/)]
23. Haggerty KD. Ethics creep: governing social science research in the name of ethics. *Qual Sociol.* 2004;27(4):391-414. [doi: [10.1023/b:quas.0000049239.15922.a3](https://doi.org/10.1023/b:quas.0000049239.15922.a3)]
24. Bosk CL, De Vries RG. Bureaucracies of mass deception: institutional review boards and the ethics of ethnographic research. *Ann Am Acad Pol Soc Sci.* 2004;595(1):249-263. [doi: [10.1177/0002716204266913](https://doi.org/10.1177/0002716204266913)]
25. Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ Digit Med.* 2020;3:53. [FREE Full text] [doi: [10.1038/s41746-020-0262-2](https://doi.org/10.1038/s41746-020-0262-2)] [Medline: [32285013](https://pubmed.ncbi.nlm.nih.gov/32285013/)]
26. Munung N, Royal CD, de Kock C, Awandare G, Nembaware V, Nguetack S, et al. Genomics and health data governance in Africa: democratize the use of big data and popularize public engagement. *Hastings Cent Rep.* 2024;54 Suppl 2:S84-S92. [doi: [10.1002/hast.4933](https://doi.org/10.1002/hast.4933)] [Medline: [39707935](https://pubmed.ncbi.nlm.nih.gov/39707935/)]

27. Tindana P, Singh JA, Tracy CS, Upshur REG, Daar AS, Singer PA, et al. Grand challenges in global health: community engagement in research in developing countries. *PLoS Med*. 2007;4(9):e273. [FREE Full text] [doi: [10.1371/journal.pmed.0040273](https://doi.org/10.1371/journal.pmed.0040273)] [Medline: [17850178](https://pubmed.ncbi.nlm.nih.gov/17850178/)]
28. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*. 2019;10(1):3069. [FREE Full text] [doi: [10.1038/s41467-019-10933-3](https://doi.org/10.1038/s41467-019-10933-3)] [Medline: [31337762](https://pubmed.ncbi.nlm.nih.gov/31337762/)]
29. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science*. 2013;339(6117):321-324. [doi: [10.1126/science.1229566](https://doi.org/10.1126/science.1229566)] [Medline: [23329047](https://pubmed.ncbi.nlm.nih.gov/23329047/)]
30. Mwaka ES, Bagenda G, Sebatta DE, Nabukenya S, Munabi I. Benefit sharing in genomic and biobanking research in Uganda: perceptions of researchers and research ethics committee members. *Front Genet*. 2022;13:1037401. [FREE Full text] [doi: [10.3389/fgene.2022.1037401](https://doi.org/10.3389/fgene.2022.1037401)] [Medline: [36468002](https://pubmed.ncbi.nlm.nih.gov/36468002/)]
31. Metz T. Toward an African moral theory. *J Polit Philos*. 2007;15(3):321-341. [doi: [10.1111/j.1467-9760.2007.00280.x](https://doi.org/10.1111/j.1467-9760.2007.00280.x)]
32. Molefe M. *An African Philosophy of Personhood, Morality, and Politics*. Cham, Switzerland. Palgrave Macmillan; 2019.
33. Couldry N, Mejias UA. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA. Stanford University Press; 2019.
34. Salami A. Artificial intelligence, digital colonialism, and the implications for Africa's future development. *Data Policy*. 2024;6:e67. [FREE Full text]
35. Prainsack B, Buyx A. *Solidarity in Biomedicine and Beyond*. Cambridge, UK. Cambridge University Press; 2017.
36. US Department of Health and Human Services. Disclosures for public health activities. [45 CFR 164.512(b)]. HHS.gov. 2013. URL: <https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/disclosures-public-health-activities/index.html>
37. European Union and Council. Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Union*. 2016;1-88. [FREE Full text]
38. Regulation (EU) 2025/327 of 11 February 2025 on the European health data space. European Union. 2025. URL: <https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng> [accessed 2026-07-25]
39. FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions: Guidance for Industry and FDA Staff. Silver Spring, MD. FDA; 2025.
40. Tabassi E. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD. National Institute of Standards and Technology; 2023.
41. National Academy of Medicine Science. Adams L, editor. *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action*. Washington, DC. National Academies Press; 2025.
42. World Health Organization. *WHO Guidelines on Ethical Issues in Public Health Surveillance*. Geneva. World Health Organization; 2017.
43. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*. Geneva. World Health Organization; 2025.
44. Nong P, Hamasha R, Singh K, Adler-Milstein J, Platt J. How Academic Medical Centers Govern AI Prediction Tools in the Context of Uncertainty and Evolving Regulation. *NEJM AI*. 2024;1(3):AIp2300048. [doi: [10.1056/AIp2300048](https://doi.org/10.1056/AIp2300048)]
45. Owens K, Griffen Z, Damaraju L. Managing a 'responsibility vacuum' in AI monitoring and governance in healthcare: a qualitative study. *BMC Health Serv Res*. 2025;25:1217. [FREE Full text] [doi: [10.1186/s12913-025-13388-z](https://doi.org/10.1186/s12913-025-13388-z)]
46. Griffen Z, Owens K. From "Human in the Loop" to a participatory system of governance for AI in healthcare. *Am J Bioeth*. 2024;24(9):81-83. [FREE Full text] [doi: [10.1080/15265161.2024.2377114](https://doi.org/10.1080/15265161.2024.2377114)] [Medline: [39226015](https://pubmed.ncbi.nlm.nih.gov/39226015/)]
47. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. 2023;6(1):120. [FREE Full text] [doi: [10.1038/s41746-023-00873-0](https://doi.org/10.1038/s41746-023-00873-0)] [Medline: [37414860](https://pubmed.ncbi.nlm.nih.gov/37414860/)]
48. Vokinger KN, Feuerriegel S, Kesselheim AS. Continual learning in medical devices: FDA's action plan and beyond. *Lancet Digit Health*. 2021;3(6):e337-e338. [FREE Full text] [doi: [10.1016/S2589-7500\(21\)00076-5](https://doi.org/10.1016/S2589-7500(21)00076-5)] [Medline: [33933404](https://pubmed.ncbi.nlm.nih.gov/33933404/)]
49. Lee LM, Heilig CM, White A. Ethical justification for conducting public health surveillance without patient consent. *Am J Public Health*. 2012;102(1):38-44. [doi: [10.2105/AJPH.2011.300297](https://doi.org/10.2105/AJPH.2011.300297)] [Medline: [22095338](https://pubmed.ncbi.nlm.nih.gov/22095338/)]
50. Klingler C, Silva DS, Schuermann C, Reis AA, Saxena A, Streh D. Ethical issues in public health surveillance: a systematic qualitative review. *BMC Public Health*. 2017;17(1):295. [FREE Full text] [doi: [10.1186/s12889-017-4200-4](https://doi.org/10.1186/s12889-017-4200-4)] [Medline: [28376752](https://pubmed.ncbi.nlm.nih.gov/28376752/)]
51. Centers for Disease Control and Prevention. *IIS policy and legislation*. Centers for Disease Control and Prevention. 2026. URL: <https://www.cdc.gov/iis/policy-legislation/index.html>
52. National Academies of Sciences, Engineering, and Medicine. *Wastewater-Based Disease Surveillance for Public Health Action*. Washington, DC. National Academies Press; 2023.
53. Ni K, Korfmacher KS. A scoping review on the ethics of wastewater surveillance for COVID-19. *J Public Health Manag Pract*. 2025;31(6):E387-E395. [doi: [10.1097/PHH.0000000000002206](https://doi.org/10.1097/PHH.0000000000002206)] [Medline: [40839804](https://pubmed.ncbi.nlm.nih.gov/40839804/)]
54. Persad G, Barnhill A, MacKay D. Displaying wastewater surveillance data: an ethics framework. *J Law Biosci*. 2025;12(1):lsaf001. [FREE Full text] [doi: [10.1093/jlb/lsaf001](https://doi.org/10.1093/jlb/lsaf001)] [Medline: [40491867](https://pubmed.ncbi.nlm.nih.gov/40491867/)]
55. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med*. 2020;26(4):463-464. [FREE Full text] [doi: [10.1038/s41591-020-0832-5](https://doi.org/10.1038/s41591-020-0832-5)] [Medline: [32284619](https://pubmed.ncbi.nlm.nih.gov/32284619/)]

56. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51(4):584-591. [FREE Full text] [doi: [10.1038/s41588-019-0379-x](https://doi.org/10.1038/s41588-019-0379-x)] [Medline: [30926966](https://pubmed.ncbi.nlm.nih.gov/30926966/)]
57. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium Methods Working Group, et al. Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet.* Jan 2024;25(1):8-25. [FREE Full text] [doi: [10.1038/s41576-023-00637-2](https://doi.org/10.1038/s41576-023-00637-2)] [Medline: [37620596](https://pubmed.ncbi.nlm.nih.gov/37620596/)]
58. Duncan L, Shen H, Gelaye B, Meijssen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun.* 2019;10(1):3328. [FREE Full text] [doi: [10.1038/s41467-019-11112-0](https://doi.org/10.1038/s41467-019-11112-0)] [Medline: [31346163](https://pubmed.ncbi.nlm.nih.gov/31346163/)]
59. National Academies of Sciences, Engineering, and Medicine. Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field. Washington, DC. National Academies Press; 2023.
60. Smith JL, Adebamowo CA, Adebamowo SN, Darst BF, Fullerton SM, Gogarten SM, Polygenic Risk Methods Development (PRIMED) Consortium, et al. Recommendations for responsible use of population descriptors in polygenic risk score development. *Nat Genet.* 2025;57(12):2962-2971. [doi: [10.1038/s41588-025-02395-9](https://doi.org/10.1038/s41588-025-02395-9)] [Medline: [41286102](https://pubmed.ncbi.nlm.nih.gov/41286102/)]
61. Khan AT, Adebamowo C, Fullerton SM, Hirbo J, Konigsberg IR, Kraft P, Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium, et al. A data model for population descriptors in genomic research. *Am J Hum Genet.* 2025;112(7):1504-1514. [FREE Full text] [doi: [10.1016/j.ajh.2025.05.011](https://doi.org/10.1016/j.ajh.2025.05.011)] [Medline: [40513563](https://pubmed.ncbi.nlm.nih.gov/40513563/)]
62. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. *BMC Med Ethics.* 2020;21(1):17. [FREE Full text] [doi: [10.1186/s12910-020-0457-8](https://doi.org/10.1186/s12910-020-0457-8)] [Medline: [32075640](https://pubmed.ncbi.nlm.nih.gov/32075640/)]
63. Lowenthal J, Lipnick S, Rao M, Hull SC. Specimen collection for induced pluripotent stem cell research: harmonizing the approach to informed consent. *Stem Cells Transl Med.* 2012;1(5):409-421. [FREE Full text] [doi: [10.5966/sctm.2012-0029](https://doi.org/10.5966/sctm.2012-0029)] [Medline: [23197820](https://pubmed.ncbi.nlm.nih.gov/23197820/)]
64. Isasi R, Bentzen HB, Fabbri M, Fuhr A, Glover JC, Mah N, et al. Dynamic governance: a new era for consent for stem cell research. *Stem Cell Rep.* 2024;19(9):1233-1241. [FREE Full text] [doi: [10.1016/j.stemcr.2024.07.006](https://doi.org/10.1016/j.stemcr.2024.07.006)] [Medline: [39151430](https://pubmed.ncbi.nlm.nih.gov/39151430/)]
65. Ikhane PA, Yusuf T, Adeyemo O, Ogundiran TO, Adebamowo SN, Adebamowo CA. Community and bioethicists' perspectives on iPSC research with biobanked samples collected using broad consent. *Stem Cell Reports.* Dec 09, 2025;20(12):102721. [FREE Full text] [doi: [10.1016/j.stemcr.2025.102721](https://doi.org/10.1016/j.stemcr.2025.102721)] [Medline: [41270747](https://pubmed.ncbi.nlm.nih.gov/41270747/)]
66. International Society for Stem Cell Research. 2025 Guidelines for Stem Cell Research and Clinical Translation. Skokie, IL. International Society for Stem Cell Research; 2025.
67. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med.* 2020;3:119. [FREE Full text] [doi: [10.1038/s41746-020-00323-1](https://doi.org/10.1038/s41746-020-00323-1)] [Medline: [33015372](https://pubmed.ncbi.nlm.nih.gov/33015372/)]
68. Pati S, Kumar S, Varma A, Edwards B, Lu C, Qu L, et al. Privacy preservation for federated learning in health care. *Patterns (N Y).* 2024;5(7):100974. [FREE Full text] [doi: [10.1016/j.patter.2024.100974](https://doi.org/10.1016/j.patter.2024.100974)] [Medline: [39081567](https://pubmed.ncbi.nlm.nih.gov/39081567/)]
69. Suver C, Thorogood A, Doerr M, Wilbanks J, Knoppers B. Bringing code to data: do not forget governance. *J Med Internet Res.* 2020;22(7):e18087. [FREE Full text] [doi: [10.2196/18087](https://doi.org/10.2196/18087)] [Medline: [32540846](https://pubmed.ncbi.nlm.nih.gov/32540846/)]
70. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep.* 2020;10(1):12598. [FREE Full text] [doi: [10.1038/s41598-020-69250-1](https://doi.org/10.1038/s41598-020-69250-1)] [Medline: [32724046](https://pubmed.ncbi.nlm.nih.gov/32724046/)]
71. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell.* 2020;2(6):305-311. [doi: [10.1038/s42256-020-0186-1](https://doi.org/10.1038/s42256-020-0186-1)]
72. Bender EM, Gebru T, McMillan-Major A. On the dangers of stochastic parrots: can language models be too big? 2021. Presented at: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency; 2021 March 3 - 10:610-623; Virtual Event Canada.
73. Fareed M, Fatima M, Uddin J, Ahmed A, Sattar MA. A systematic review of ethical considerations of large language models in healthcare and medicine. *Front Digit Health.* 2025;7:1653631. [FREE Full text] [doi: [10.3389/fdgh.2025.1653631](https://doi.org/10.3389/fdgh.2025.1653631)] [Medline: [41019285](https://pubmed.ncbi.nlm.nih.gov/41019285/)]
74. Zhong X, Li S, Chen Z, Ge L, Yu D, Wang S, et al. Considerations for patient privacy of large language models in health care: scoping review. *J Med Internet Res.* 2025;27:e76571. [FREE Full text] [doi: [10.2196/76571](https://doi.org/10.2196/76571)] [Medline: [41269747](https://pubmed.ncbi.nlm.nih.gov/41269747/)]
75. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* 2023;29(8):1930-1940. [doi: [10.1038/s41591-023-02448-8](https://doi.org/10.1038/s41591-023-02448-8)] [Medline: [37460753](https://pubmed.ncbi.nlm.nih.gov/37460753/)]
76. Bourtole L, Chandrasekaran V, Choquette-Choo CA, Jia H, Travers A, Zhang B, et al. Machine unlearning. *IEEE*; 2021. Presented at: IEEE Symposium on Security and Privacy (SP); 2021 May 24-27:141; San Francisco, CA, USA. [doi: [10.1109/SP40001.2021.00019](https://doi.org/10.1109/SP40001.2021.00019)]
77. Xu H, Zhu T, Zhang L, Zhou W, Yu PS. Machine unlearning: a survey. *ACM Comput Surv.* 2023;56(1):1-36. [doi: [10.1145/3603620](https://doi.org/10.1145/3603620)]
78. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018. [FREE Full text] [doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)] [Medline: [26978244](https://pubmed.ncbi.nlm.nih.gov/26978244/)]

Abbreviations

HIPAA: US Health Insurance Portability and Accountability Act

NIST: US National Institute of Standards and Technology

Edited by S Brini; submitted 17.Apr.2026; peer-reviewed by M Ijaz, C Schmit, S Nelson; comments to author 15.May.2026; revised version received 18.Jul.2026; accepted 23.Jul.2026; published 19.Aug.2026

Please cite as:

Adebamowo C, Adebamowo SN, Akintola A, Ikhane P, Akintola S, Ogundiran T, Jegede A, Adeyemo O, Callier S, Imam-Tamim M, Uthman I, BridgELSI Project as part of the DS-I Africa Consortium

The Continuity Trap in Data Science Health Research

J Med Internet Res 2026;28:e98699

URL: <https://www.jmir.org/2026/1/e98699>

doi: [10.2196/98699](https://doi.org/10.2196/98699)

PMID:

©Clement Adebamowo, Sally Nneoma Adebamowo, Adeola Akintola, Peter Ikhane, Simisola Akintola, Temidayo Ogundiran, Ayodele Jegede, Olusegun Adeyemo, Shawneequa Callier, Muhammad Imam-Tamim, Ibrahim Uthman, BridgELSI Project as part of the DS-I Africa Consortium. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 19.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.