

Commentary

# From Data Stewardship to Model Stewardship: Extending Governance Frameworks for AI Era Health Data Use

Leon Rozenblit<sup>1,2</sup>, JD, PhD; Steven Labkoff<sup>2,3</sup>, MD; Charles Safran<sup>2,4</sup>, MS, MD

<sup>1</sup>Q.E.D. Institute, New Haven, CT, United States

<sup>2</sup>Division of Clinical Informatics, Beth Israel Deaconess Medical Center, Boston, MA, United States

<sup>3</sup>Luminant Consulting, Stamford, CT, United States

<sup>4</sup>Department of Medicine, Harvard Medical School, Boston, MA, United States

**Corresponding Author:**

Leon Rozenblit, JD, PhD  
Division of Clinical Informatics  
Beth Israel Deaconess Medical Center  
133 Brookline Avenue, HVMA Annex, Suite 2200  
Boston, MA 02215  
United States  
Phone: 1 617-278-8162  
Email: [lrozenbl@bidmc.harvard.edu](mailto:lrozenbl@bidmc.harvard.edu)

## Abstract

Maris et al document important ethical challenges at the intersection of electronic health record data and artificial intelligence development, but existing governance frameworks designed for secondary data use are categorically insufficient for artificial intelligence model training, which creates persistent deployable artifacts that encode local clinical patterns as generalizable knowledge. Drawing on two decades of stewardship framework development, we propose extending governance from data stewardship to model stewardship.

*J Med Internet Res* 2026;28:e97859; doi: [10.2196/97859](https://doi.org/10.2196/97859)

**Keywords:** data stewardship; model stewardship; AI governance; electronic health records; decontextualization; secondary use; clinical tropism; health data ethics

## Introduction

Maris et al [1] make an empirical contribution to the ethics of health data use for artificial intelligence (AI), grounding four cross-cutting themes (privacy, public trust, fair representation, and responsible integration) in stakeholder perspectives from the LEAPfROG project. Their identification of “decontextualization” as a central challenge deserves particular attention. We write from the vantage of two decades of work on stewardship frameworks for health data. The American Medical Informatics Association (AMIA) national framework for secondary use [2], the National Committee on Vital and Health Statistics (NCVHS) stewardship report to the Department of Health and Human Services [3], and the elaboration of data stewardship principles [4] established core principles (accountability, chain of trust, transparency, data quality) that Maris et al’s [1] stakeholders independently rediscover. This convergence is validating but concerning: the principles hold, yet remain unoperationalized. We argue that AI model training represents a fundamentally new form

of data use requiring a shift from data stewardship to model stewardship.

## AI Model Training Is Not Your Grandfather’s Secondary Use

Nearly 20 years ago, a national expert panel defined the secondary use of health data as uses beyond direct patient care, including research, quality measurement, public health surveillance, and commercial applications [2]. AI model training falls under this broad umbrella, but it differs from every use the framework’s architects envisioned. Traditional secondary uses analyze data and produce bounded findings; AI training creates persistent, deployable artifacts: models that may be commercialized globally, influence clinical decisions at scale, and embed the assumptions of their training context into every future prediction. A research study produces conclusions bounded by its methods and sample;

a model trained on the same data produces an artifact with unbounded downstream reach and no expiration date.

The NCVHS [3] recommended abandoning “secondary use” as too imprecise for meaningful governance; advice even more apt today. AI training should be recognized as a qualitatively distinct category of secondary use, with stewardship requirements reflecting its unique characteristics: persistence, scalability, commercial deployment, and the encoding of institutional context as generalizable knowledge.

## The Challenge of Decontextualization

Maris et al [1] identify decontextualization as a cross-cutting ethical concern. We argue it is more fundamental than their analysis suggests: not one challenge among several, but the mechanism through which the others arise.

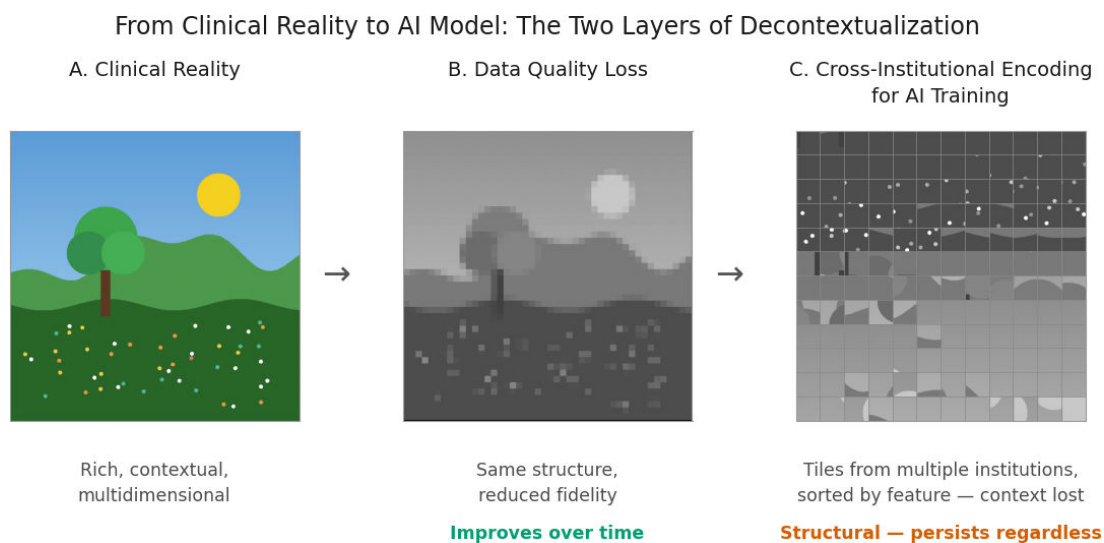
Electronic health record data encode not only clinical facts but institutional workflows, documentation practices, coding conventions, billing incentives, and resource constraints. Van der Lei’s [5] first law of medical informatics, that data should

be used only for the purpose for which they were collected, takes on new force when the reuse creates persistent, deployable artifacts rather than bounded research findings.

Two distinct challenges are at work. The first is data quality, and it is improvable: advances in ambient documentation, terminology that captures clinical intent, and better problem list governance will strengthen electronic health record data over time. The second is structural and persists regardless of data quality: every dataset carries the institutional fingerprint of its origin. What Maris et al [1], following Alami et al [6], term “clinical tropism,” the tendency of AI to reproduce narrow training environment practices, is a symptom of this structural layer. A model trained at an academic medical center with aggressive sepsis protocols learns different signals than one at a community hospital, not because the data are poor, but because they faithfully reflect different contexts (Figure 1).

Deploying models trained in one context across settings that differ systematically risks disadvantaging patients in predictable, preventable ways, the kind of harm that stewardship frameworks were designed to address.

**Figure 1.** Progressive decontextualization of clinical data. Panel A: a recognizable scene representing the rich context of clinical reality. Panel B: the same scene reduced to a grayscale grid of discrete tiles. Panel C: the tiles resorted by value, severing all spatial relationships—a visual metaphor for how electronic health record data lose institutional context when extracted for model training. AI: artificial intelligence.



## From Data Stewardship to Model Stewardship

The AMIA and NCVHS frameworks established stewardship principles for health data: accountability, chain of trust, transparency, oversight, data quality, and individual participation [3,4]. These principles must now extend to AI models and the datasets used to train them.

Consider the chain of trust, a core NCVHS concept. When data flow from hospital to aggregator to AI company to commercial model to clinical deployment across institutions, with the training data never represented, the chain does not

merely stretch; it breaks. Who is the steward of a model trained on data from five health systems and deployed in 50?

Recent work offers concrete starting points that are achievable with existing infrastructure. Multistakeholder governance frameworks [7,8] propose domain-specific approaches: clinical decision support, real-world evidence generation, and consumer health AI each require distinct governance structures. The Safe, Effective, Equitable, Trustworthy (SEET) framework provides organizing principles [8], while recommendations for AI-enabled clinical decision support specify validation, certification, safety monitoring, adverse event reporting, and provenance documentation requirements [9,10]. Real-world data

governance standards, including metadata requirements and bias documentation [9], offer complementary infrastructure.

Model stewardship, we propose, should encompass at minimum training data provenance documentation, so downstream users know what populations and practice settings a model reflects; (2) cross-institutional validation before deployment beyond the training context; (3) ongoing monitoring for context drift as clinical practices evolve; and (4) accountability structures that follow the model through its life cycle, not merely the data at its origin. All these requirements are technically feasible: provenance documentation and validation protocols exist in other regulated domains. Do we have the will to mandate them?

## Conclusion

Maris et al [1] are right that stakeholder-led governance is essential. But governance must evolve to match its target. AI models are not simply a new use of data; they are new artifacts with their own life cycle, risks, and accountability requirements. The stewardship frameworks built over two decades provide a proven foundation; extending them is the challenge. The immediate task is clear: require training data provenance and cross-institutional validation as preconditions for clinical AI deployment, just as we require evidence of efficacy before deploying therapeutics.

## Acknowledgments

Generative artificial intelligence tools (Anthropic Claude, Opus 4.6 model) were used to assist with literature organization, outline structuring, prose drafting, and generating illustrative figures during the preparation of this commentary. All intellectual content, arguments, and conclusions are the authors' own. The authors reviewed, edited, and take full responsibility for the final manuscript.

## Funding

The authors declared that no financial support was received for this work.

## Data Availability

This commentary reports no original research data.

## Authors' Contributions

LR conceived the commentary, developed the argument structure, and wrote the first draft. SL contributed to the historical stewardship framework analysis and codeveloped the thesis. CS reviewed the outline and draft, contributed to the data quality improvement perspective, and provided critical revisions. All authors reviewed and approved the final manuscript.

## Conflicts of Interest

None declared.

## References

1. Maris MT, Klopotoska JE, Cornet R, et al. The ethics of leveraging routinely collected patient data for AI development: mixed methods study. *J Med Internet Res*. Mar 2, 2026;28:e79863. [doi: [10.2196/79863](https://doi.org/10.2196/79863)] [Medline: [41814967](https://pubmed.ncbi.nlm.nih.gov/41814967/)]
2. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc*. 2007;14(1):1-9. [doi: [10.1197/jamia.M2273](https://doi.org/10.1197/jamia.M2273)] [Medline: [17077452](https://pubmed.ncbi.nlm.nih.gov/17077452/)]
3. National Committee on Vital and Health Statistics. Enhanced protections for uses of health data: a stewardship framework for "secondary uses" of electronically collected and transmitted health data. Department of Health & Human Services; 2007. URL: <https://ncvhs.hhs.gov/wp-content/uploads/2013/12/071221lt.pdf> [Accessed 2026-04-09]
4. Bloomrosen M, Detmer DE. Advancing the framework: use of health data--a report of a working conference of the American Medical Informatics Association. *J Am Med Inform Assoc*. 2008;15(6):715-722. [doi: [10.1197/jamia.M2905](https://doi.org/10.1197/jamia.M2905)] [Medline: [18755988](https://pubmed.ncbi.nlm.nih.gov/18755988/)]
5. van der Lei J. Use and abuse of computer-stored medical records. *Methods Inf Med*. Apr 1991;30(2):79-80. [doi: [10.1055/s-0038-1634831](https://doi.org/10.1055/s-0038-1634831)] [Medline: [1857252](https://pubmed.ncbi.nlm.nih.gov/1857252/)]
6. Alami H, Lehoux P, Auclair Y, et al. Artificial intelligence and health technology assessment: anticipating a new level of complexity. *J Med Internet Res*. Jul 7, 2020;22(7):e17707. [doi: [10.2196/17707](https://doi.org/10.2196/17707)] [Medline: [32406850](https://pubmed.ncbi.nlm.nih.gov/32406850/)]
7. Rozenblit L, Price A, Solomonides A, et al. Towards a multi-stakeholder process for developing responsible AI governance in consumer health. *Int J Med Inform*. Mar 2025;195:105713. [doi: [10.1016/j.ijmedinf.2024.105713](https://doi.org/10.1016/j.ijmedinf.2024.105713)] [Medline: [39642592](https://pubmed.ncbi.nlm.nih.gov/39642592/)]
8. Rozenblit L, Price A, Solomonides A, et al. Toward responsible AI governance: balancing multi-stakeholder perspectives on AI in healthcare. *Int J Med Inform*. Nov 2025;203:106015. [doi: [10.1016/j.ijmedinf.2025.106015](https://doi.org/10.1016/j.ijmedinf.2025.106015)] [Medline: [40680319](https://pubmed.ncbi.nlm.nih.gov/40680319/)]

9. Koski E, Das A, Hsueh PYS, et al. Towards responsible artificial intelligence in healthcare-getting real about real-world data and evidence. *J Am Med Inform Assoc*. Nov 1, 2025;32(11):1746-1755. [doi: [10.1093/jamia/ocaf133](https://doi.org/10.1093/jamia/ocaf133)] [Medline: [40999782](https://pubmed.ncbi.nlm.nih.gov/40999782/)]
10. Labkoff S, Oladimeji B, Kannry J, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. *J Am Med Inform Assoc*. Nov 1, 2024;31(11):2730-2739. [doi: [10.1093/jamia/ocae209](https://doi.org/10.1093/jamia/ocae209)] [Medline: [39325508](https://pubmed.ncbi.nlm.nih.gov/39325508/)]

## Abbreviations

**AI:** artificial intelligence

**AMIA:** American Medical Informatics Association

**NCVHS:** National Committee on Vital and Health Statistics

**SEET:** Safe, Effective, Equitable, Trustworthy

*Edited by Tiffany Leung; This is a non-peer-reviewed article; submitted 10.Apr.2026; final revised version received 01.May.2026; accepted 03.May.2026; published 05.Jun.2026*

*Please cite as:*

*Rozenblit L, Labkoff S, Safran C*

*From Data Stewardship to Model Stewardship: Extending Governance Frameworks for AI Era Health Data Use*

*J Med Internet Res 2026;28:e97859*

URL: <https://www.jmir.org/2026/1/e97859>

doi: [10.2196/97859](https://doi.org/10.2196/97859)

© Leon Rozenblit, Steven Labkoff, Charles Safran. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 05.Jun.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.