



Dimensions of health data Integrity

Jochen K. Lennerz¹ · Nick Schneider² · Karl Lauterbach^{2,3}

Received: 27 January 2024 / Accepted: 31 January 2024 / Published online: 15 February 2024
© Springer Nature B.V. 2024

Abstract

Health data integrity, as an emergent concept, stands to reshape the lifecycle of data-driven healthcare and research, ensuring a shared commitment to ethical practices and improved patient care.

Health data integrity has emerged as a pivotal concept in the realm of health data science and research. In a recent policy publication in *Issues in Science and Technology* [1], the authors explored some the challenges of health data utilization and propose *health data integrity* as one guiding principle toward moving the field forward. In our research, we have uncovered more than a dozen forms of integrity across various technical, professional, and other contexts. Importantly, due to the policy-oriented nature of our prior work, we intentionally omitted detailing the forms of integrity or listing the specific concepts [1]. Admittedly, these concepts can serve as instruments to facilitate an interdisciplinary dialogue, a communication tool with concrete examples from different disciplines, and as a catalyst for a more nuanced understanding. Here, we aim to share these resources as a collection of concepts from various disciplines (Table 1) to enrich the discourse on this emergent concept.

Health data integrity, in its essence, encapsulates the overall accuracy, completeness, and consistency of health-related data throughout the health data utilization life cycle. Specifically, health data integrity reflects the collective responsibility for the human rights and dignity of the individual at any stage of the life cycle from data collection until the concrete benefits of its use for the individual citizen or

the health system at large (Fig. 1a). It goes beyond the technical and legal realms, becoming a cultural imperative that demands adherence to ethical values, ensuring patient confidentiality, and respecting individual autonomy. Our understanding extends from the ALCOA principles (Table 1) to encompass the broader attributes of dependability, transparency, loyalty, honesty, good judgment, and respect [2, 3].

The multidimensional nature of health data integrity is pivotal to its applicability. From the integrity of healthcare providers to political integrity, and from safety-related data integrity to end-to-end data integrity, each dimension contributes to the overarching goal of fostering fairness, reliability in healthcare practices and ultimate trust for health data use for the individual or common good [1]. Integrity in regulatory sciences and authenticity as a special case further amplifies the complexity of this emerging guiding principle.

We therefore share, with the field, Table 1 that lists various integrity concepts established in diverse fields, providing a comprehensive resource for interdisciplinary collaboration. In contrast to the initial work [1], the table now provides concrete definitions of established concepts in various disciplines. These concepts serve as building blocks aiming to instill a collaborative spirit into ongoing interdisciplinary discussions about health data integrity. By providing these concepts, we envision that *health data integrity* can evolve into a unified framework that transcends disciplinary boundaries, fostering a shared commitment to ethics, transparency, and responsibility.

As we delve deeper into the dimensions of health data integrity, it becomes apparent that its comprehensive understanding requires transdisciplinary collaboration [4–6]. The concepts exemplify the diversity of skill sets required from

✉ Jochen K. Lennerz
Joe.Lennerz@bostongene.com

¹ BostonGene, Waltham, MA, USA

² German Federal Ministry of Health, Berlin, Germany

³ Department of Health Policy and Management, Harvard T. H. Chan School of Public Health, Boston, MA, USA

Table 1 Overview of relevant integrity concepts

Context	Description
<i>Technical</i>	
Integrity of data	• The overall accuracy, completeness, and consistency of data • Integrity means that data can be trusted • Complete, consistent, and accurate data should be Atttributable, Legible, Contemporaneously recorded, Ori- ginal or a true copy, and Accurate (so-called ALCOA principle)
Data integrity related to safety	• Safety of data regarding regulatory compliance and protecting data from unauthorized use by the data control- lers or processors (e.g., GDPR compliance); • Not to be confused with data security: protecting data from unauthorized access and use.
End-to-end data integrity	• A data science concept that ensures that the data read is the same as the data recorded • Also includes that the system can use, in case of an error, another true copy of the data
Design integrity	• Design integrity is maintained by a collection of processes, rules, and standards • Implemented before and during the design phase
Physical integrity	• Essential for storing and retrieval of data (e.g., integrity of a repository) - e.g., if a hard drive or memory is damaged, its stored data is unavoidably affected
Logical integrity	• Correctness and rationality of data in a particular context • Logical integrity ensures that the data “makes sense”
Referential integrity	• A property in data science that states that all data references are valid, e.g., referential integrity guarantees that a target referred to will be found
User integrity	• A set of rules specified by a user which do not belong to the data itself
Data-integrity system	• A type of data quality management system that ensures stability, performance, re-usability, and maintenance
<i>Professional</i>	
Integrity of action	• “Practice of being honest and showing a consistent and uncompromising adherence to strong moral and ethical principles and values” • Relies on 5 attributes: dependability, loyalty, honesty, good judgement, and respect
Integrity of a healthcare provider	• Refers to the quality of being honest, reliable, and trustworthy in the care they provide to their patients • High-quality care is based on the best available evidence and the individual needs of each patient. • Maintenance of patient confidentiality and respect their patients’ autonomy • Committed to ongoing learning and professional development, staying up to date on the latest research and best practices in their field (competence)
Integrity with respect to diversity, equity, inclusion, and minorities	• All individuals are treated fairly and with dignity • Recognizing and challenging discrimination, bias, and prejudice • Acknowledging and rectifying past injustices and working towards creating a more equitable future • Ability to address systemic barriers
Political integrity	• Political integrity is the requirement that the law be made in a way which is morally coherent.
Professional integrity (Science)	• Refers to the adherence to ethical principles and standards while conducting research and communicating scientific findings (e.g., Declarations of Helsinki and Taipei) • Encompasses accountability in the scientific process through honesty, objectivity, transparency, impartial evaluation of the quality and validity of research papers, ensuring that the review process is fair and unbiased, maintaining confidentiality, providing constructive feedback, disclosing conflicts of interest, and upholding the principles of the scientific method
<i>Legal & regulatory</i>	
Integrity of Law	• Moral coherence and the coherence of the fundamental moral principles that justify the relevant moral, politi- cal, and legal past decisions in each domain.
Integrity in regulatory sciences	• FDA Data Integrity Guidance, ISO 13,485, EU Guidance on good manufacturing practice and good distribu- tion practice, UK (2015), Switzerland (2016), Australia (2017) • Most countries have adopted the ALCOA principles • ALCOA+ (spoken ALCOA plus) also includes Completeness, Consistency, Endurance, and Availability
<i>Other</i>	
Integrity in security	• Data or information in a system is maintained so that it is not modified or deleted by unauthorized parties • Data integrity in security is an important element of data hygiene, reliability, and accuracy.
Integrity as authenticity	• Authenticity is a special case of integrity, where the “reference data version” is defined as “whatever it was when it was under control of a specific entity”
Entity integrity	• A primary key of an entity, or any part of it, can never take a null value • A requirement of E.F. Codd in his seminal paper “A relational model of data for large share data banks” (1970)

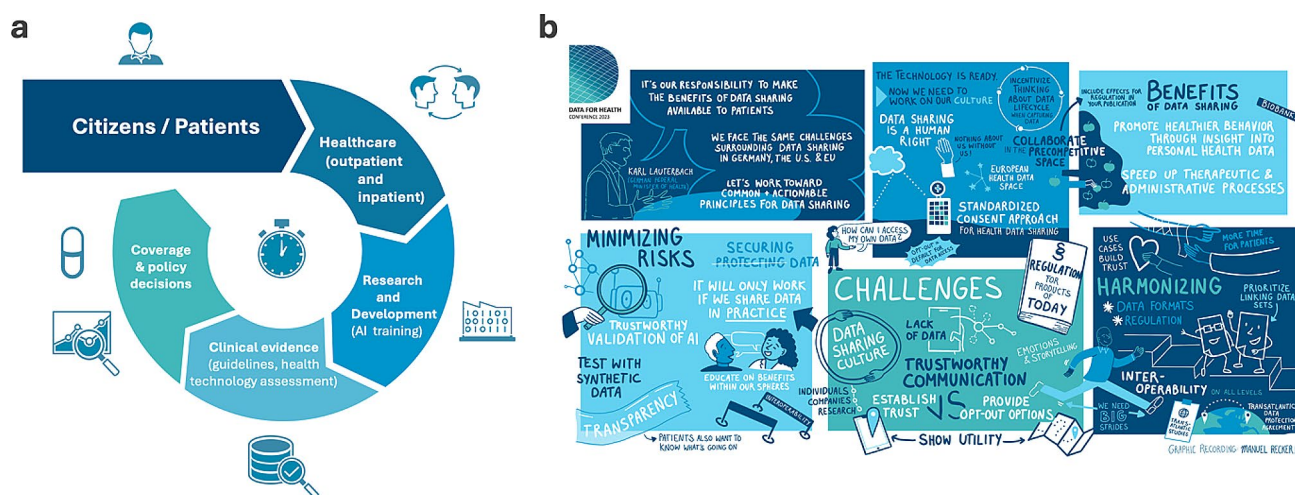


Fig. 1 Health data integrity. **a** Health data integrity and utilization life cycle: **b** example of a graphic recording from the *Data for Health* Initiative (see also: <https://projekttraeger.dlr.de/media/events/dfh23/index.html>)

scientists across various disciplines, both within and outside of medicine. Thereby *health data integrity* evolves into a multidisciplinary approach for unlocking the full potential and ensuring secure practices in the utilization of health data [4, 6].

For instance, the ‘Data for Health’ initiative, a collaboration between the German Federal Ministry of Health and the Harvard T.H. Chan School of Public Health, showcases successful transdisciplinary collaborations in the pre-competitive space [7]. The Data for Health Conference in Berlin and the subsequent Workshop in Boston exemplify how such approaches yield tangible outcomes, visually summarized through graphic recordings [7]. These transparent and accessible findings align with the initiative’s goal of promoting the secure and responsible use of health data (Fig. 1b).

Simultaneously, we assert that health data integrity, while serving as *one* guiding principle, is unlikely to comprehensively address all challenges in the health data sciences [1, 4–6]. Thus, we emphasize the need for multiple overarching principles. We advocate and foresee the emergence of multiple such principles that collectively aim to aid us in capturing the full value of health data. We hope that by sharing the concepts exemplifying health data integrity, our table serves as a starting point to facilitate collaborative power in

ongoing discussions to increase individual and public trust to optimize ethical and value-based health data utilization.

References

1. Lennerz J, Schneider N, Lauterbach K. Issues Sci Technol. 2024;1(1). <https://doi.org/10.58875/GTRD9795>. Health Data Integrity Can Earn Trust to Advance Health.
2. Council GE. Big Data und Gesundheit – Datensozialveränität als informationelle Freiheitsgestaltung 2017.
3. Council GE. Mensch und Maschine - Herausforderungen durch Künstliche Intelligenz. 2023.
4. Halamka JD, Tripathi M. The HITECH era in Retrospect. N Engl J Med. 2017;377(10):907–9. <https://doi.org/10.1056/NEJMp1709851>.
5. Sing RK, Agrawal S, Sahu A, Kazancoglu Y. Strategic issues of big data analytics applications for managing health-care sector: a systematic literature review and future research agenda. TQM J. 2023.
6. Tripathy M, Yeager M. TECCA Live! The future of Network Interoperability is Here. Health Affairs Forefr. 2023. <https://doi.org/10.1377/forefront.20231211.267976>.
7. DLR. Data for Health. 2023. <https://projekttraeger.dlr.de/media/events/dfh23/index.html>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.