Opinion Paper

Andrea Padoan*, Janne Cadamuro, Glynis Frans, Federico Cabitza, Alexander Tolios, Sander De Bruyne, William van Doorn, Johannes Elias, Zeljko Debeljak, Salomon Martin Perez, Habib Özdemir and Anna Carobene

Data flow in clinical laboratories: could metadata and peridata bridge the gap to new AI-based applications?

An investigation on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI)

Abstract: In the last decades, clinical laboratories have significantly advanced their technological capabilities, through the use of interconnected systems and advanced software. Laboratory Information Systems (LIS), introduced in the 1970s, have transformed into sophisticated information technology (IT) components that integrate with various digital tools, enhancing data retrieval and exchange. However, the current capabilities of LIS are not sufficient to rapidly save the extensive data, generated during the total testing process (TTP), beyond just test results. This opinion paper discusses qualitative types of TTP data, proposing how to divide laboratory-generated information into two categories, namely metadata and peridata. Being both metadata and peridata information derived from the testing process, it is proposed that the first is useful to describe the characteristics of data, while the second is for interpretation of test results. Together with standardizing preanalytical coding, the subdivision of laboratory-generated information into metadata or peridata might enhance ML studies, also by facilitating the adherence of laboratory-derived data to the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles. Finally, integrating metadata and peridata into LIS can improve data usability, support clinical utility, and advance AI model development in healthcare, emphasizing the need for standardized data management practices.

Keywords: metadata; peridata; artificial intelligence; clinical laboratory; total testing process; laboratory medicine

*Corresponding author: Andrea Padoan, Department of Medicine (DIMED), University of Padova and Laboratory Medicine Unity, University Hospital of Padova, Padova, Italy, E-mail: andrea.padoan@unipd.it. https://orcid.org/0000-0003-1284-7885
Janne Cadamuro, Department of Laboratory Medicine, Paracelsus Medical University Salzburg, Salzburg, Austria. https://orcid.org/0000-0002-6200-9831
Glynis Frans, Department of Laboratory Medicine, UZ Leuven, Leuven, Belgium; and Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium. https://orcid.org/0000-0001-8528-5719
Federico Cabitza, DISCo, Università degli Studi di Milano-Bicocca, Milano, Italy; and IRCCS Istituto Ortopedico Galeazzi, Milan, Italy
Alexander Tolios, Department of Transfusion Medicine and Cell Therapy, Medical University of Vienna, Vienna, Austria
Sander De Bruyne, Department of Diagnostic Sciences, Ghent University, Ghent, Belgium; and Department of Laboratory Medicine, AZ Sint-Blasius, Dendermonde, Belgium

William van Doorn, Central Diagnostic Laboratory, Department of Clinical Chemistry, Maastricht University Medical Center+, Maastricht, The Netherlands
Johannes Elias, MDI Limbach Berlin GmbH, Berlin, Germany; and HMU Health and Medical University GmbH, Potsdam, Germany
Zeljko Debeljak, Faculty of Medicine, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia; and Clinical Institute of Laboratory Diagnostics, University Hospital Center Osijek, Osijek, Croatia
Salomon Martin Perez, Unidad de Bioquímica Clínica, Hospital Universitario Virgen Macarena, Sevilla, Spain. https://orcid.org/0000-0002-2086-3597
Habib Özdemir, Türkiye Health Data Research and Artificial Intelligence Applications Institute, Health Institutes of Türkiye (TUSEB), İstanbul, Türkiye. https://orcid.org/0000-0002-1267-7233
Anna Carobene, IRCCS San Raffaele Scientific Institute, Milan, Italy

Introduction

In recent years, clinical laboratories have experienced significant improvements in several technological tools and instrumentation. Laboratory information systems (LIS), initially introduced in the 1970s, have rapidly evolved from simple software to a sophisticated part of the entire information technology (IT) system, available in clinical laboratories, able to retrieve and exchange information with several instrumental middleware, other laboratories LIS as well as hospital information systems and regional databases [1]. In general, the increased capabilities of certain LIS, coupled with the advancements in various information technologies, the introduction of cost-effective chemical, physical and mechanical sensors into analytical instruments, and the improved integration with other digital tools, have led to an increase in the volume of data produced in clinical laboratories [2].

Considering the total testing process (TTP) or the so-called "brain-to-brain loop" [3], the analytical phase includes only a part of the extensive information stream generated during the entire cycle of testing. Apart from patients' test results and demographic details, LIS might also register many additional pieces of information, such as the plain names of tests, audit trail records, technical or medical validations, as well as ward names for inpatients or the general practitioner for outpatients. Data generated during the TTP may also include other details or information of importance [4]. As an example, considering the pre-preanalytical and the preanalytical phases, a large volume of data may contain the time and types of samples collected, the handling conditions, including the temperature and time of transportation, the centrifugation conditions, the number of aliquots generated, etc. [5, 6]. During the analytical phases, relevant information is often recorded by LIS in addition to test results. These might include hemolysis (H), icterus (I) and lipemia (L) indexes (which might be a unique value, or a value associated for each executed test), the assay calibration curves, sample dilutions, analysis repetitions and technical validation rules (accepted or violated). For analyses based on "-omics", other data, often of large volume, are generated, encompassing mass spectra or proteomics, metabolomics and lipidomics, and sequence files for next-generation sequencing. Certain LIS might incorporate or integrate data from the laboratory quality control (QC) system; these include not only internal quality controls (IQC), but also external quality assessments (EQA) as well as additional resources about the entire process of verification and validation of analytical methods [7]. Finally, there are some LIS, more specifically designed for genetic testing or for microbiology or transfusion medicine, which are capable of integrating with different instrumental software or sophisticated pipelines to convert thousands of sequences into clinically useful information [8].

During the post-analytical phase, interpretative comments, the sample storage or sample re-analyses (e.g. in case of follow-up testing), additional tests requested by clinical wards and details on the scrutiny of urgent results with the issue of provisional report and its delivery to the requesting clinician are usually generated and are also registered by LIS.

The aim of this opinion paper from the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI) is to discuss qualitative types of data produced in each phase of TTP, and propose how to divide laboratory-generated information into distinct data categories.

The importance of collecting and sharing high-quality data

The short statement "garbage in, garbage out" recognizes that poor quality of data would lead to unreliable outputs, emphasizing the utmost importance of improving and maintaining data quality [9]. In healthcare research, the collection of high-quality, reliable data is essential in the field of statistics and artificial intelligence (AI)-models, especially machine learning (ML) methods, which also include deep learning techniques for generating new clinically relevant algorithms for advancing medical knowledge, illness detection and personalized treatments [10]. The development of highly reliable ML models usually requires large sets of data for training, testing, and validation; however, these large datasets can often only be obtained by combining a multitude of sources. Additional reasons motivate the merging of data. Firstly, in many circumstances, even for large clinical laboratories, it is unrealistic to retrieve large datasets as required by ML, and the under-representation of specific patient cohorts in testing data could represent a biasing issue; examples are cases where ML algorithms are applied to pediatric records, rare diseases or in situations with a high imbalance in the number of studied individuals within groups. On the other hand, outside sources (e.g. from other centers or other laboratories) have to be merged with local data to support external validation. Third, the *in praxi* continuous performance monitoring of ML-based algorithms (e.g. accuracy, sensitivity and specificity) over time might be of higher quality when using externally gathered data, different from the source (e.g. the laboratory) where they are usually applied [11, 12]. This latter consideration may be of particular interest to engineers and data scientists, since the complex interaction between AI-models and real-world conditions (including the instrument

variability and bias) can lead to unexpected and harmful behavior [13]. Thus, the possibility of collecting high-quality data from multiple, unparalleled resources is of utmost importance and should be encouraged by scientific communities, and through national and international guidelines [14].

Defining properties of data attributable to the total testing process

It is undoubtedly true that laboratory TTP-generated data could benefit from a redefinition of terms, especially considering their potential use outside clinical purposes, such as in the fields of data science for statistical applications and the generation of ML tools.

The term *metadata* was first announced by the National Information Standards Organization [15] and it was defined as "attributes that are necessary to locate, fully characterize, and ultimately reproduce other attributes that are identified as data" [16]. Thus, metadata are additional information that can be enclosed with data, which clearly and unambiguously describe the data as well as their full provenance [17]. This definition is reminiscent of the statement given by the National Institute of Standards and Technology (NIST): "Information describing the characteristics of data including, for example, structural metadata describing data structures (e.g. data format, syntax, and semantics) and descriptive metadata describing data contents" [18]. Metadata are often called "data about information" or "information about information" [19]. Both definitions are not contradictory and highlight some possible relevant features of metadata applicable to the laboratory medicine field. These valuable technical pieces of information improve the "reproducibility", "comparability" and "harmonization" of data, requirements that are relevant for assessing the quality and the validity of laboratory test results. In addition, considering the scientific fields related to data science, such as ML approaches, metadata not only represent useful information on the robustness of data but may also facilitate data sharing and improve usability. Interestingly, the concepts of comparability and reliability are commonly associated with clinical laboratory findings as well.

Beyond metadata, other forms of data (excluding laboratory test results, which are deemed the "primary data"), e.g. reference intervals, clinical decision limits, etc., may in some way affect the TTP and thus warrant special consideration. Our suggestion is to designate data, which neither constitute primary data nor metadata, as "peridata". Peridata can be characterized as data facilitating the accurate interpretation of test results. Hence, peridata may enrich mostly the clinical

validity of test results, also by offering comparative frameworks, benchmarks, or additional layers of validation. Hence, peridata are information relevant for the interpretation of the results within the clinical context, making that data actionable for the patients' care. It is essential to consider that since peridata may also be derived from databases different from those typically obtained from clinical laboratories, it has to be ensured that informed consent has been obtained from the patient or a legal representative of the patient. This consideration applies only to certain peridata. However, the requirement for obtaining consent may vary depending on the regulations of different European and non-European countries.

Based on the previous definitions of metadata and peridata in the clinical laboratory setting, it should be clearly stated which kind of information generated throughout the TTP could be defined as metadata or peridata. Therefore, through consensus among the members of the EFLM WG-AI, a distinction between metadata and peridata was identified, as reported in Tables 1–2 and Figure 1.

A practical, real-world example is provided in the Supplementary Material to demonstrate the data records retrieved from the LIS, instrument middleware, quality management system (QMS), or electronic health records (EHR). This was done for a synthetic patient result derived from the clinical laboratory of one of the authors. All available data registrations were categorized into meta- or peridata. The results are reported in Supplementary Tables 1–4. It should be noted that not all data types were retrievable directly from the LIS, and that additional data sources needed to be consulted to obtain the different meta- and peridata missing values. This means that for the registration of relevant meta- and peridata in the LIS, laboratories might need to invest time, efforts and costs for their creation, storage, and retrieval. These investments will have to be expended by the LIS user (who needs to enter data that currently remains unregistered), the LIS producer (who needs to create additional storage and interfaces for registration and retrieval), or the LIS buyer (who needs to pay a surplus for a meta- and peridata-capable LIS or additional instrument driver connections that feed data into the LIS).

Table 1: Short definitions of metadata and peridata.

Primary data	Laboratory test results
Metadata	Data derived from the testing process that describe the characteristics and the requirements that are relevant for assessing the quality and the validity of laboratory test results.
Peridata	Data derived from the testing process that are relevant for the interpretation of the results within the clinical context, making that data actionable for the patients' care.

Table 2: Detailed categorization of data management in clinical laboratory testing across pre-preanalytical, preanalytical, analytical, and post-analytical phases.

Category	Data description	Data categorization
Patient's data	Patient ID	Peridata
	In-patient/out-patient	Peridata
	Department or clinical ward	Peridata
	Demographic characteristics (e.g. age, sex, ethnicity)	Peridata
	Co-morbidity/clinical data/health status	Peridata
	Clinical indications or diagnosis	Peridata
	Test request indication/purpose in the clinical pathway (e.g. screening, diagnosis, monitoring, …)	Peridata
	Lifestyle factors	Peridata
	Status of the patient during the collection (e.g. fasting or lying down)	Peridata
Pre-pre-analytical	Test name (named or coded, e.g. using LOINC) [29]	Metadata
	Processing lab name and geographical location	Metadata
	Type of tubes	Metadata
	Sample type	Metadata
	Time of collection	Metadata
	Location of sample collection	Metadata
	Transport and storage condition	Metadata
Pre-analytical	Centrifugation (time and temperature)	Metadata
	Sample preparation (e.g. dilution or treatment)	Metadata
	Time between collection and analysis	Peridata
	Hemolysis, icterus, and lipemia (HIL) indexes values	Peridata
	Intra-lab processing, e.g. manual, pre-analytical workstation or total lab automation	Metadata
Analytical	Calibration curves	Metadata
	Internal quality controls (brand and types, e.g. commutable or not)	Metadata
	Internal quality control results	Metadata
	Reagents references and lots, and types of equipment used for the test	Metadata
	Analytical method (principle of the method, dilution factors, etc.)	Metadata
	Interferences (heterophile antibodies, human anti-animal antibodies, etc.)[a]	Metadata/Peridata
	Intermediate calculations (including measurement units' conversion)	Metadata
	Results generated by reflex testing algorithms	Peridata
	Data quality metrics (accuracy, precision, metrological traceability of calibrators)	Metadata
	Measurement units	Peridata
	Samples dilutions/was the sample reprocessed after dilution?	Metadata
	Measurement uncertainty or total error	Peridata
	Biological variation	Peridata
Post-analytical	Reference intervals or thresholds for interpreting results (including whether these are based on population studies, manufacturer guidelines, or laboratory-specific validations)	Peridata
	Previous test results for the same patient for comparative analysis (historical results stored within the laboratory information system or require external retrieval)	Peridata
	Result communication (how results are reported: electronic health records, printed reports, direct communication)	Peridata
	External quality controls (results, providers, reference systems)	Metadata/Peridata[b]
	Data analysis and interpretation (how data is analyzed, interpreted, and integrated into clinical decision-making)	Peridata
	Turnaround time (TAT)	Metadata

[a]Depends on whether interference from immunoassay heterophile or human anti-animal antibodies is only reported by the manufacturer within the insert (metadata), or whether endogenous heterophile and human anti-animal antibodies have been previously identified in the patient (peridata). [b]Depends on whether results can have an impact in the clinical setting.

The FAIR data approach

The concepts of findability, accessibility, interoperability, and reusability (FAIR) within the context of scientific data are paramount for making healthcare data (and thus laboratory medicine data) optimally usable for clinical and research purposes [20]. The FAIR principles aim to address challenges associated with data sharing and reuse in research, ensuring that data are not only generated for a specific purpose but also has the potential to contribute to
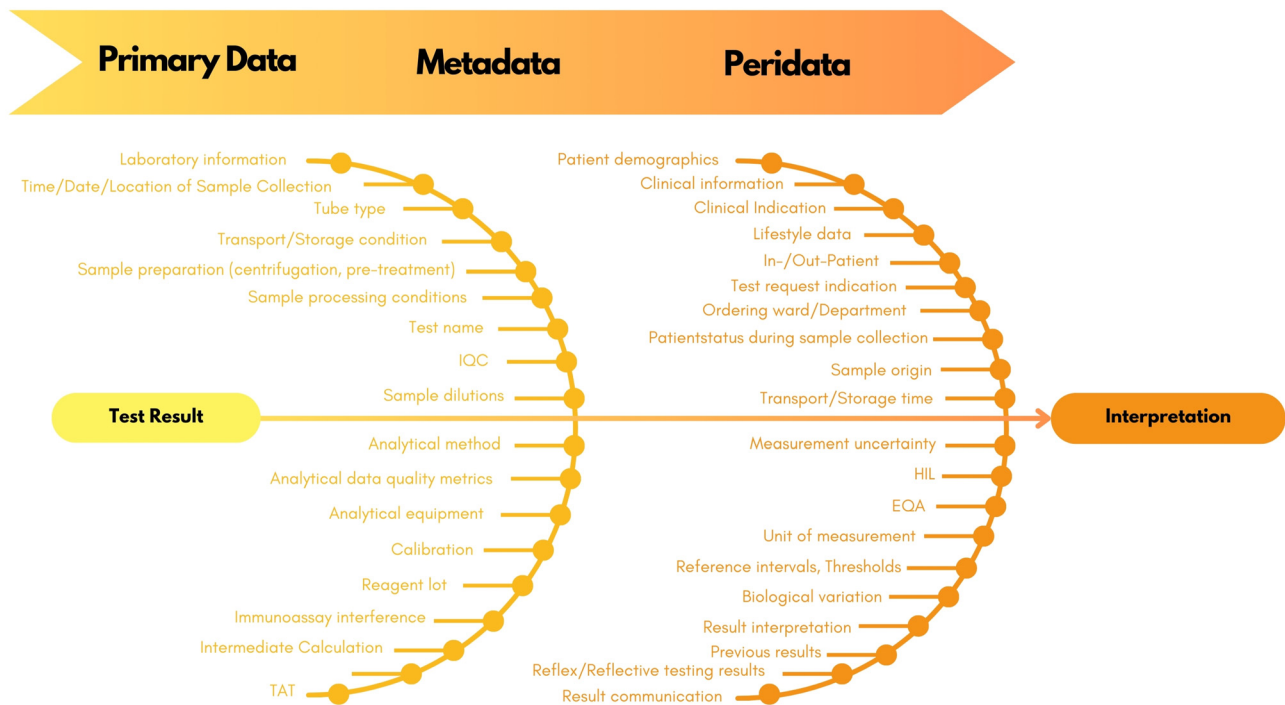
Figure 1: Types of data generated during the phases of the total testing process, subdivided by primary data (test results) and metadata/peridata. HIL, hemolysis, icterus and lipemia indexes; TAT, turnaround time; EQA, external quality assessment; IQC, internal quality control.

broader scientific advancements across disciplines [21]. It is noteworthy that the FAIR principles could be extended beyond the realm of collected data to encompass the entirety of a research study, offering even greater applicability and benefits [22].

Ensuring the findability of data establishes a foundation for effective data discovery, enabling researchers to locate and access relevant datasets efficiently. Indeed, datasets should be easily findable (e.g. by using unique identifiers for the dataset and by promoting use of online repositories); second, once data are found, they should be easily accessible. This means that the data, along with its meta- and peridata, should be promptly accessible using adequate permissions.

Finally, data should be interoperable and reusable [23]. Interoperability involves the capability for data to be seamlessly integrated and merged with other datasets. Reusability, on the other hand, implies that data are designed to be reused for future purposes [24]. In the context of medical laboratory data, ensuring interoperability and reusability entails providing clear and comprehensive meta- and peridata, documenting the data, the steps involved in data creation and processing, and adopting widely accepted standard formats to ease interpretation and promote reuse. Adopting the FAIR principles could create positive synergy with the validation of AI models. Specific benefits of FAIR AI research could be a) an

improvement of generalizability, by exposing AI models to diverse, FAIR datasets, b) the guarantee of scientific correctness of AI algorithms, especially when they are applied to different contexts or different labs [25], c) reduce the overall cost of research [26], d) the evaluation of real-world performance monitoring of AI models [27]. The latter point gains significance in assessing the potential growing volume of AI tools available both for improving instrumental technologies and in clinical practice. This attention has grown in response also to recent studies highlighting concerns about the robustness and generalizability of FDA-cleared AI models for research and clinical use [28]. The establishment of a publicly available international repository of laboratory-medicine datasets that may serve as the benchmark sets for AI/ML model performance evaluation may resolve this problem. Further solutions could be represented by synthetic datasets, as already demonstrated for hematological laboratory analyses [29].

Necessity of standard preanalytical coding for biospecimens to improve data reliability

Standardization and harmonization are key elements pursued by national and international laboratory societies and

agencies to render medical diagnostics comparable and usable for national and international databases (e.g. European Health Data Space). Systems like *Logical Observation Identifiers Names and Codes* (LOINC) [30] or *Systematized Nomenclature of Medicine Clinical Terms* (SNOMED-CT) [31] have been around for some time and have been improved over time. While LOINC focuses on variables such as the type of sample material, the analytical method, the unit of measurement, etc., SNOMED-CT includes also clinical content for medical documentation with greater granularity than LOINC. However, none of these systems take pre- or post-analytical meta- or peridata of laboratory test results into account. Terminology standards are crucial for ensuring semantic interoperability in healthcare data and for guaranteeing efficient interactions across databases. However, the existence of multiple standards limits the achievement of a consensus, and data interoperability can be enhanced by cross-mapping LOINC and SNOMED-CT or using SNOMED-CT as a reference terminology to bridge various standards [32].

Interestingly, both LOINC and SNOMED-CT prove to be highly pertinent, since they encompass valuable information that could be regarded as either meta- or peridata. The latter consideration emphasizes the significance of meta- and peridata; while certain features of LOINC and SNOMED-CT are recognized as relevant for ensuring data robustness, their applicability in the context of data analysis or for ML modeling remains relatively limited. Another important system is the Observational Health Data Sciences and Informatics (OHDSI), which is fundamental for the interrelationship among databases, assuring anonymization and aggregation of data. OHDSI offers many open-source tools to support the standardization of data analytics allowing the interaction across two or more databases using the Common Data Model (CDM) [33].

As of today, to the best of our knowledge, the only system that may be considered a valid approach to document pre-analytical variables in a standardized fashion is the *Sample PREanalytical Code* (SPREC) system for biobanking samples, developed by the *International Society for Biological and Environmental Repositories* (ISBER) Biospecimen Science Working Group [34]. A standardized methodology for constructing a coding system tailored to medical laboratories could be developed enlisting national or international laboratory associations and working groups active in these topics. However, in anticipation of the ever-increasing use of data for AI modeling, it is crucial that variables potentially influencing the test results and/or its interpretation be documented alongside the result itself, as they could represent important meta- or peridata. Therefore, immediate actions have to be taken to develop such a catalog, ideally by implementation into existing coding systems. With respect to the data structure, either SNOMED-CT or a SNOMED-CT/LOINC combination would be the most reasonable candidates.

Conclusions and remarks

This study delves into key facets regarding data produced in clinical laboratories and the distinctive attributes defining them in the context of AI applications. Significant advancements in LIS and IT have ushered in an era of unprecedented data availability, emphasizing the pivotal role of laboratories in generating patient data, promptly applicable in ML models. Nevertheless, data derived from laboratories TTP extend beyond mere test results; it encompasses various additional pieces of information, with a substantial portion involving meta- or peridata. Understanding and appreciating the significance of additional information linked to test results may be improved by referring to a part of TTP laboratory data as meta- and peridata. This information is truly valuable for many applications, including those using ML specifically. Initially, this approach could facilitate the design of new clinical studies, feature collections, and information recording. However, achieving this purpose can be challenging due to limited access to clinical data. A recent European survey organized by WG-AI of the EFLM, collecting 195 replies from European countries, revealed that about 50 % of laboratory professionals lack full access to clinical data, and another study reported that about 65 % of laboratories do not have such access [35, 36].

Secondly, metadata could improve the reusability of data, aligning with the FAIR principles and creating a positive synergy for the validation of AI models. The adoption of these could yield several advantages in the development of future AI applications in healthcare, mitigating financial waste associated with fragmented efforts. For this purpose, clinical laboratories and Scientific Societies should stimulate the sharing of original experimental data, for reproducibility and comparability studies and other uses. Thirdly, standardization processes could be enhanced by focusing on a combination of meta- and peridata, whilst the development of LIS software could be improved by enabling the prompt visualization of all peridata available, particularly in clinical or technical validation. Metadata and peridata are, therefore, crucial for ensuring data quality, results accuracy, and their proper interpretation, ultimately aiming to benefit patient care [37, 38].

While the terms meta- and peridata are usually more confined to the domain of data science, they should be viewed as essential elements within the data management ecosystem, enabling researchers and clinicians to comprehend the nuances and context of primary data. For these reasons, it is the responsibility of the domain experts to

define the kind of information useful for enriching the data with meta- or peridata to comprehensively describe the data production process (see Table 2). The redefinition of data types in the field of laboratory medicine is becoming increasingly essential, especially with the growing utilization of advanced computational methods, such as ML algorithms, to facilitate the creation of clinically valuable datasets.

References

1. Kammergruber R, Durner J. Laboratory information system and necessary improvements in function and programming. LaboratoriumsMedizin 2018;42:277–87.
2. Padoan A, Plebani M. Flowing through laboratory clinical data: the role of artificial intelligence and big data. Clin Chem Lab Med 2022;60: 1875–80.
3. Plebani M, Laposata M, Lundberg GD. The brain-to-brain loop concept for laboratory testing 40 years after its introduction. Am J Clin Pathol 2011;136:829–33.
4. Bellini C, Padoan A, Carobene A, Guerranti R. Moving towards total health data integration including quality management: insights from the SIBioC Working Group "Big Data and Artificial Intelligence" survey. Biochim Clin 2024;48:46–52.
5. Plebani M. Exploring the iceberg of errors in laboratory medicine. Clin Chim Acta 2009;404:16–23.
6. Cadamuro J, Simundic AM. The preanalytical phase – from an instrument-centred to a patient-centred laboratory medicine. Clin Chem Lab Med 2023;61:732–40.
7. Sepulveda JL, Young DS. The ideal laboratory information system. Arch Pathol Lab Med 2013;137:1129–40.
8. Aronson S, Mahanta L, Ros LL, Clark E, Babb L, Oates M, et al. Information technology support for clinical genetic testing within an academic medical center. J Personalized Med 2016;6:1–9.
9. Kilkenny MF, Robinson KM. Data quality: "garbage in – garbage out". Health Inf Manag J 2018;47:103–5.
10. Javaid M, Haleem A, Pratap Singh R, Suman R, Rab S. Significance of machine learning in healthcare: features, pillars and applications. Int J Intell Network 2022;3:58–73.
11. Carobene A, Milella F, Famiglini L, Cabitza F. How is test laboratory data used and characterised by machine learning models? A systematic review of diagnostic and prognostic models developed for COVID-19 patients using only laboratory data. Clin Chem Lab Med 2022;60:1887–901.
12. Agnello L, Vidali M, Padoan A, Lucis R, Mancini A, Guerranti R, et al. Machine learning algorithms in sepsis. Clin Chim Acta 2024;553:117738.
13. Azimi V, Zaydman MA. Optimizing equity: working towards fair machine learning algorithms in laboratory medicine. J Appl Lab Med 32023;8:113–28.
14. Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, et al. The importance of being external methodological insights for the external validation of machine learning models in medicine. Comput Methods Progr Biomed 2021; 208:106288.
15. Riley J. Understanding metadata what is metadata, and what is it for? A primer publication of the National Information Standards Organization. National Information Standards Organization (NISO); 2017. Available from: https://groups.niso.org/higherlogic/ws/public/download/17446/Understanding%20Metadata.pdf.
16. ISO/IEC 2382. Information technology — vocabulary; 2015. Available from: https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v2:en.
17. Ghiringhelli LM, Baldauf C, Bereau T, Brockhauser S, Carbogno C, Chamanara J, et al. Shared metadata for data-centric materials science. Sci Data 2023;10:1–18.
18. Johnson CS, Badger ML, Waltermire DA, Snyder J, Skorupka C. Guide to cyber threat information sharing. National Institute of Standards and Technology (NIST); 2016. Available from: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-150.pdf.
19. Grassi PA, Lefkovitz NB, Nadeau EM, Galluzzo RJ, Dinh AT. Attribute metadata: a proposed schema for evaluating federated attributes. National Institute of Standards and Technology (NIST); 2018.
20. Blatter TU, Witte H, Nakas CT, Leichtle AB. Big data in laboratory medicine—FAIR quality for AI? Diagnostics 2022;12:1–13.
21. GO FAIR. FAIR principles. https://www.go-fair.org/fair-principles/ [Accessed 23 July 2024].
22. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.
23. Overmars LM, Niemantsverdriet MSA, Groenhof TKJ, De Groot MCH, Hulsbergen-Veelken CAR, Van Solinge WW, et al. A wolf in sheep's clothing: reuse of routinely obtained laboratory data in research. J Med Internet Res 2022;24:e40516.

24. Ravi N, Chaturvedi P, Huerta EA, Liu Z, Chard R, Scourtas A, et al. FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy. Sci Data 2022;9:1–9.

25. Huerta EA, Blaiszik B, Brinson LC, Bouchard KE, Diaz D, Doglioni C, et al. FAIR for AI: an interdisciplinary and international community building perspective. Sci Data 2023;10:1–10.

26. European Commission. Cost-benefit analysis for FAIR research data: cost of not having FAIR research data; 2019. [Online]. Available from: https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en.

27. Allen B, Dreyer K, Stibolt R, Agarwal S, Coombs L, Treml C, et al. Evaluation and real-world performance monitoring of artificial intelligence models in clinical practice: try it, buy it, check it. J Am Coll Radiol 2021;18:1489–96.

28. Dreyer KJ, Allen B, Wald C. Real-world surveillance of FDA-cleared artificial intelligence models: rationale and logistics. J Am Coll Radiol 2022;19:274–7.

29. D'Amico S, Dall'Olio D, Sala C, Dall'Olio L, Sauta E, Zampini M, et al. Synthetic data generation by artificial intelligence to accelerate research and precision medicine in hematology. JCO Clin Cancer Inf 2023:e2300021. https://doi.org/10.1200/CCI.23.00021.

30. The international standard for identifying health measurements, observations, and documents. https://loinc.org/ [Accessed 23 July 2024].

31. Use SNOMED CT. SNOMED International. https://www.snomed.org/use-snomed-ct [Accessed 23 July 2024].

32. Park HA. Why terminology standards matter for data-driven artificial intelligence in healthcare. Ann Lab Med 2024;44:467–71.

33. The Observational Health Data Sciences and Informatics (OHDSI). https://www.ohdsi.org [Accessed 17 September 2024].

34. Lehmann S, Guadagni F, Moore H, Ashton G, Barnes M, Benson E, et al. Standard preanalytical coding for biospecimens: review and implementation of the Sample PREanalytical Code (SPREC). Biopreserv Biobanking 2012;10:366–74.

35. Cadamuro J, Carobene A, Cabitza F, Debeljak Z, De Bruyne S, van Doorn W, et al. A comprehensive survey of artificial intelligence adoption in European Laboratory Medicine: current utilization and prospects. Clin Chem Lab Med 2024;63:692–703.

36. Bellini C, Padoan A, Carobene A, Guerranti R. A survey on Artificial Intelligence and Big Data utilisation in Italian clinical laboratories. Clin Chem Lab Med 2022;60:2017–26.

37. Badrick T, Banfi G, Bietenbeck A, Cervinski MA, Loh TP, Sikaris K. Machine learning for clinical chemists. Clin Chem 2019;65:1350–6.

38. Ferrari A, Pennestrì F, Bonciani M, Banfi G, Vainieri M, Tomaiuolo R. The role of patient-reported experiences in disclosing genetic prenatal testing: findings from a large-scale survey on pregnant women. Eur J Obstet Gynecol Reprod Biol X 2024;23:100327.