IAES Journal of Electrical and Electronics Engineering
Vol. 1, No. 2, December 2014, pp. 44~48
ISSN: 2338-7432, DOI: 10.11591/jeee.v1i2.pp44-48

Data quality, interoperability, and information integration in public health informatics: a critical review

Joshua R. Vest, Thomas R. Campion Jr.
Weill Cornell Medical College, United States

Article Info

Article history:

Received Feb 14, 2014
Revised Jul 1, 2014
Accepted Aug 15, 2014

Keywords:

Data quality
Health information exchange
Health information systems
Information integration
Interoperability
Public health informatics

ABSTRACT

The increasing use of digital technologies in healthcare has generated substantial amounts of health information from hospitals, laboratories, public health agencies, surveillance systems, and population-based databases. However, the availability of large volumes of information does not necessarily guarantee its reliability, accessibility, or usefulness for public health decision-making. Data quality, interoperability, and information integration are therefore fundamental components of effective public health informatics. This review critically examines the concepts, challenges, and technological approaches associated with these three components, emphasizing literature published up to 2013. Particular attention is given to data quality dimensions, health information exchange (HIE), standardized data structures, semantic interoperability, and integration of public health information systems. The review identifies technical, organizational, and governance-related barriers that limit effective information integration. It argues that sustainable public health informatics requires an integrated approach combining standardized data management, continuous quality assessment, interoperable technologies, and institutional collaboration.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Information is a fundamental resource for effective public health management because decisions concerning disease prevention, health promotion, surveillance, and healthcare planning depend on reliable and timely data. Public health organizations obtain information from various sources, including hospitals, laboratories, primary healthcare facilities, demographic databases, immunization registries, and disease surveillance systems. However, differences in data formats, collection procedures, reporting standards, and technological infrastructure frequently create fragmented information environments.

Public health informatics addresses these challenges through the systematic application of information science and technology to public health practice. Its effectiveness depends not only on the availability of digital systems but also on the quality and compatibility of the information they produce. Vest *et al.* [1] emphasized that the quality and integration of public health information systems are closely related to the ability of organizations to exchange and utilize information from multiple independent sources.

Three interconnected issues are particularly important in this context: data quality, interoperability, and information integration. Data quality determines whether information is sufficiently reliable for its intended purpose, while interoperability enables different information systems to exchange and interpret data consistently. Information integration brings these capabilities together to support comprehensive analysis, coordinated public health activities, and evidence-based decision-making.

This review critically examines the conceptual foundations, technological approaches, practical applications, and persistent challenges associated with data quality, interoperability, and information integration in public health informatics. The discussion focuses exclusively on scientific literature published no later than 2013.

2. DATA QUALITY IN PUBLIC HEALTH INFORMATICS

Data quality is a fundamental requirement for effective public health information systems because inaccurate or incomplete information can produce misleading epidemiological findings and inappropriate policy decisions. Public health data are frequently collected through decentralized systems involving different healthcare institutions, administrative organizations, and reporting mechanisms. Consequently, variations in data collection, recording, validation, and management can substantially influence the reliability of information.

2.1. Dimensions of data quality

Data quality is a multidimensional concept that extends beyond simple accuracy. Weiskopf and Weng [2] identified five important dimensions of electronic health record data quality: completeness, correctness, concordance, plausibility, and currency. These dimensions provide a useful framework for assessing whether information is sufficiently reliable and suitable for secondary use in research and public health applications.

Completeness refers to whether required information is available, whereas correctness concerns whether recorded values accurately represent the underlying reality. Concordance describes agreement between different data elements or information sources, while plausibility evaluates whether recorded values are reasonable within their clinical or epidemiological context. Currency concerns whether information is sufficiently up to date for its intended application, particularly in time-sensitive public health surveillance.

These dimensions are interconnected and should not be evaluated independently. For example, a disease surveillance database may contain accurate information but still be unsuitable for immediate intervention if reporting is delayed. Similarly, complete records may produce unreliable conclusions when information is inconsistent across healthcare institutions.

2.2. Data quality assessment and management

Effective data quality management requires systematic assessment throughout the information lifecycle, beginning with data collection and continuing through processing, storage, analysis, and reporting. Weiskopf *et al.* [3] demonstrated that data completeness can be defined and measured in different ways depending on the intended use of electronic health records. This finding highlights the importance of establishing explicit assessment criteria before using health information for research or administrative purposes.

Common assessment approaches include comparison with reference standards, examination of missing values, consistency checks, distribution analysis, and verification against independent information sources. Automated validation procedures can identify unusual values, incomplete records, duplicate entries, and inconsistencies during data entry. Nevertheless, automated procedures cannot independently establish whether information accurately represents real-world health conditions, making human review and institutional data governance necessary.

A critical limitation of conventional data quality assessment is its tendency to emphasize technical attributes while overlooking the circumstances under which information is generated. Data quality is also influenced by healthcare workers' recording practices, institutional reporting policies, patient access to services, and the availability of technological resources. Therefore, data quality should be understood as both a technical and organizational responsibility.

3. INTEROPERABILITY IN PUBLIC HEALTH INFORMATION SYSTEMS

Interoperability refers to the ability of different information systems to exchange, process, and meaningfully use information without requiring extensive manual transformation. In public health, interoperability is particularly important because epidemiological surveillance and population health management require information from multiple independent institutions. Without effective interoperability, potentially valuable information remains fragmented within individual organizations.

3.1. Technical and syntactic interoperability

Technical interoperability concerns the ability of systems to establish communication and transfer information through compatible technological infrastructures. It involves communication protocols, network connectivity, data exchange mechanisms, and compatible system interfaces. These components establish the basic conditions required for information exchange between healthcare institutions and public health organizations.

Syntactic interoperability focuses on the structure and format of exchanged information. Standardized message structures and data formats allow different applications to interpret the organization of

incoming information consistently. Health level seven (HL7) standards represent an important example of efforts to facilitate structured information exchange across healthcare information systems.

However, technical connectivity and standardized message formats do not necessarily guarantee that exchanged information will be interpreted consistently. Different systems may use different definitions, classifications, abbreviations, or coding practices for apparently identical health concepts. Consequently, additional approaches are needed to ensure that information retains its intended meaning across organizational and technological boundaries.

3.2. Semantic interoperability

Semantic interoperability refers to the ability of different information systems to interpret and use exchanged information according to consistent meanings. This capability is particularly important in public health because disease classifications, diagnostic terminology, laboratory observations, and population health indicators must be interpreted consistently. Standardized terminologies and classification systems can reduce ambiguity and improve the comparability of information.

Liaw *et al.* [4] emphasized the importance of semantic interoperability and ontology-based approaches in managing data quality across integrated chronic disease management systems. Ontologies can provide structured representations of concepts and relationships, allowing information systems to interpret health-related terminology more consistently. Such approaches can also facilitate information integration, knowledge representation, and the reuse of routinely collected health data.

Nevertheless, semantic interoperability remains challenging because health concepts and clinical terminology may vary between institutions and geographical settings. Differences in local clinical practices, coding systems, language, and organizational requirements can complicate the adoption of universal standards. Effective semantic interoperability therefore requires not only standardized terminology but also agreement among the institutions responsible for generating and using health information.

4. Information integration in public health

Information integration involves combining data from different sources into a coherent information environment that supports analysis, reporting, and decision-making. Public health systems frequently require integrated information from clinical records, laboratory reports, immunization registries, demographic databases, and disease surveillance platforms. Integrating these sources can provide a more comprehensive understanding of population health conditions than examining individual databases separately.

4.1. Health information exchange

Health information exchange (HIE) provides an organizational and technological mechanism for sharing health information across institutions. It can support communication among hospitals, laboratories, primary healthcare facilities, and public health agencies. Vest and Gamm [5] identified persistent challenges in HIE, emphasizing that barriers extend beyond technical limitations to include organizational and institutional factors.

Effective HIE requires compatible information systems, agreed data-sharing procedures, appropriate governance arrangements, and sufficient organizational incentives. Institutions must also address concerns regarding confidentiality, ownership, access rights, and the responsibilities associated with shared information. Therefore, successful information exchange depends on coordinated institutional arrangements rather than technology alone [6].

4.2. Integration of public health information systems

Public health information systems often operate independently, with separate databases and administrative structures for specific programs. Immunization information systems and vital records systems provide examples of platforms that require information from multiple data providers. Vest *et al.* [7] demonstrated that fragmentation, provider autonomy, technological heterogeneity, and data quality issues can limit the effective integration of these systems.

Integrating information systems can improve the availability of population-level data and reduce duplication in reporting processes. For example, linking immunization records with demographic and vital statistics information can support more comprehensive assessment of vaccination coverage and population health indicators. However, successful integration requires consistent identifiers, standardized data definitions, reliable record linkage procedures, and appropriate mechanisms for resolving conflicting information [8].

4.3. Secondary use of health information

The secondary use of health information involves reusing data originally collected for clinical, administrative, or operational purposes in research, surveillance, and public health planning. Electronic health records provide valuable opportunities for this purpose because they contain information about diagnoses, treatments, laboratory results, and patient characteristics. Nevertheless, data collected for individual patient care may not automatically be appropriate for population-level analysis.

Weiskopf and Weng [3] highlighted the importance of evaluating data quality before reusing electronic health records for research. Differences in documentation practices, missing information, and variations in recording frequency can introduce systematic errors into secondary analyses. Consequently, researchers must evaluate the suitability of available information according to the specific objectives and requirements of each investigation.

5. CRITICAL CHALLENGES IN DATA QUALITY AND INFORMATION INTEGRATION

Despite technological progress, several interconnected challenges continue to limit the effective implementation of integrated public health information systems. These challenges can be classified into technical, organizational, methodological, and governance-related dimensions. Addressing only the technological dimension is insufficient because information systems operate within broader institutional and regulatory environments.

First, heterogeneous data structures and inconsistent terminology complicate the exchange and interpretation of information across different platforms. Second, incomplete reporting, inaccurate records, delayed submissions, and duplicate entries can undermine the reliability of integrated databases. Third, inadequate infrastructure and limited technical expertise may prevent smaller healthcare institutions from participating effectively in information-sharing networks.

Organizational fragmentation represents another major obstacle because healthcare institutions may have different priorities, administrative procedures, and incentives for sharing information. Vest and Jasperson [6] discussed the importance of measuring actual HIE utilization rather than focusing exclusively on the availability of exchange infrastructure. This distinction is important because the existence of a technically connected system does not necessarily indicate meaningful information sharing or practical improvements in public health activities.

Privacy, confidentiality, and information security also require careful consideration. The integration of health information increases the potential exposure of sensitive patient and population-level data, particularly when multiple institutions gain access to shared databases. Appropriate access controls, data protection policies, anonymization procedures, and institutional accountability are therefore essential components of sustainable information integration.

Finally, differences in institutional capacity can produce unequal benefits from digital transformation. Organizations with limited financial resources, inadequate infrastructure, or insufficiently trained personnel may struggle to adopt standardized systems and maintain reliable information exchange. This situation can perpetuate existing inequalities in public health information availability and reduce the completeness of population-level surveillance.

6. FUTURE DIRECTIONS

Improving public health informatics requires a coordinated strategy that treats data quality, interoperability, and information integration as mutually dependent priorities. Standardized data definitions, common terminologies, consistent reporting procedures, and reliable data validation mechanisms should be incorporated into information system development from the beginning. Continuous monitoring is also necessary to identify errors and maintain the usefulness of information throughout its lifecycle.

Greater collaboration between healthcare institutions, public health agencies, technology developers, and regulatory authorities is equally important. Shared governance frameworks can establish clear responsibilities for data ownership, access, exchange, quality assurance, and information security. In addition, investment in institutional infrastructure and workforce development is essential to ensure that interoperability initiatives can be implemented consistently across different healthcare settings.

Another important direction is the development of information systems capable of supporting integrated epidemiological surveillance and population health analysis. Linking clinical, laboratory, demographic, and administrative information can strengthen the ability of public health organizations to identify emerging health problems and evaluate intervention outcomes. However, these developments must be supported by rigorous data quality assessment and transparent information governance to avoid generating misleading conclusions from integrated datasets.

7. CONCLUSION

Data quality, interoperability, and information integration are essential foundations of effective public health informatics. Reliable information is necessary for accurate epidemiological analysis, while interoperability enables different institutions to exchange and interpret information consistently. Information integration connects these capabilities to support comprehensive public health surveillance, research, planning, and decision-making.

This review demonstrates that the principal barriers to effective integration are not exclusively technological but also involve organizational fragmentation, inconsistent data standards, inadequate governance, privacy concerns, and unequal institutional capacity. Sustainable improvements therefore require coordinated investments in data quality management, standardized information exchange, semantic interoperability, and institutional collaboration. Ultimately, public health informatics should be developed as an integrated information ecosystem in which reliable and meaningful data can be transformed into actionable evidence for improving population health.

REFERENCES

[1] J. R. Vest, H. M. Kirk, and L. M. Issel, "Quality and integration of public health information systems: A systematic review focused on immunization and vital records systems," *Online Journal of Public Health Informatics*, vol. 4, no. 2, p. e61035, Sep. 2012, doi: 10.5210/ojphi.v4i2.4198.

[2] N. G. Weiskopf and C. Weng, "Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 144–151, Jan. 2013, doi: 10.1136/amiajnl-2011-000681.

[3] N. G. Weiskopf, G. Hripcsak, S. Swaminathan, and C. Weng, "Defining and measuring completeness of electronic health records for secondary use," *Journal of Biomedical Informatics*, vol. 46, no. 5, pp. 830–836, Oct. 2013, doi: 10.1016/j.jbi.2013.06.010.

[4] S. T. Liaw *et al.*, "Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature," *International Journal of Medical Informatics*, vol. 82, no. 1, pp. 10–24, Jan. 2013, doi: 10.1016/j.ijmedinf.2012.10.001.

[5] J. R. Vest and L. D. Gamm, "Health information exchange: Persistent challenges and new strategies," *Journal of the American Medical Informatics Association*, vol. 17, no. 3, pp. 288–294, May 2010, doi: 10.1136/jamia.2010.003673.

[6] J. R. Vest and J. Jasperson, "What should we measure? Conceptualizing usage in health information exchange," *Journal of the American Medical Informatics Association*, vol. 17, no. 3, pp. 302–307, May 2010, doi: 10.1136/jamia.2009.000471.

[7] J. R. Vest, H. Zhao, J. Jaspserson, L. D. Gamm, and R. L. Ohsfeldt, "Factors motivating and affecting health information exchange usage," *Journal of the American Medical Informatics Association*, vol. 18, no. 2, pp. 143–149, Mar. 2011, doi: 10.1136/jamia.2010.004812.

[8] J. R. Vest, T. R. Campion, and R. Kaushal, "Challenges, alternatives, and paths to sustainability for health information exchange efforts," *Journal of Medical Systems*, vol. 37, no. 6, p. 9987, Dec. 2013, doi: 10.1007/s10916-013-9987-7.