

UDC 004.65:004.62(045)

DOI:10.18372/1990-5548.89.21477

¹Dmitry Persunov,²Natalia Apenko

EVALUATION CRITERIA SYSTEMS FOR DATA INTEGRITY IN MULTI-SOURCE INFORMATION ENVIRONMENTS

Department of Intelligent Cybernetic Systems, National University "Kyiv Aviation Institute", Kyiv, Ukraine

E-mails: ¹5812850@stud.kai.edu.ua ORCID 0009-0000-5877-3493

²nataliia.apenko@npp.kai.edu.ua ORCID 0000-0001-6891-0869

Abstract—It is proposed a system of criteria for assessing data integrity, based on the formalised description of inter-source consistency, information completeness, and temporal relevance of data. A mathematical model of an integral integrity indicator has been developed, allowing quantitative assessment of the data state in distributed information environments while considering weighting coefficients of sources and attributes. The obtained results confirmed the adequacy of the proposed approach, its sensitivity to conflicting and incomplete data, as well as its suitability for automated monitoring, identification of problematic records, and decision support in data management systems. The simulation results showed that the integral data integrity indicator varied approximately from 0.30 to 0.90 depending on the level of inter-source discrepancies, missing data, and temporal inconsistency.

Keywords—Data integrity, data consistency, completeness of information, composite indicator, information completeness, mathematical model, information systems.

I. INTRODUCTION

Modern information systems operate in environments where data are simultaneously generated, transmitted, and processed from multiple independent sources. Such environments are characterised by heterogeneity of data formats, semantic differences between attributes, asynchronous updates, incomplete records, and varying levels of source reliability. The integration of distributed information flows significantly complicates the process of ensuring data integrity, since inconsistencies, duplicated records, temporal conflicts, and missing values may occur during aggregation and synchronisation of information. Under these conditions, the reliability of analytical results and decision-making processes directly depends on the quality and integrity of integrated data.

Traditional approaches to data integrity control are mainly focused on maintaining structural constraints within individual databases, including key dependencies, referential integrity, and transaction consistency. However, such approaches are insufficient for multi-source information environments, where integrity should be considered not only as structural correctness, but also as a complex property that combines consistency, completeness, relevance, and trustworthiness of data obtained from heterogeneous sources. As a result, there is a need for new approaches capable of quantitatively evaluating the integrity of integrated

information resources under conditions of uncertainty and distributed data interaction.

A significant number of recent studies are devoted to data quality assessment, distributed storage reliability, cloud integrity verification, and secure information integration. Existing methods mainly address isolated aspects of integrity control, including verification of consistency, protection against unauthorised modifications, and detection of missing values.

Nevertheless, most existing approaches do not provide a unified criterion capable of simultaneously considering inter-source consistency, completeness of information, temporal relevance, and source reliability within a single integrated model. This limitation reduces the efficiency of automated monitoring and decision-support mechanisms in distributed information systems.

Therefore, the development of a formalised system of criteria and a mathematical model for integrated assessment of data integrity in multi-source environments is an important scientific and practical task. The proposed approach aims to provide quantitative evaluation of integrity indicators and support automated detection of conflicting, incomplete, or outdated data in distributed repositories and information systems.

II. LITERATURE ANALYSIS

Modern scientific studies devoted to data integrity in distributed information systems mainly focus on ensuring reliability of information exchange, secure

storage of data, consistency verification, and intelligent processing of heterogeneous information flows [1] – [5].

Existing approaches investigate the problems of integrity preservation in cloud environments, distributed repositories, communication networks, and adaptive information systems operating under conditions of uncertainty and asynchronous updates [4], [5], [10].

Research related to adaptive information exchange and telemetry-based analysis demonstrates the importance of maintaining stable and reliable data transmission in distributed systems with dynamically changing communication conditions [1].

Such approaches confirm the necessity of integrity monitoring mechanisms capable of functioning under heterogeneous and unstable information flows.

A significant number of studies are devoted to integrity verification mechanisms in cloud and distributed storage systems [2], [4], [5], [11].

Existing methods mainly focus on detecting unauthorised modifications, maintaining consistency of distributed repositories, and improving the security of information storage. Blockchain technologies and decentralised verification mechanisms are also actively investigated as tools for increasing reliability and transparency of information processing [2].

However, most current approaches are concentrated on storage security and cryptographic protection, while insufficient attention is paid to integrated assessment of consistency, completeness, and temporal relevance of heterogeneous data obtained from multiple independent sources.

Separate scientific works investigate cryptographic methods and secure communication protocols for distributed environments [3], [8].

Existing approaches propose mechanisms for protected information exchange, pseudo-random sequence generation, and quantum-secure communication.

These methods increase resistance to external interference and unauthorised access but do not solve the problem of comprehensive quantitative assessment of data integrity in multi-source information environments.

Modern intelligent information systems increasingly apply machine learning methods, adaptive filtering algorithms, and optimisation approaches for processing heterogeneous information flows [9], [12].

Such methods improve the efficiency of information analysis, conflict detection, and decision-making under uncertainty.

In addition, multidimensional time-series analysis approaches are actively used for modelling asynchronous and distributed data streams [13]. Nevertheless, existing intelligent approaches generally focus on isolated data-processing tasks and do not provide a unified mathematical criterion for integrity assessment.

Current studies also consider mathematical support methods for decision-support systems and adaptive information technologies operating under conditions of incomplete and uncertain information [6], [7].

These approaches provide mechanisms for processing heterogeneous data and improving the reliability of analytical results. However, the problem of integrated evaluation of data integrity in multi-source environments remains insufficiently investigated.

Therefore, the analysis of existing scientific publications shows that most current approaches address isolated aspects of integrity control, including cryptographic protection, secure storage, consistency verification, or intelligent filtering of information [2] – [15].

At the same time, insufficient attention is paid to the development of integrated mathematical models capable of simultaneously considering inter-source consistency, completeness, temporal relevance, and trustworthiness of data within a unified evaluation framework.

This confirms the relevance of developing a formalised system of criteria and an integral indicator for assessing data integrity in multi-source information environments.

III. PROBLEM STATEMENT

The problem considered in this study is the quantitative assessment of data integrity in a multi-source information environment under conditions of inter-source discrepancies, missing values, asynchronous updates, and different levels of source reliability. Unlike conventional database integrity control, the assessment should reflect the actual consistency and relevance of alternative representations of the same information object.

Let the multi-source information environment be represented by a set of data sources:

$$S = \{s_1, s_2, \dots, s_M\},$$

and a set of attributes:

$$A = \{a_1, a_2, \dots, a_K\}.$$

For an information object with identifier i , the value of attribute k obtained from source s is denoted

$x_{ik}^{(s)}$. Each source is characterised by a reliability weight w_s while the state of the data is additionally determined by the availability of attribute values and the time of their update.

The task is to determine an integral data integrity indicator I_i as a mapping of the set of multi-source observations:

$$I_i = \mathcal{F}(\{x_{ik}^{(s)}\}, \{w_s\}, \{m_{ik}^{(s)}\}, \{t_{ik}^{(s)}\}), \quad I_i \in [0, 1].$$

where $m_{ik}^{(s)}$ characterises the availability of the corresponding attribute value and $t_{ik}^{(s)}$ is its last update time. The required indicator should satisfy the condition that increasing discrepancies between data sources, the number of missing values, or temporal inconsistency leads to a decrease in I_i whereas increasing consistency, completeness, and relevance leads to its increase.

The required indicator must decrease with increasing inter-source discrepancy, proportion of missing values, and temporal inconsistency, while increasing with higher consistency, completeness, and relevance of the data. Thus, the research task is to develop a system of criteria and a mathematical model that provide an integrated quantitative assessment of data integrity and enable the identification of conflicting, incomplete, or outdated records.

IV. MATHEMATICAL MODEL FOR DATA INTEGRITY ASSESSMENT

In modern information environments, data are generated, stored, and processed simultaneously from multiple independent sources. Such environments are characterised by heterogeneity of formats, differences in attribute semantics, asynchronous updates, varying quality of sources, and the possibility of value conflicts. Under these conditions, the classical understanding of integrity as merely compliance with database constraints (keys, referential integrity) is insufficient.

In a multi-source information environment, data integrity acquires a multidimensional nature and should be considered as an integral property of data that includes structural consistency, logical coherence, completeness, relevance, reliability, and inter-source consistency.

Therefore, the assessment of data integrity requires the formation of a system of criteria that makes it possible to quantitatively describe each of these components and integrate them into a unified indicator.

The system of criteria for assessing data integrity is a formalised set of interrelated indicators that characterise various aspects of consistency, completeness, reliability, and relevance of information in a multi-source environment.

From the perspective of the systems approach, a multi-source environment is considered as a set of data flows interacting through integration mechanisms.

Each object attribute may have multiple representations obtained from different sources. Therefore, integrity assessment should take into account not only compliance with constraints but also the degree of consistency between alternative representations of the same information fragment.

At the theoretical level, the proposed evaluation framework consists of three interconnected criteria systems: the inter-source consistency criteria system, the data completeness criteria system, and the temporal relevance criteria system. Each system evaluates a separate component of data integrity, while their joint use provides an integrated assessment of the state of data in a multi-source information environment.

The inter-source consistency criteria system characterises the degree of correspondence between the values of the same attribute obtained from different sources. Its purpose is to quantify discrepancies between alternative representations of the same information object.

The data completeness criteria system characterises the availability of the required attribute values and reflects the influence of missing data on the overall integrity assessment.

The temporal relevance criteria system characterises the validity of data with respect to their update time and makes it possible to take into account asynchronous changes in information received from different sources.

The proposed system combines three interconnected components: inter-source consistency, data completeness, and temporal relevance, while taking into account the reliability of individual information sources.

On this basis, a mathematical model for forming an integral integrity indicator is developed, which provides a continuous normalised assessment of the data state and enables the identification of conflicting, incomplete, or outdated records.

In this study, three interconnected evaluation criteria systems are developed for assessing data integrity in multi-source information environments: an inter-source consistency criteria system, a data completeness criteria system, and a temporal relevance criteria system.

Their integration, together with source and attribute weighting mechanisms, forms a mathematical model for calculating a normalised integral data integrity indicator.

Thus, the proposed solution provides both component-wise assessment of individual integrity properties and their aggregation into a single quantitative criterion.

Let there be a set of sources ($S = 1, \dots, M$).

The data are described by a set of attributes $A = \{1, \dots, K\}$.

For an object with identifier i , the source s provides the value of attribute k : $(x_{i,k}^{(s)})$.

Step 1. The proposed mathematical model of data integrity assessment is defined as a mapping. The computational procedure for evaluating the components of the proposed mathematical model is given below.

A normalisation mapping is introduced:

$$y_{i,k}^{(s)} = f_k(x_{i,k}^{(s)}).$$

Step 2. Source quality assessment.

For each source, a trust weight is introduced:

$$w_s = \frac{\alpha r_s + (1 - \alpha) q_s}{\sum_{j=1}^M (\alpha r_j + (1 - \alpha) q_j)}.$$

Step 3. Determination of the consistent (reference) value from multiple sources.

The reference value for attribute k is defined as:

$$\hat{y}_{i,k} = \sum_{s=1}^M w_s y_{i,k}^{(s)}.$$

Step 4. Measure of discrepancy (conflict) between sources.

A distance (discrepancy) function d_k is introduced for attribute k . The weighted average discrepancy is calculated as:

$$D_{i,k} = \sum_{s=1}^M w_s d_k(y_{i,k}^{(s)}, \hat{y}_{i,k}), (d_k(a, b) = (a - b)^2).$$

Step 5. Consistency criterion as a component of integrity.

The criterion is normalised within the interval $[0; 1]$ the greater the discrepancy, the lower the integrity value.

Exponential smoothing is applied:

$$I_{i,k} = \exp(-\beta_k D_{i,k}).$$

Step 6. Completeness criterion (missing values as a component of integrity).

The presence indicator of a value is defined as:

$$C_{i,k} = \sum_{s=1}^M w_s m_{i,k}^{(s)}.$$

Step 7. Relevance criterion (temporal consistency).

Let $t_{i,k}^{(s)}$ be the last update time of attribute k from source s . Then the temporal relevance criterion is:

$$R_{i,k} = \sum_{s=1}^M w_s \exp(-\lambda(T - t_{i,k}^{(s)})).$$

Step 8. Integral integrity indicator for a record. Weights of attributes are introduced:

$$\gamma_k \geq 0, \sum_{k=1}^K \gamma_k = 1.$$

Step 9. Decision-making rule:

$$i = \begin{cases} 1, & J_i > \tau \\ 0, & J_i < \tau \end{cases}.$$

The proposed mathematical model provides quantitative assessment of data integrity while taking into account the multidimensional nature of the information environment.

Unlike classical approaches that consider integrity as a binary characteristic (satisfied/not satisfied), the proposed model forms a continuous integral assessment.

The main properties of the model include normalisation of indicators, which ensures comparability of results; weighting coefficients for sources and attributes; resistance to conflicts due to the use of weighted aggregated values; scalability, which allows the model to be applied to an arbitrary number of sources; and the possibility of automated decision-making through threshold classification.

The integral integrity indicator makes it possible to rank records according to their reliability level, identify problematic data requiring clarification, form mechanisms for information recovery or correction, and ensure monitoring of the data state in a dynamic environment.

The obtained system of criteria and its mathematical implementation can be used as a basic component of information technology for ensuring data integrity in multi-source repositories.

In order to verify the adequacy of the proposed system of criteria for assessing data integrity, numerical modelling of a multi-source information environment was carried out.

Within the simulation framework, a set of records was generated in which attribute values were obtained from several independent sources with

different characteristics of quality, completeness, and relevance.

For each record, an integral integrity indicator was calculated, aggregating the criteria of consistency, completeness, and temporal relevance while taking into account trust weights of the sources.

The obtained results make it possible to evaluate the distribution of integrity values and identify records that do not satisfy the established threshold level. The simulation results of the integral integrity indicator are presented in Fig. 1.

In a multi-source information environment, the level of data integrity is determined not by isolated characteristics but by the combined influence of several factors that have a nonlinear effect on the generalised assessment.

From the perspective of systems theory, integrity is an integrative property that arises as a result of interaction between alternative representations of the same information object.

Even minor discrepancies between sources, when accumulated, can significantly reduce the reliability of the integrated result.

In addition, information in distributed environments is characterised by dynamism and incompleteness, which determines the necessity of considering temporal relevance and structural completeness of data.

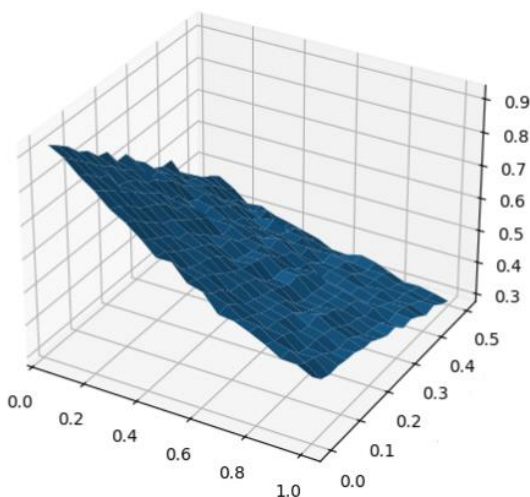


Fig. 1. Modelling results

Theoretically, the integral integrity indicator should be sensitive to variations of these parameters and demonstrate predictable changes when the level of conflict increases or the proportion of missing values becomes larger.

The values of the integral indicator vary within the interval $[0;1]$, which confirms the normalisation of the proposed system of criteria.

A part of the records exceeds the threshold value, indicating an acceptable level of data consistency and completeness. At the same time, a group of records with reduced values of the integral indicator was identified, which can be classified as potentially conflicting or incomplete.

The obtained distribution demonstrates the sensitivity of the model to discrepancies between sources and the presence of missing values, confirming its suitability for monitoring the state of data in a multi-source information environment.

VI. CONCLUSIONS

The main result of the study is a mathematical model for integrated data integrity assessment in multi-source information environments based on the criteria of inter-source consistency, completeness, temporal relevance, and source reliability.

The developed system of criteria considers consistency, completeness, temporal relevance, and reliability of information sources within a unified mathematical framework, which allows integrated evaluation of heterogeneous data under conditions of uncertainty and asynchronous updates.

A mathematical model of the integral integrity indicator has been developed to provide continuous quantitative assessment of data quality and consistency.

Unlike traditional approaches based on binary integrity verification, the proposed model enables detection of conflicting, incomplete, or outdated records and supports automated monitoring of distributed information repositories.

The conducted numerical simulation confirmed the adequacy and sensitivity of the proposed approach to discrepancies between sources, missing values, and temporal inconsistencies.

The obtained results demonstrated that records with reduced integrity values can be identified using the threshold decision rule, which enables the detection of conflicting, incomplete, or outdated data and their subsequent prioritisation for verification or correction.

The developed system of criteria and its mathematical implementation can be used as a basic component of information technologies for ensuring data integrity in distributed and multi-source information systems.

Further research may be focused on the development of adaptive mechanisms for integrity assessment, intelligent conflict resolution methods, and integration of machine learning approaches for automated data quality management.

REFERENCES

- [1] S. Kashkevych, N. Marchenko, and Y. Dehtiar, "Method for Predicting the Communication Channel Capacity of Unmanned Aerial Vehicles Based on Telemetry Data," *Cybernetics and Computer Technologies*, pp. 120–127, 2026. <https://doi.org/10.34229/2707-451X.26.1.10>
- [2] G. R. Annapureddy, "The Impact of Blockchain on Data Integrity and Security in Distributed Systems," *The Eastasouth Journal of Information System and Computer Science*, vol. 2, no. 2, 2024. <https://doi.org/10.58812/esiscs.v2i02.486>
- [3] S. O. Gnatyuk and Y. A. Burmak, "Method for developing pseudo-random number generators for cryptographic applications in 5G networks," *Scientific journal article*, (ISSN 2663-4023, 2021). <https://doi.org/10.28925/2663-4023.2021.12.151162>
- [4] P. Goswami et al., "Investigation on Storage Level Data Integrity Strategies in Cloud Computing," *Journal of Cloud Computing*, 2024.
- [5] B. A. Jalil et al., "A Survey on Data Integrity," *Verification Schemes in the Cloud. AIP Conference Proceedings*, 2023. <https://doi.org/10.1063/5.0103392>
- [6] S. Kashkevich (Ed.), "Decision support systems: mathematical support," *Kharkiv: TECHNOLOGY CENTER PC*, 2025, 202 p. <https://doi.org/10.15587/978-617-8360-13-9>.
- [7] D. O. Persunov and O. P. Nechyporuk, "Ensuring data integrity in distributed information environments," *Information Technologies: Theory and Practice: Proceedings of the III (IX) International Internet Conference of Higher Education Students and Young Scientists*, 25–27 March 2026, pp. 583–585.
- [8] T. Okhrimenko, S. Dorozhynskiy, and B. Horbakha, "Analysis of quantum secure direct communication protocols," *CSIT Journal*, vol. 5, no.1, 2023. <https://doi.org/10.31891/csit-2023-1-8>
- [9] O. Poliarus, S. Krepych, and I. Spivak, "Hybrid approach for data filtering and machine learning inside content management system," *Advanced Information Systems*, vol. 7(4), pp. 70–74, 2023. <https://doi.org/10.20998/2522-9052.2023.4>.
- [10] C. Shi, A. Manschula, T. Mahmud et al., *Revisiting Computational Storage for Data Integrity and Security*, 2025.
- [11] O. Tachinina, O. Lysenko, S. Ponomarenko, S. Chumachenko, and V. Kutieпов, "Engineering Methodology for the Synthesis of Control Algorithms for Digital Electric Drives of Mechatronic Devices of Flying Search Robots," *Lecture Notes in Networks and Systems*, vol. 996, pp. 427–439, 2024. https://doi.org/10.1007/978-3-031-60549-9_32
- [12] K. A. Tamer, O. Sova, O. Shaposhnikova, V. Yashchenok, I. Stanovska, S. Shostak, O. Rudenk, S. Petruk, O. Matsyi, and S. Kashkevich, "Development of a solution search method using a combined bio-inspired algorithm," *Eastern European Journal of Enterprise Technologies*, vol. 1, no. 4(127), pp. 6–13, 2024. <https://doi.org/10.15587/1729-4061.2024.298205>.
- [13] O. Nechyporuk, & S. Podelskyi, "Information System Functioning Based on Multidimensional Time Series," *Electronics and Control Systems*, 2(88), pp. 61–68, 2026. <https://doi.org/10.18372/1990-5548.88.20967>
- [14] A. E. Quossini, Y. I. Daradkeh, S. M. Al Tabib, S. Gnatyuk, T. Okhrimenko and V. Kinzeryavyy, "Improved Model of Quantum Deterministic Protocol Implementation in Channel with Noise," *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Metz, France, 2019, pp. 572–578, <https://doi.org/10.1109/IDAACS.2019.8924287>
- [15] O. Turovsky, Y. Khlaponin, M. Mohamed, T. Okhrimenko, I. Goncharenko, & M. Iavich, "Combined System of Phase Synchronization with Increased Astatism order in Frequency Monitoring Mode," *International Workshop on Control, Optimisation and Analytical Processing of Social Networks*, 2020.

Received: Yune 28, 2026

Accepted: August 16, 2026

Published: September 18, 2026

Persunov Dmitry. ORCID 0009-0000-5877-3493. Postgraduate student.

Department of Intelligent Cybernetic Systems, National University "Kyiv Aviation Institute", Kyiv, Ukraine.

Education: Kyiv Aviation Institute, Kyiv, Ukraine. (2024).

Research area: information systems; artificial intelligence in network technologies.

Publications: 10.

E-mail: 5812850@stud.kai.edu

Apenko Natalia. ORCID 0000-0001-6891-0869. Candidate of Technical Sciences. Associate Professor.

Department of Intelligent Cybernetic Systems, National University "Kyiv Aviation Institute", Kyiv, Ukraine.

Education: National Aviation University, Kyiv, Ukraine. (2006).

Research area: information systems; decision support systems; artificial intelligence.

Publications: 40.

E-mail: nataliia.apenko@npp.kai.edu.ua

Д. О. Персунов, Н. В. Апенко. Системи критеріїв оцінювання цілісності даних у багатоджерельних інформаційних середовищах

У статті запропоновано систему критеріїв оцінки цілісності даних, що базується на формалізованому описі міжджерельної узгодженості, повноти інформації та часової релевантності даних. Розроблено математичну модель інтегрального показника цілісності, що дозволяє кількісно оцінити стан даних у розподілених інформаційних середовищах з урахуванням вагових коефіцієнтів джерел та атрибутів. Отримані результати підтвердили адекватність запропонованого підходу, його чутливість до суперечливих та неповних даних, а також його придатність для автоматизованого моніторингу, ідентифікації проблемних записів та підтримки рішень у системах управління даними. Результати моделювання показали, що інтегральний показник цілісності даних змінювався приблизно від 0,30 до 0,90 залежно від рівня міжджерельних розбіжностей, відсутніх даних та часової неузгодженості.

Ключові слова: цілісність даних, узгодженість даних, повнота інформації, актуальність даних, інтегральний показник, математична модель, інформаційні системи.

Персунов Дмитро Олександрович. ORCID 0009-0000-5877-3493. Аспірант.

Кафедра інтелектуальних кібернетичних систем, Національний університет «Київський авіаційний інститут», Київ, Україна.

Освіта: Київський авіаційний інститут, Київ, Україна, (2024).

Напрямок наукової діяльності: інформаційні системи; штучний інтелект у мережевих технологіях.

Кількість публікацій: 10.

E-mail: 5812850@stud.kai.edu

Апенко Наталія Вікторівна. ORCID 0000-0001-6891-0869. Кандидат технічних наук. Доцент.

Кафедра інтелектуальних кібернетичних систем, Національний університет «Київський авіаційний інститут», Київ, Україна.

Освіта: Національний авіаційний університет, Київ, Україна, (2006).

Напрямок наукової діяльності: інформаційні системи; системи підтримки прийняття рішень; штучний інтелект.

Кількість публікацій: 40.

E-mail: nataliia.apenko@npp.kai.edu.ua