УДК 005.591.6:004
DOI: https://doi.org/10.32782/2708-0366/2026.30.9

Dymova Hanna
PhD in Technical Sciences,
Department of Management, Marketing and Information Technologies,
Kherson State Agrarian and Economic University
(Kherson / Kropyvnytskyi)
ORCID: https://orcid.org/0000-0002-5294-1756
Zaitsev Ivan
PhD Student,
Odesa State Agrarian University
ORCID: https://orcid.org/0009-0009-4911-7529

Димова Г.О.
Херсонський державний аграрно-економічний університет
(м. Херсон / м. Кропивницький)
Зайцев І.О.
Одеський державний аграрний університет

ТРАНСФОРМАЦІЯ МЕТОДІВ ФІНАНСОВОЇ ОЦІНКИ МАСИВІВ ВЕЛИКИХ ДАНИХ У ПРОЦЕСІ УПРАВЛІННЯ ЦИФРОВИМИ ПЛАТФОРМАМИ

TRANSFORMATION OF FINANCIAL VALUATION METHODS FOR BIG DATA ARRAYS IN DIGITAL PLATFORM MANAGEMENT

The article conducts a comprehensive scientific study of the transformation of financial valuation methods for Big Data arrays within the operational and strategic management of modern digital platforms. It is substantiated that amid global digitalization and widespread adoption of artificial intelligence technologies, information assets have evolved from an operational byproduct into a primary intangible asset. The study analyzes the key limitations of traditional valuation approaches (cost, market, and income) when applied to large-scale data arrays. The paper systematizes novel customer-centric valuation models based on Customer Lifetime Value (CLV) and audience engagement metrics. It proves the high efficiency of algorithmic game-theory models, specifically Data Shapley, for ensuring equitable financial distribution and measuring the marginal contribution of specific datasets to machine learning model accuracy. Furthermore, the direct impact of Data Governance frameworks, Data Quality attributes, and modern storage architectures (Data Lakehouse) on reducing infrastructure costs and maximizing return on investment is thoroughly revealed.

Keywords: *Big Data, digital platform, financial valuation, Data Shapley, Customer Lifetime Value.*

У статті здійснено комплексне наукове дослідження процесів трансформації підходів та методів фінансової оцінки масивів великих даних у системі операційного та стратегічного управління сучасними корпоративними цифровими платформами. Обґрунтовано, що в умовах глобальної цифровізації та масового впровадження технологій штучного інтелекту інформаційні масиви еволюціонували від побічного продукту операційної діяльності до статусу найбільш цінного нематеріального активу. Проаналізовано фундаментальний парадокс сучасної фінансової звітності, який полягає у нездатності класичних міжнародних стандартів обліку (зокрема МСФЗ 38) повноцінно відображати реальну ринкову вартість органічно згенерованих цифрових активів. Досліджено критичні обмеження традиційних методів фінансової оцінки (витратного, ринкового та прибуткового) при їх застосуванні до масивів Big Data. Систематизовано та адаптовано новітні клієнтоорієнтовані моделі оцінки, зокрема теорію оцінки

© Dymova Hanna, Zaitsev Ivan, 2026
Стаття поширюється на умовах ліцензії відкритого доступу (CC BY 4.0)

клієнтів (CVT), капіталізацію показника пожиттєвої цінності (CLV) та моделювання залученості аудиторії. Доведено високу ефективність застосування алгоритмічного апарату кооперативної теорії ігор, а саме вектора Шеплі (Data Shapley), для забезпечення справедливого розподілу фінансової цінності та визначення граничного внеску окремих дата-сетів у точність функціонування моделей машинного навчання. Розкрито прямий взаємозв'язок між зрілістю процесів корпоративного управління даними (Data Governance), атрибутами їх якості (точність, повнота, узгодженість, своєчасність) та підсумковою фінансовою вартістю інформаційного активу. Доведено економічну доцільність переходу цифрових платформ до новітніх архітектурних парадигм зберігання й обробки даних (Data Lakehouse та потоковий аналіз у режимі реального часу) з метою мінімізації інфраструктурних витрат та оптимізації показника повернення інвестицій (ROI). Запропоновано концептуальну схему інтеграції технологічної інфраструктури, систем управління даними та алгоритмів оцінки.

Ключові слова: *великі дані, цифрова платформа, фінансова оцінка, вектор Шеплі, пожиттєва цінність клієнта.*

Formulation of the problem. In the modern global economic landscape, the paradigm of information perception has undergone fundamental and irreversible changes. Over recent decades, information has evolved from the status of an exclusively operational byproduct into one of the most valuable strategic assets of a modern enterprise. The majority of assets belonging to large multinational corporations today are intangible, which is directly driven by rapid digital transformation processes and the deep integration of artificial intelligence (AI) systems into business processes. Successful tech leaders such as Amazon, Meta, Tencent, and Walmart have built their dominant business models on the ability to continuously aggregate, systematically analyze, and effectively monetize colossal arrays of data regarding users, transactions, and behavioral patterns.

The concept of "Big Data" goes far beyond the simple accumulation of information. According to the "5V" concept (Volume, Velocity, Variety, Veracity, Value), data acquires the status of big data only when its volume, rate of arrival (in real time), and variety (structured and unstructured formats) require the implementation of fundamentally new technological architectures, such as Data Lakehouse or streaming brokers. The global market for big data solutions, which was valued at $185 billion in 2023, is expected to reach $383.4 billion by 2030, demonstrating a compound annual growth rate (CAGR) of 11.0%.

However, despite the general recognition of the potential of big data to stimulate economic growth, optimize processes, and create new markets, a significant portion of companies face serious difficulties regarding its effective commercialization and overcoming technological and organizational barriers [9, 11]. A fundamental obstacle for managing digital platforms is the problem of an adequate financial valuation of accumulated information arrays. Without an understanding of the exact value of a digital asset, it is impossible to build an effective strategy for its capitalization, optimize infrastructure costs for storing petabytes of information, and fairly distribute profits in complex ecosystems (for example, AI training consortia). This makes the development, adaptation, and implementation of new methods for financial data valuation a critically important scientific and practical task for modern management of digital platforms.

Analysis of recent research and publications. The issue of recognition and valuation of intangible assets, specifically digital data, is actively discussed in the scientific and professional financial environment. From the perspective of classical economic theory, the institutionalization of big data as a full-fledged asset on a company's balance sheet faces significant bureaucratic and methodological obstacles. Traditional approaches are governed by International Financial Reporting Standards (IFRS), particularly IAS 38 "Intangible Assets" [1] and US GAAP. Researchers, notably D. Laney, the founder of the "Infonomics" concept, point out the limitations of these standards [2]. They require clear identification of the costs incurred to create the asset and guarantees of future economic benefits, which is

extremely difficult to apply to organically generated (internally created) digital arrays of a company.

Modern scientific discourse highlights attempts to adapt basic valuation approaches (cost, market, income) to the needs of the digital economy. However, leading tech companies and academic institutions are developing their own customer-centric models, such as Customer Valuation Theory – research by V. Kumar [3], and algorithmic approaches based on game theory – the Data Shapley concept proposed by A. Ghorbani and J. Zou [4]. A separate line of research focuses on the link between the value of data, its quality, and the infrastructure level – the works of J. Rowley regarding the DIKW hierarchy [5].

Unresolved parts of the general problem: despite the presence of a wide range of theoretical approaches, the understanding of how exactly to integrate traditional financial methods with novel algorithmic models (for instance, the Shapley Vector) remains fragmented. There is a lack of comprehensive studies linking the transformation of financial valuation methods directly with the operational management of digital platforms, architectural evolution (Data Lakehouse), and corporate Data Governance systems.

Formulation of the purpose of the article. The objective of the article is a comprehensive study of the transformation process of financial valuation methods for Big Data arrays, as well as the systematization of novel algorithmic, customer-centric, and infrastructure valuation models that are critically necessary for effective management, monetization, and asset capitalization of modern corporate digital platforms.

Presentation of the main material. In accordance with applicable International Financial Reporting Standards, information generated internally by a digital platform (e.g., transaction logs, on-site customer behavior patterns, IoT device telemetry) is generally not recognized as an intangible asset in financial statements. This paradox of the digital economy leads to the market capitalization of technology companies often exceeding the book value of their tangible assets tenfold.

The evolution of corporate data valuation practices is gradually shifting from conservative methods to complex computational algorithms that account for network effects [6]. Basic approaches borrowed from physical asset valuation practice have been modified for big data arrays; however, they retain significant limitations (Table 1).

Table 1

Comparative analysis of basic methods for the financial evaluation of data sets

Assessment method	Basic calculation principle	Benefits for platform management	Conceptual Limitations in the Era of Big Data
Cost-Based	The sum of historical costs for collection infrastructure (IoT, API), cleaning, storage (cloud servers), and administration of the data array.	Objectivity, audit transparency, ease of calculation in the presence of receipts for IT services.	Complete ignorance of economic benefit. Storage costs for a petabyte of "junk" logs equal the costs for high-value customer profiles.
Market-Based	Comparative analysis of cost with similar transactions on open data marketplaces (Data Marketplaces).	Reflects the actual purchasing power of the market and demand for specific information segments.	Lack of highly liquid and transparent markets for specific corporate databases, along with their unique structure.
Income-Based	Calculation of net present value (*NPV*) of future cash flows generated due to data.	Directly links the data array with the business result (sales growth, profit margin).	High complexity of forecasting; it is difficult to isolate the impact of a specific dataset from other product success factors.

Source: author's development

The income method for digital corporations most frequently uses the discounted cash flow (*DCF*) model, which is mathematically expressed as:

$$NPV = \sum_{t=1}^{T} \frac{CF_t}{(1+r)^t} - C_0$$

where CF_t is the additional cash flow in period *t*, generated through data monetization (for example, via an improved recommendation system); *r* is the discount rate that accounts for project risks (including risks of leakage or loss of information relevance); C_0 is the initial investment in the collection and processing of the dataset.

Aware of the limitations of traditional accounting, e-commerce platforms and social networks are shifting their focus from the data itself to the value of the subjects of this data.

Customer Valuation Theory (*CVT*) is based on the Customer Lifetime Value (*CLV*) metric [3, 8]. This metric is modeled by machine learning algorithms based on big data analytics, predicting future revenue from cross-selling. In the context of platform management, the value of the data array equals the aggregate increase in *CLV* across all users achieved through personalization:

$$CLV = \sum_{t=1}^{n} \frac{(p_t - c_t) \cdot r_t}{(1+d)^t} - AC$$

where p_t is the revenue from the customer in period *t*; c_t is the transaction costs; r_t is the retention rate, which is directly improved by data-driven algorithms; *d* is the discount rate; *AC* is the Customer Acquisition Cost.

For purely media platforms, audience engagement is capitalized through the Assetization of Users model. Dynamic interaction metadata is evaluated, forming an "attention" metric (eyeballs metric) that allows investors to assess the market capitalization of a startup even in the absence of a current direct cash flow.

Today, the most valuable data is used to train predictive machine learning models. In complex ecosystems (for example, a consortium of banks aggregating transaction logs to create a shared anti-fraud system), the problem arises of fairly distributing financial profit among participants, given that the volume and quality of their data differ.

For such scenarios, leading platforms apply cooperative game theory, specifically the calculation of the Shapley Vector (Data Shapley). This mathematical framework allows one to precisely isolate and quantitatively evaluate the marginal contribution of each specific data array to the overall result (for example, improving the accuracy of an AI algorithm by 5%). The Shapley vector formula for evaluating dataset i is as follows:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} \left(v(S \cup \{i\}) - v(S) \right)$$

where *N* is the total set of all datasets (suppliers); *S* is any coalition (subset) of datasets that does not include dataset *i*; $|S|$ is the number of datasets in the coalition; *n* is the total number of datasets; $v(S)$ is the utility function (for example, the accuracy of a neural network trained only on data *S*).

This algorithmic method of financial valuation guarantees fairness: if adding supplier *i*'s data does not increase the model's accuracy (because it duplicates already existing information), its financial value (Shapley Value) will equal zero, regardless of the costs incurred to collect it.

Financial valuation is not static; it is directly proportional to the maturity of Data Governance processes on the digital platform [7]. Systems theory utilizes the DIKW hierarchy (Data – Information – Knowledge – Wisdom), emphasizing that value assessment must account for Utility Stages and the general context of data lifecycle management (Fig. 1) [10]. Raw "Data" in Data Lakes costs less than transformed "Knowledge" in the form of trained AI models.

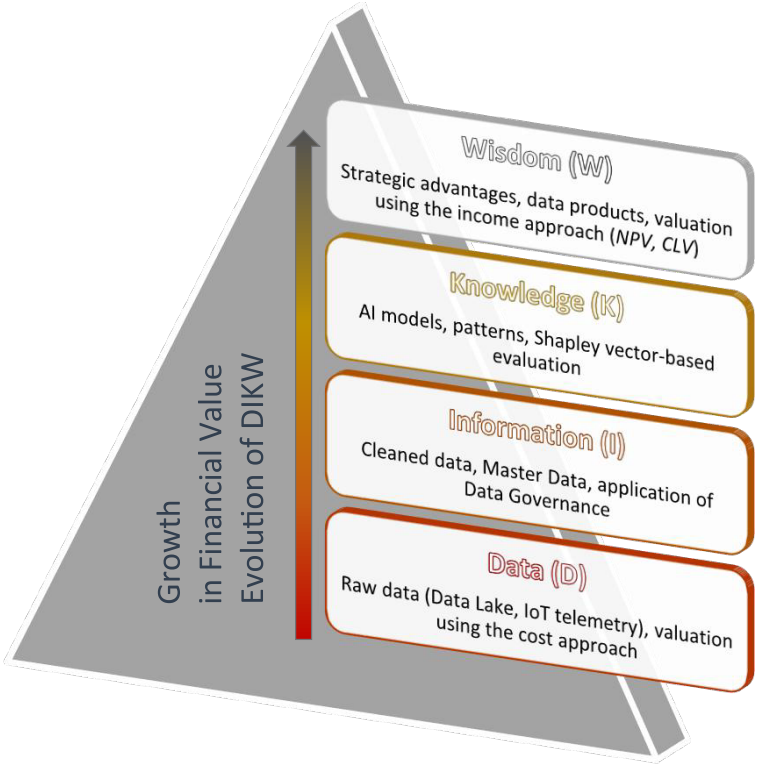


Figure 1. Hierarchical model of the evolution of big data financial value (based on the DIKW hierarchy)

Source: author's development

A critical component that can both multiply and reduce the value of an array to zero is Data Quality [11]. Information with high levels of noise or duplication does not merely lose value, but acquires negative value due to the risk of making loss-making management decisions and algorithmic "hallucinations" of autonomous AI agents (Table 2).

Finally, the chosen storage architecture has a colossal impact on the book value of maintaining information assets (Table 3).

Table 2

Impact of Data Quality attributes on their financial value

Data Quality dimension	Meaning for a digital platform	Impact on financial valuation (monetization)
Accuracy	Objective representation of real objects without distortions.	Low accuracy destroys predictive models, turning *NPV* into a negative value due to losses.
Completeness	Absence of critical gaps (NULL) in important customer attributes.	Limits the possibilities of algorithmic cross-selling, artificially understating the *CLV* metric.
Consistency	Absence of conflicts when comparing data across CRM, ERP, and billing systems.	Increases costs (cost-based valuation) for reconciliation processes and Master Data Management (MDM) operations.
Timeliness	Real-time information updates (Streaming).	Is a key factor for dynamic pricing and anti-fraud systems; outdated data has zero value.

Source: author's development

Table 3

Economic impact of storage architecture evolution on data array management costs

Architectural paradigm	Technological features (Stack)	Economic impact on the platform
Data Warehouse	Rigid structure (schema-on-write), ETL. Example: Oracle, Teradata.	Astronomical costs for storing petabytes due to proprietary hardware. Inability to store audio/video for AI.
Data Lake	Raw format storage (schema-on-read), ELT. Example: Hadoop, Amazon S3.	Extremely cheap storage. However, without Data Governance, it quickly turns into a loss-making "data swamp" (Data Swamp).
Lakehouse	Combining cheap disks with open table formats (Apache Iceberg) and ACID transactions.	Optimal ROI: eliminates infrastructure duplication, reduces operating costs, and allows running BI and ML workloads simultaneously.

Source: author's development

Today, the transition of digital platforms to the Lakehouse architecture (using Apache Iceberg) and stream processing (Apache Kafka, Flink) is an essential prerequisite for optimizing costs and maximizing net profit from data operations [6], forming a unified integrated management mechanism (Fig. 2).

Conclusions. The evolution of the big data concept from a technical abstraction to the status of a strategic intangible asset requires a fundamental shift in the financial management paradigm from digital platforms. The study proved that traditional accounting methods (cost, market, income) demonstrate significant limitations when valuing digital information, as they are unable to fully account for dynamic value, network effects, and utility stages of the asset.

It has been proven that effective management of digital platforms today is impossible without the synergy of modern infrastructure (Data Lakehouse, Stream Processing), governance policies (Data Governance), and the implementation of algorithmic valuation models. Calculating Customer Lifetime Value (CLV), capitalizing user "attention", and

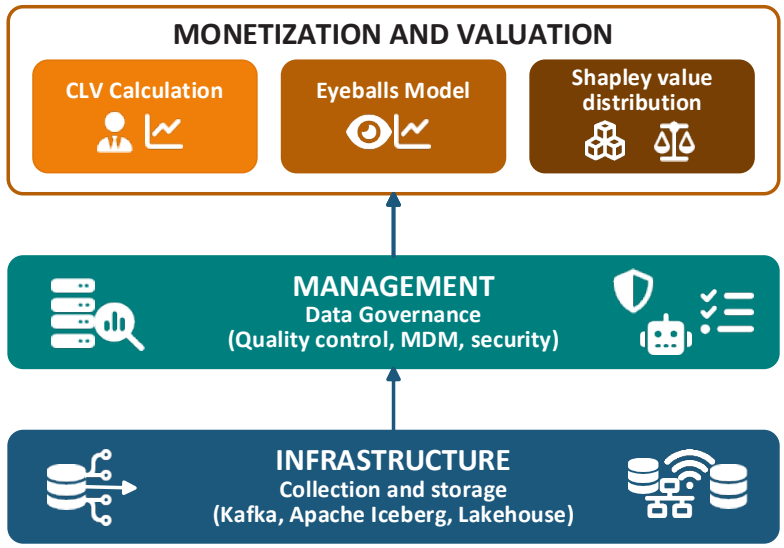


Figure 2. Mechanism for integrating technological architecture and financial valuation algorithms on a digital platform

Source: author's development

using the Shapley vector to assess the marginal contribution of data to the accuracy of AI models are the most relevant tools for the data economy era.

Further research perspectives in this direction consist of developing standardized industry protocols for integrating Data Quality metrics (Data Quality KQI) directly into enterprise financial models, as well as forming a unified regulatory framework for the audit and valuation of data products (Data Products) in global digital markets.

References:

1. Mizhnarodni standarty finansovoi zvitnosti (MFSZ) (2026). [International Financial Reporting Standards (IFRS)]. *IFRS Foundation*. Available at: https://www.ifrs.org (accessed: 10.10.2026). (in Ukrainian)

2. Laney, D. (2018). *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*. Routledge, 352 p.

3. Kumar, V. (2018). Customer Valuation Theory. *Customer Relationship Management*, Springer, Pp. 127–152.

4. Ghorbani, A., & Zou, J. (2019). Data Shapley: Equitable Valuation of Data for Machine Learning. *Proceedings of the 36th International Conference on Machine Learning*, PMLR, Vol. 97, Pp. 2242–2251.

5. Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, vol. 33, No. 2, Pp. 163–180.

6. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.

7. Slota, S. C., et al. (2020). Data governance and data valuation. *Journal of Data and Information Quality*, vol. 12, No. 3, Pp. 1–15.

8. Geva, T., et al. (2017). Using Big Data to evaluate the customer lifetime value. *Journal of Interactive Marketing*, Vol. 38, Pp. 1–15.

9. Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, vol. 60, No. 3, Pp. 285–292.

10. Pejić-Bach, M., et al. (2019). Big data in the context of data lifecycle management. *Strategic Management*, vol. 24, No. 2, Pp. 16–25.

11. Dymova, H. (2026). Information Technologies and Foundations Of Enterprise Digital Transformation in the Context of European Integration. Science and Education in the Third Millennium: information technologies, education, law, psychology, social security and work, management. *Collective monograph. Instytut Spraw Administracji Publiczney (Lublin, Poland)*. Pp. 84–117, DOI: https://doi.org/10.5281/zenodo.20809894

Список використаних джерел:

1. Міжнародні стандарти фінансової звітності (МСФЗ). Фонд МСФЗ. URL: https://www.ifrs.org (дата звернення: 10.10.2026)

2. Laney D. *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*. Routledge, 2018. 352 p.

3. Kumar V. Customer Valuation Theory. *Customer Relationship Management.* Springer, 2018.

4. Ghorbani A., Zou J. Data Shapley: Equitable Valuation of Data for Machine Learning. *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 2019. Vol. 97. P. 2242–2251.

5. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*. 2007. No. 33 (2). P. 163–180.

6. Wirth R., Hipp J. CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. 2000.

7. Slota S. C. et al. Data governance and data valuation. *Journal of Data and Information Quality*. 2020. No. 12 (3). P. 1–15.

8. Geva T. et al. Using Big Data to evaluate the customer lifetime value. *Journal of Interactive Marketing.* 2017. No. 38. P. 1–15.

9. Alharthi A., Krotov V., Bowman M. Addressing barriers to big data. *Business Horizons*. 2017. No. 60 (3). P. 285–292.

10. Pejić-Bach M. et al. Big data in the context of data lifecycle management. *Strategic Management*. 2019. No. 24 (2). P. 16–25.

11. Dymova H. Information Technologies and Foundations Of Enterprise Digital Transformation in the Context of European Integration. Science and Education in the Third Millennium: information technologies, education, law, psychology, social security and work, management. Collective monograph. Instytut Spraw Administracji Publiczney (Lublin, Poland), 2026. Pp.84–117, DOI: https://doi.org/10.5281/zenodo.20809894

Дата надходження статті: 10.10.2026
Дата прийняття статті до публікації: 08.11.2026
Дата публікації статті: 27.11.2026