Data Quality Assessment in Cryptocurrency Exchanges: A Big Data Analytics Approach

Jun-Kuan Tsao*
Department of Data Science
Soochow University
Taipei, Taiwan
junkuancao35@gmail.com
**Corresponding author*

Chiang-Yu Cheng
Department of Data Science
Soochow University
Taipei, Taiwan
taicanccy@gmail.com

***Abstract*— Data quality in cryptocurrency exchanges significantly impacts trading efficiency and market fairness. This study presents a comprehensive multi-dimensional assessment framework to quantitatively evaluate exchange infrastructure quality using big data analytics techniques. Through systematic data collection at 60-second intervals over eight consecutive hours, we obtained 463 valid Bitcoin price samples from three major exchanges (Binance, MEXC, and Bitget) covering Asian and European trading sessions. The framework integrates three core dimensions: consistency (price synchronization with 50,000 coefficient for micro-level variation capture), freshness (information timeliness with 0.2-second threshold), and availability (service reliability). Using weighted averaging (40% consistency, 40% freshness, 20% availability), we compute a composite Data Quality Index (DQI). Results reveal significant disparities: Binance achieves 92.50 points, Bitget 88.61 points, and MEXC 85.36 points, with data freshness emerging as the primary differentiating factor. This framework provides quantitative tools for exchange evaluation, regulatory oversight, and platform selection.**

Keywords—Data Quality Assessment, Cryptocurrency Exchanges, Big Data Analytics, API Performance Evaluation, Real-time Data Analysis

I. INTRODUCTION

Cryptocurrency exchanges serve as critical infrastructure for digital asset trading, processing billions of dollars in transactions daily through automated systems that rely on high-quality real-time data. As of 2025, the global cryptocurrency market capitalization exceeds $4 trillion [1], with trading activity distributed across numerous platforms operating 24/7 globally without interruption. These digital assets have evolved into significant financial instruments [2], with exchanges performing essential functions including price discovery, liquidity provision, and market information transmission [3], making their data service quality fundamental to market efficiency and participant decision-making [4].

However, significant variations exist in data quality across cryptocurrency platforms, creating substantial challenges for automated trading systems, arbitrage algorithms, and market analysis tools. Unlike traditional financial markets with centralized data providers and standardized protocols, cryptocurrency exchanges operate independently with heterogeneous technical architectures, resulting in inconsistent data delivery performance. Research demonstrates that price inconsistencies across Bitcoin markets can persist due to regulatory fragmentation and technical infrastructure differences [5], while arbitrage opportunities arise from systematic latency variations [6]. The high-frequency nature of cryptocurrency trading amplifies these data quality impacts—microsecond-level delays can significantly affect arbitrage returns [7], while price inconsistencies across platforms may lead to suboptimal execution decisions. This information asymmetry problem [8], well-documented in traditional financial markets, manifests more acutely in cryptocurrency trading environments.

Assessing cryptocurrency exchange data quality presents unique big data challenges. Exchanges generate continuous high-frequency streaming data with varying levels of consistency, timeliness, and reliability across platforms and time zones. Traditional data quality assessment methods, designed for batch processing or smaller-scale systems [9], prove inadequate for real-time evaluation of exchange API performance in globally distributed environments. Existing research addresses cryptocurrency market microstructure [3] and general data quality frameworks [9], yet systematic evaluation methodologies specifically targeting exchange infrastructure remain limited. The need for standardized, quantitative assessment frameworks becomes critical as market participants increasingly rely on API data for automated trading strategies and real-time risk management.

This research fills these gaps by establishing a comprehensive multi-dimensional data quality assessment framework for cryptocurrency exchanges. Our contributions include: (1) a systematic evaluation methodology integrating consistency, freshness, and availability metrics tailored for high-frequency trading environments; (2) empirical analysis of three major exchanges (Binance, MEXC, and Bitget) through 463 samples collected at 60-second intervals over eight consecutive hours, covering multiple trading sessions; (3) quantitative evidence of significant quality disparities (7.14-point DQI gap) with data freshness as the primary differentiating factor; and (4) publicly available implementation and dataset [10] ensuring reproducibility and facilitating future research extensions. The framework provides actionable tools for exchange operators to guide infrastructure investments, regulators to establish service quality standards, and traders to make informed platform selection decisions. Beyond technical advancement, this work supports broader objectives of market fairness and sustainable digital financial infrastructure development.

II. LITERATURE REVIEW

A. Market Microstructure Theory and Information Asymmetry

The theoretical foundation for analyzing data quality in financial markets stems from classical market microstructure theory. Kyle (1985) established seminal work linking information asymmetry to market efficiency, demonstrating how informed traders exploit information advantages to obtain excess profits, thereby reducing overall market efficiency [8]. This framework remains relevant in cryptocurrency markets where information asymmetry manifests through differential access to high-quality data streams. Modern cryptocurrency markets exhibit unique microstructural features including continuous 24/7 operation, fragmented liquidity across multiple venues, and varying technical capabilities across platforms [2].

Research on cryptocurrency market microstructure reveals hierarchical price discovery patterns and persistent arbitrage opportunities. Brandvold et al. (2015) documented that larger exchanges often lead price movements while smaller platforms follow [3], establishing a foundation for understanding how exchange size correlates with data quality and market influence. Makarov and Schoar (2020) provided comprehensive evidence of trading patterns and arbitrage activities across cryptocurrency exchanges, finding that price deviations persist even between major platforms due to technical and regulatory barriers [6]. Their analysis of cross-exchange arbitrage documented that profitable opportunities exist when traders have access to superior data feeds and lower latency connections.

B. Data Quality Assessment Frameworks

Data quality assessment methodologies provide the theoretical foundation for this research. Wang and Strong (1996) established a seminal multi-dimensional information quality framework, categorizing data quality into four dimensions: intrinsic, contextual, representational, and accessibility [9]. This framework emphasizes that effective assessment must consider objective indicators across multiple dimensions, as single-dimension evaluations cannot comprehensively reflect data quality states. Their work demonstrates that data consumers' requirements extend beyond simple accuracy to encompass consistency, timeliness, and completeness—dimensions particularly critical in high-frequency trading environments.

Our study builds upon this theoretical foundation by adapting these dimensions to the unique characteristics of cryptocurrency exchange APIs. We operationalize consistency through price synchronization metrics that capture micro-level variations across platforms, timeliness through precise request-response latency measurements, and accessibility through service availability rates. This adaptation addresses the specific challenges of continuous real-time data streams in globally distributed trading environments.

C. Cryptocurrency Data Quality and Market Efficiency

Recent empirical research highlights systematic data quality problems across cryptocurrency platforms. Schwenkler et al. (2025) analyzed multiple cryptocurrency data providers from 2018-2024, finding that approximately 21% of coin identifiers were replaced without disclosure and 70% of daily closing prices deviated more than 5% from peer median values [11]. These findings underscore the critical need for standardized data quality assessment frameworks and validate our focus on establishing reproducible evaluation methodologies. The documented inconsistencies extend beyond price data to include metadata reliability, trading volume accuracy, and timestamp synchronization.

Pieters and Vivanco (2017) examined how financial regulations and technical infrastructure differences create price inconsistencies across Bitcoin markets, documenting that regulatory fragmentation leads to persistent arbitrage opportunities and market segmentation [5]. Their work demonstrates that institutional factors interact with technical capabilities to influence data quality and market integration. Baur et al. (2018) investigated whether Bitcoin functions as a medium of exchange or speculative asset, finding that its market characteristics—including high volatility and price sensitivity to information—amplify the importance of data quality for both traders and investors [12].

D. Data Timeliness and High-Frequency Trading

Data freshness assumes particular importance in cryptocurrency markets due to high volatility and continuous trading. Alexander (2025) provides empirical evidence through analysis of high-frequency order flows at institutional-grade exchanges, demonstrating that microsecond-level delays significantly impact arbitrage returns and liquidity estimation [7]. The study documented latency variations of 30-40 milliseconds between exchanges and quantified that high-frequency traders with lower latency captured profit margins of 15-20 basis points per trade relative to competitors with higher latency. These findings directly inform our framework's emphasis on data freshness as a core assessment dimension, with our 0.2-second threshold derived from documented trading performance impacts.

Market volatility research further supports the criticality of timely data. Hafner (2018) documented significant time-varying volatility in cryptocurrency markets, demonstrating that price movements exhibit regime-switching behavior with periods of extreme volatility [13]. During these high-volatility periods, even small delays in data delivery can result in substantial trading losses or missed opportunities. Liu and Tsyvinski (2021) analyzed risk-return characteristics of cryptocurrencies, finding that their unique risk exposures and return predictability patterns require timely and accurate data for effective portfolio management [14].

E. Research Gaps and Contributions

Despite substantial progress in cryptocurrency market research and data quality assessment methodologies, significant gaps persist. First, existing data quality research predominantly addresses traditional financial institutions or general database systems [9], with limited adaptation to the unique characteristics of cryptocurrency exchanges—continuous global operation, API-based access, heterogeneous technical architectures, and absence of centralized oversight. Second, while cryptocurrency market research addresses price discovery [3, 6], arbitrage [6], and risk characteristics [14], systematic evaluation of exchange infrastructure quality differences and their operational impacts remains underdeveloped.

Third, methodological frameworks for continuous real-time assessment of API performance remain limited. While Wang and Strong's (1996) framework provides conceptual foundations [9], operationalizing these concepts for high-frequency streaming data requires novel measurement approaches. Our study addresses these gaps by: (1) developing cryptocurrency-specific quality metrics accounting for 24/7 operation and microsecond-level timing requirements; (2) establishing systematic evaluation procedures deployable for continuous monitoring; (3) providing empirical evidence of quality variations across platforms through 463 real-time observations; (4) quantifying the magnitude of quality disparities (7.14-point DQI gap) with data freshness as the primary differentiator; and (5) releasing open-source implementations [10] enabling reproducible research and practical applications. The framework advances both academic understanding of cryptocurrency infrastructure and practical tools for market participants and regulators.

III. RESEARCH METHOD

A. Research Design and Framework

This study employs a quantitative empirical analysis approach, collecting real-time data through standardized API interfaces and establishing a multi-dimensional evaluation framework to systematically compare data quality performance across cryptocurrency exchanges. The research methodology integrates big data analytics techniques for continuous high-frequency data acquisition, real-time quality assessment, and statistical analysis. The evaluation framework operationalizes data quality theory [9] into measurable metrics specifically designed for cryptocurrency exchange infrastructure, addressing the unique challenges of 24/7 global operation and microsecond-level timing requirements. Figure 1 illustrates the comprehensive data quality assessment framework architecture.

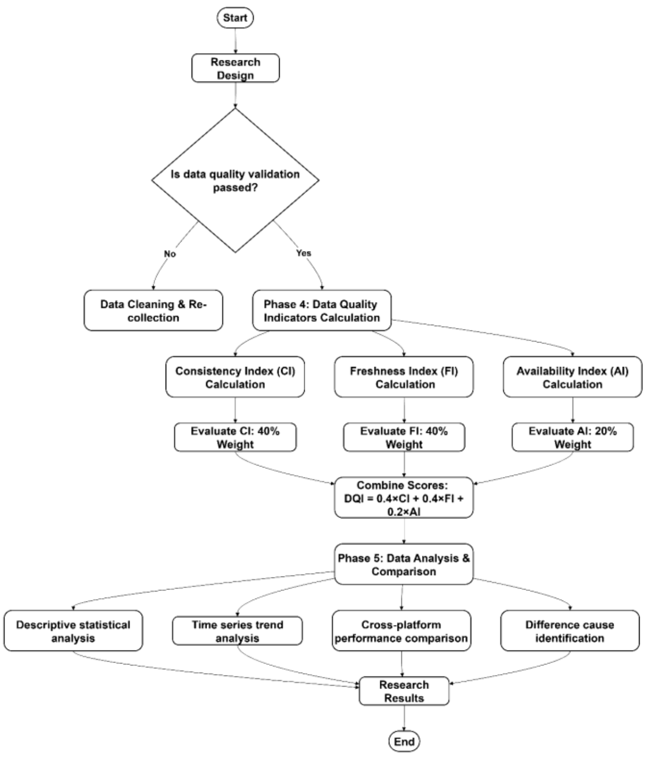

Fig. 1. Data Quality Assessment Framework for Cryptocurrency Exchanges

The framework comprises four core components: (1) Data Acquisition Layer implementing concurrent API requests with precise timestamp synchronization; (2) Quality Metrics Computation Layer calculating consistency, freshness, and availability indicators; (3) Analysis Layer aggregating dimensional metrics into composite DQI scores; and (4) Evaluation Layer enabling comparative assessment and trend analysis. This modular architecture supports extensibility for additional exchanges, trading pairs, or quality dimensions.

B. Selection of Research Objects

This study examines three representative cryptocurrency exchanges selected to capture variation in market position, technical infrastructure, and operational characteristics. Binance represents the market leader segment, functioning as the world's largest exchange by trading volume with extensive global operations and sophisticated technical infrastructure. MEXC represents medium-sized regional platforms with significant Asian market presence, offering a balance between scale and agility. Bitget represents emerging exchanges embodying new-generation platform characteristics, including rapid feature deployment and competitive positioning strategies.

C. Data Collection System Architecture

Data collection spanned eight continuous hours from 11:14 to 19:14 on September 6, 2025, with systematic sampling at 60-second intervals, yielding 463 valid observations from a theoretical maximum of 480 samples (96.46% capture rate). The temporal window encompasses both Asian trading session (226 samples) and European trading session (237 samples), ensuring balanced representation across different market conditions and geographic activity patterns.

The data acquisition system implements a distributed multi-threaded architecture using Python 3.8+ to ensure simultaneous API requests across all three exchanges, minimizing inter-platform timing discrepancies critical for comparative analysis. Core system components include:

Request Handling Module: Implements RESTful API clients for each exchange with automatic retry logic, timeout handling, and error recovery mechanisms. API endpoints include Binance's /api/v3/ticker/price, MEXC's /api/v3/ticker/price, and Bitget's /api/spot/v1/market/ticker. The module employs the requests library for HTTP communication with connection pooling to optimize network efficiency.

Timestamp Synchronization Module: Records dual timestamp information for each API call—T_request representing local system time when the request is initiated, and T_exchange representing the exchange-provided timestamp embedded in the response. This dual-timestamp design enables precise latency calculation for freshness evaluation, accounting for both network transmission delays and exchange processing time. System clock synchronization using NTP ensures timestamp accuracy within milliseconds.

Concurrent Processing Module: Utilizes Python's threading module to execute simultaneous API requests to all three exchanges. The threading architecture minimizes temporal skew between platforms, typically achieving sub-10-millisecond request initiation differences. This near-simultaneous querying is essential for valid price consistency comparisons across platforms.

Data Validation Module: Implements real-time validation checks including price reasonableness constraints (detecting outliers beyond 5% deviation from moving average), timestamp consistency verification (ensuring monotonic progression and reasonable exchange timestamps), and completeness validation (confirming all required fields are populated). Invalid samples are flagged and excluded from analysis.

Storage Module: Persists collected data in dual formats—JSON for raw data preservation maintaining full response structure, and CSV for streamlined analytical processing. This dual-format approach supports both detailed forensic analysis and efficient statistical computation.

The complete system implementation, including data collection scripts, analysis algorithms, and visualization tools, is publicly available in an open-source repository [10]. This transparency ensures research reproducibility and enables community validation and extension of the methodology.

D. Evaluation Metrics Design

The evaluation framework operationalizes three core data quality dimensions into quantitative metrics: consistency, freshness, and availability. Each metric is designed to capture critical aspects of exchange data service quality relevant to trading decisions and market efficiency.

Consistency Index: The Consistency Index quantifies price synchronization across exchanges at identical time points, calculated as:

$$CI_i = \max\left(0,100 - \frac{|P_i - \bar{P}_{-i}|}{\bar{P}_{-i}} \times 50000\right) \quad (1)$$

Where P_i is the price of exchange i, and $\bar{P}_{-i}$ is the average price of other exchanges. The consistency index reflects the fairness of price information transmission and market efficiency. The coefficient of 50,000 is designed to capture micro-level price differences of 0.001%, ensuring sensitivity to changes in market microstructure. When prices across different exchanges are completely consistent at the same time point, the indicator scores 100 points; the greater the price difference, the lower the score. This indicator is significant for evaluating arbitrage opportunities and the degree of information asymmetry. Platforms with higher price consistency can provide users with a fairer trading environment and reduce trading disadvantages caused by price differences.

Freshness Index: The Freshness Index evaluates data timeliness by measuring the delay between request initiation and data timestamp:

$$\Delta t = |T_{request} - T_{exchange}| \quad (2)$$

The scoring mechanism implements a segmented penalty function reflecting differential impacts of various delay levels on trading outcomes. The function awards full points (FI = 100) for delays ≤ 0.2 seconds, with progressively increasing penalties for longer delays: 0.167 points per 0.01-second increment for delays of 0.2-0.5 seconds; 0.3 points per 0.01-second increment for delays of 0.5-2.0 seconds; 0.1 points per 0.01-second increment for delays of 2.0-5.0 seconds; and 0.08 points per 0.01-second increment for delays exceeding 5.0 seconds. This segmented design reflects empirical evidence of latency impacts on high-frequency trading performance [7], with the 0.2-second threshold derived from documented trading outcome sensitivity to microsecond-level delays.

Availability Index: The Availability Index measures API service reliability:

$$AI = \frac{N_{success}}{N_{total}} \times 100 \quad (3)$$

where $N_{success}$ represents successful API responses and N_{total} denotes total request attempts. This metric captures infrastructure stability and service continuity, essential prerequisites for continuous market participation. Perfect availability (AI = 100) indicates zero service interruptions, while lower values reflect request failures that prevent timely information access.

Composite Data Quality Index: The Composite Data Quality Index aggregates dimensional metrics through weighted averaging:

$$DQI = 0.4 \times CI + 0.4 \times FI + 0.2 \times AI \quad (4)$$

Weight allocation reflects market microstructure principles [6, 8], prioritizing consistency and freshness (40% each) as primary determinants of trading fairness in high-frequency environments, while assigning availability (20%) lower weight given minimal variation among established exchanges. The composite index provides a holistic quality assessment ranging from 0-100, facilitating platform comparison and establishing quantitative benchmarks for regulatory standards or operational targets.

IV. RESULTS

A. Data Collection Performance and Descriptive Statistics

The data collection system achieved robust performance, successfully capturing 463 valid samples over the eight-hour observation window, representing 96.46% of the theoretical maximum (480 samples). The 17-sample shortfall resulted from temporary network fluctuations rather than systematic failures. All three exchanges demonstrated perfect API availability with 100% successful response rates, indicating stable baseline infrastructure performance during the observation period.

Bitcoin price dynamics during the observation window exhibited moderate volatility with mean price of $110,742.36 and standard deviation of $108.23. Price movements ranged from $110,580 to $111,100, representing intraday volatility of approximately $520 (0.47%). The observation period encompassed both Asian trading session (226 samples, 11:14-16:00) and European trading session (237 samples, 16:00-19:14), providing balanced temporal distribution for cross-session analysis. This sampling strategy controls for potential time-of-day effects on data quality metrics while capturing different market liquidity regimes.

B. Composite Data Quality Index Results

Table 1 presents comprehensive DQI assessment results across all evaluation dimensions. The analysis reveals clear performance stratification among the three exchanges, with Binance establishing market leadership (DQI = 92.50), followed by Bitget (DQI = 88.61) and MEXC (DQI = 85.36). The 7.14-point spread between highest and lowest performers indicates substantial quality variation with practical implications for trading efficiency and information access equity.

Dimensional analysis reveals that all platforms achieve perfect availability (AI = 100), demonstrating mature baseline infrastructure stability. However, significant heterogeneity emerges in consistency (CI range: 93.58-97.72) and particularly freshness (FI range: 69.06-83.53), with the latter driving overall DQI differentiation. Average latency measurements (Binance: 0.878s, Bitget: 1.066s, MEXC: 1.362s) quantify the technical infrastructure gaps underlying freshness performance differences.

TABLE I. DATA QUALITY INDEX ASSESSMENT RESULTS FOR THREE CRYPTOCURRENCY EXCHANGES

	Composite DQI	Consistency Index	Freshness Index	Availability Index	Average Delay (sec)
Binance	92.5	97.72	83.53	100	0.878
Bitget	88.61	93.58	77.95	100	1.066
MEXC	85.36	94.36	69.06	100	1.362

C. Price Consistency Analysis

Figure 2 illustrates Bitcoin price trajectories across the three exchanges throughout the observation period. Visual inspection reveals remarkable price synchronization, with all platforms tracking nearly identical price movements across both trading sessions. The price trajectory exhibits an initial decline from approximately $111,300 at 12:00 to $110,700 by 14:00, followed by gradual recovery to $110,800 during European session opening (17:00), before stabilizing near $110,750 toward period end.

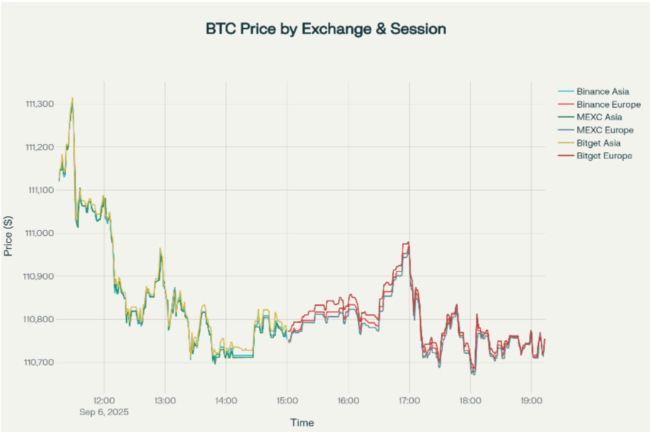

Fig. 2. BTC Price Movements by Exchange

Quantitative consistency analysis confirms visual impressions, with maximum inter-platform price deviations rarely exceeding 0.002% (approximately $2.21 on a $110,742 base). All three exchanges achieve high consistency scores (CI: 93.58-97.72), reflecting mature price discovery mechanisms in the Bitcoin market [3]. Binance's superior consistency (CI = 97.72) aligns with its market leadership position, as larger platforms often serve as price reference points for smaller exchanges [3, 6]. Interestingly, MEXC (CI = 94.36) slightly outperforms Bitget (CI = 93.58) in consistency despite lower overall DQI ranking, suggesting that price synchronization capabilities can be maintained even with infrastructure constraints affecting other quality dimensions.

The minimal price discrepancies observed reflect efficient arbitrage mechanisms [6] that rapidly equilibrate prices across platforms. These tight spreads indicate that consistency differences primarily stem from micro-level timing variations rather than systematic price deviations, supporting the framework's sensitivity calibration (50,000 coefficient in Equation 1).

D. Data Freshness Performance Differentiation

Data freshness emerges as the primary discriminator of service quality, with exchange performance spanning a 14.47-point range—significantly larger than consistency (4.14-point range) or availability (zero range). Figure 3 visualizes freshness scores, clearly illustrating the performance hierarchy: Binance (FI = 83.53, delay = 0.878s), Bitget (FI = 77.95, delay = 1.066s), and MEXC (FI = 69.06, delay = 1.362s).

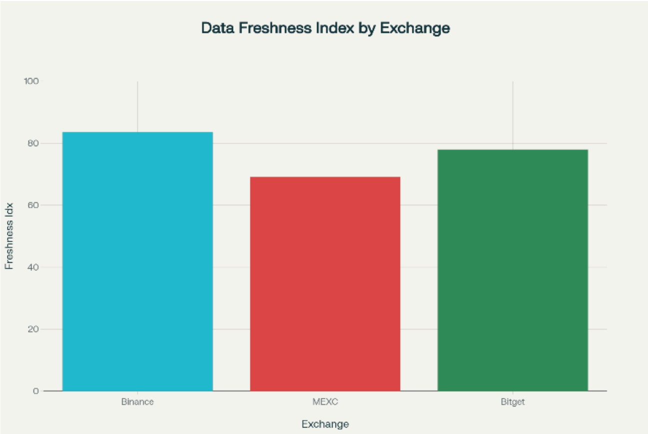

Fig. 3. Data Freshness Index by Exchange

Delay distribution analysis reveals distinct infrastructure characteristics. Binance exhibits concentrated delay patterns with 90% of samples within 1.2 seconds, reflecting optimized API architecture and geographically distributed servers that minimize latency. Bitget shows more uniform delay distribution (0.5-1.5 seconds range), suggesting adequate but less optimized infrastructure. MEXC demonstrates greater delay variability with some samples exceeding 2 seconds, indicating potential infrastructure bottlenecks or less aggressive performance optimization.

The 0.484-second latency difference between Binance and MEXC represents substantial competitive disparity in high-frequency trading contexts. Empirical evidence [7] documents that latency advantages of this magnitude translate to 15-20 basis points profit margin differences per trade in arbitrage strategies. For automated trading systems executing hundreds or thousands of daily transactions, cumulative impacts become economically significant. These findings demonstrate that data freshness directly affects market participation equity, with slower platforms systematically disadvantaging their users in time-sensitive strategies.

E. Temporal Stability and Cross-Session Analysis

Time series analysis reveals stable DQI performance throughout the observation period without significant cyclical patterns or sudden discontinuities. Binance maintains DQI scores within 92.1-92.8 range (standard deviation = 0.23), Bitget within 88.2-89.1 range (SD = 0.28), and MEXC within 85.0-85.8 range (SD = 0.24). These low standard deviations indicate consistent service quality independent of short-term demand fluctuations or network conditions.

Cross-session comparison between Asian and European trading periods reveals minimal performance variation. DQI differences between sessions measure 0.6 points for Binance, 0.6 points for Bitget, and 0.7 points for MEXC—all within measurement error margins. This cross-session consistency suggests that exchanges maintain stable infrastructure performance across varying trading volumes and geographic activity distributions, a positive indicator of operational maturity.

Importantly, relative platform rankings remain stable throughout the entire observation window without rank reversals. This ranking persistence reflects fundamental infrastructure capability differences rather than transient operational issues, validating the framework's ability to capture durable quality characteristics. The temporal stability also confirms measurement reliability and supports the framework's utility for ongoing monitoring applications.

V. DISCUSSION AND CONCLUSIONS

A. *Principal Findings and Technical Implications*

This research establishes a comprehensive, reproducible framework for quantitative assessment of cryptocurrency exchange data quality using big data analytics techniques. Through systematic evaluation of 463 real-time observations across three major platforms, the study documents significant infrastructure quality disparities with direct implications for market efficiency and trading fairness. The 7.14-point DQI gap between Binance (92.50) and MEXC (85.36) quantifies meaningful service quality differences, with data freshness (14.47-point range) emerging as the primary differentiating dimension.

The empirical findings reveal a clear correlation between market position and infrastructure investment. Binance's superior performance (0.878-second average latency, 83.53 freshness score) reflects substantial technical infrastructure investment including geographically distributed servers, optimized API architectures, and high-bandwidth network connectivity. Conversely, MEXC's comparatively lower performance (1.362-second latency, 69.06 freshness score) suggests resource constraints or different strategic prioritization. The 0.484-second latency difference represents substantial competitive disparity in high-frequency trading environments, potentially affecting arbitrage profitability by 15-20 basis points per trade [7].

These infrastructure quality variations create systematic information asymmetries [8] that advantage participants using higher-quality platforms. For retail investors, such disparities may be imperceptible in low-frequency trading but become consequential in time-sensitive strategies. For institutional traders and market makers operating at scale, quality differences directly impact execution efficiency and profitability. The framework's quantitative approach enables objective platform comparison, supporting informed decision-making by all market participants.

B. *Practical Applications and Stakeholder Implications*

The evaluation framework supports multiple practical applications across stakeholder groups:

Exchange Operators: The multi-dimensional assessment identifies specific infrastructure improvement priorities. Operators can use DQI metrics for continuous quality monitoring, competitive benchmarking, and validation of infrastructure investments. The framework's modular design enables integration into operational dashboards for real-time quality tracking.

Regulatory Agencies: The standardized methodology provides objective tools for establishing minimum service quality standards and monitoring regulatory compliance. Regulators can mandate periodic DQI reporting, set quality thresholds for licensing, or implement tiered regulatory frameworks based on demonstrated quality levels. The quantitative approach supports evidence-based policymaking and transparent enforcement.

Market Participants: Traders benefit from transparent platform quality comparisons informing selection decisions. High-frequency traders can optimize platform selection based on latency requirements, while retail investors gain awareness of service quality differences. The framework reduces information asymmetry between sophisticated and unsophisticated market participants.

Technology Developers: The open-source implementation [10] enables technology providers to integrate quality assessment into trading infrastructure, API aggregation services, or market monitoring tools. The methodology serves as a reference architecture for building quality-aware trading systems.

C. *Methodological Contributions and Reproducibility*

This research advances cryptocurrency infrastructure evaluation through several methodological innovations. First, the framework operationalizes theoretical data quality concepts [9] into concrete metrics specifically calibrated for cryptocurrency market characteristics—continuous operation, API-based access, and microsecond-level timing sensitivity. Second, the dual-timestamp approach enables precise latency measurement independent of network configuration. Third, the composite index aggregates multiple quality dimensions into a single interpretable metric while preserving dimensional granularity for detailed analysis.

Critically, the complete methodology—data collection system, analysis algorithms, and evaluation framework—is publicly available as open-source software [10]. This transparency ensures reproducibility, enables community validation, and facilitates extensions by other researchers. The released dataset comprising 463 timestamped observations provides empirical foundation for secondary analyses and methodology refinements. This commitment to open science supports cumulative knowledge development and practical application of research findings.

D. *Limitations and Future Research Directions*

Several limitations bound the generalizability of findings. First, the eight-hour observation window, while capturing cross-session variation, represents a limited temporal sample. Long-term monitoring would reveal whether quality patterns persist across days, weeks, or months, and whether performance varies with market conditions (e.g., high volatility periods, major news events). Second, the three-exchange sample, though representative of different market segments, constitutes a small fraction of global cryptocurrency platforms. Extension to additional exchanges, including decentralized exchanges (DEXs) and smaller regional platforms, would enhance generalizability. Third, the single trading pair (BTC/USDT) may not represent quality patterns for altcoins with lower liquidity or different market dynamics.

Fourth, this study employs REST API polling at 60-second intervals rather than WebSocket streaming connections. The measured latency ($\Delta t = |T_{request} - T_{exchange}|$) conflates client network conditions, routing characteristics, and exchange processing time, preventing full isolation of exchange infrastructure quality from network transmission factors. Future research should collect parallel REST and WebSocket measurements from multiple geographic locations to decompose these effects. Fifth, key parameters warrant validation: the consistency coefficient (50,000) in Equation 1 aligns with HFT precision thresholds [6, 7] but lacks systematic sensitivity analysis, while the freshness threshold (0.2 seconds) derives from documented trading impacts [7] yet was not empirically validated against actual execution outcomes. Perfect 100% availability across platforms also eliminates discriminative power; finer metrics (rate-limiting events, P95/P99 tail latency) would better distinguish infrastructure quality.

Future research should pursue several extensions. First, develop real-time monitoring systems implementing continuous quality assessment for operational deployment by exchanges or monitoring services. Second, investigate quality variations across different market conditions—volatility regimes, liquidity cycles, and crisis periods—to understand conditional performance. Third, extend the framework to decentralized exchanges and automated market makers (AMMs) where quality assessment presents unique challenges due to blockchain-based settlement. Fourth, analyze the relationship between data quality and trading outcomes through causal inference designs linking quality metrics to execution performance. Fifth, explore machine learning approaches for anomaly detection in quality metrics, enabling predictive maintenance or early warning systems. Despite these limitations, the study provides a transparent and reproducible framework [10] with actionable tools for stakeholders.

E. Broader Context and Sustainable Infrastructure Development

Beyond technical contributions, this work supports broader objectives of building sustainable and equitable digital financial infrastructure. The quantitative framework enables objective assessment of infrastructure development, supporting UN Sustainable Development Goals 9.1 (reliable infrastructure) and 10.5 (financial market regulation) [15]. By providing transparent evaluation methods, the research promotes accountability and continuous improvement in cryptocurrency exchange operations.

As cryptocurrency markets mature and integrate with traditional financial systems, ensuring equitable access to high-quality data becomes increasingly critical for market integrity. The correlation between technical investment and service quality underscores that infrastructure development requires sustained commitment and resources. Regulatory frameworks should incentivize quality improvement while ensuring that infrastructure disparities do not create systematic disadvantages for certain market participants.

F. Concluding Remarks

This research establishes that standardized, quantitative data quality assessment is both feasible and essential for cryptocurrency exchange evaluation. The multi-dimensional framework provides actionable tools for operators, regulators, and market participants while advancing academic understanding of digital financial infrastructure. The documented quality disparities—particularly the 14.47-point freshness gap—demonstrate that even among established exchanges, meaningful infrastructure differences persist with practical consequences for trading efficiency and market fairness.

The path forward requires continued collaboration among researchers, industry practitioners, and regulatory bodies. Methodological refinements, extended empirical validation, and practical implementation through monitoring systems will strengthen the framework's utility. Open-source release of implementations and datasets [10] facilitates this collaborative development, ensuring that research findings translate into practical improvements in cryptocurrency market infrastructure. By combining technical rigor with practical applicability, this work contributes to building more transparent, efficient, and equitable digital financial ecosystems.

REFERENCES

[1] BlockBeats. (2025, September 10). "Crypto market fully recovers, total market cap rebounds to exceed $4 trillion" (加密市場全面復甦， 總市值回升突破 4 兆美元), cnYES. [Online]. Available: https://m.cnyes.com/news/id/6150456

[2] S. Corbet, B. Lucey, A. Urquhart, and L. Yarovaya, "Cryptocurrencies as a financial asset: A systematic analysis," Int. Rev. Financial Analysis, vol. 62, pp. 182-199, 2019.

[3] M. Brandvold, P. Molnár, K. Vagstad, and O. C. A. Valstad, "Price discovery on Bitcoin exchanges," J. Int. Financial Markets, Institutions and Money, vol. 36, pp. 18-35, 2015.

[4] P. K. Ozili, "Impact of digital finance on financial inclusion and stability," Borsa Istanbul Rev., vol. 18, no. 4, pp. 329-340, 2018.

[5] G. Pieters and S. Vivanco, "Financial regulations and price inconsistencies across Bitcoin markets," Information Economics and Policy, vol. 39, pp. 1-14, 2017.

[6] I. Makarov and A. Schoar, "Trading and arbitrage in cryptocurrency markets," J. Financial Economics, vol. 135, no. 2, pp. 293-319, 2020.

[7] A. Alexander, "Latency arbitrage in cryptocurrency markets: Analyzing execution speeds & liquidity dynamics," SSRN Working Paper No. 5143158, 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5143158

[8] A. S. Kyle, "Continuous auctions and insider trading," Econometrica, vol. 53, no. 6, pp. 1315-1335, 1985.

[9] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," J. Management Information Systems, vol. 12, no. 4, pp. 5-33, 1996.

[10] J.-K. Tsao, "Binance and MEXC and Bitget price data analytic," GitHub, 2025. [Online]. Available: https://github.com/kuan35/Binance_and_MEXC_and_Bitget_price_data_analytic. [Accessed: Sep. 24, 2025].

[11] G. Schwenkler, A. Shah, and D. Yang, "Aggregate confusion in crypto market data," SSRN Working Paper No. 4673905, 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4673905

[12] D. G. Baur, K. Hong, and A. D. Lee, "Bitcoin: Medium of exchange or speculative assets?" J. Int. Financial Markets, Institutions and Money, vol. 54, pp. 177-189, 2018.

[13] C. M. Hafner, "Testing for bubbles in cryptocurrencies with time-varying volatility," J. Financial Econometrics, vol. 18, no. 2, pp. 233-249, 2020.

[14] Y. Liu and A. Tsyvinski, "Risks and returns of cryptocurrency," Rev. Financial Studies, vol. 34, no. 6, pp. 2689-2727, 2021.

[15] United Nations General Assembly, "Transforming our world: the 2030 Agenda for Sustainable Development," 2015.