Metrics for Quality Assessment in Blockchain-based Systems: A Systematic Mapping Study

Raphael Saraiva
State University of Ceará
Fortaleza, Ceará, Brasil
raphael.saraiva@aluno.uece.br

Allysson Allex Araújo
Federal University of Cariri
Juazeiro do Norte, Ceará, Brazil
allysson.araujo@ufca.edu.br

Pamella Soares
State University of Ceará
Fortaleza, CE, Brazil
pamella.soares@aluno.uece.br

João Carlos Pontes
State University of Ceará
Fortaleza, CE, Brazil
joao.carlos@aluno.uece.br

Jerffeson Souza
State University of Ceará
Fortaleza, CE, Brazil
jerffeson.souza@uece.br

Abstract

Context: Blockchain technology has been widely adopted across various industries, such as finance, healthcare, and logistics, due to its ability to ensure transaction integrity and transparency. **Problem:** The growing adoption of blockchain systems demands rigorous quality assessment to ensure compliance with requirements. While software metrics for centralized systems are well-studied, a clear research gap remains in understanding their systematic adoption for evaluating blockchain-based systems. **Solution:** This study identifies, analyzes, and categorizes software metrics used in blockchain-based systems assessment grounded on the quality characteristics of the ISO/IEC 25010 standard. **IS Theory:** The study uses the ISO/IEC 25010 as a theoretical foundation for classifying and analyzing quality metrics. **Method:** A Systematic Mapping Study was conducted, identifying 128 primary studies extracted from different academic databases. **Results:** The findings highlight performance and security as fundamental to quality in blockchain systems, focusing on throughput and latency metrics. Additionally, Hyperledger Caliper is a frequent tool for evaluating blockchain network performance. Smart contracts, on the other hand, emerge as a promising area, given the lack of studies exploring their security and efficiency. **Contributions:** This work enhances the understanding of quality assessment in blockchain systems by mapping software metrics to the ISO/IEC 25010, covering predominant metrics, quality characteristics, and evaluation methods.

CCS Concepts

• **General and reference → Metrics**; • **Information systems → Data management systems**.

Keywords

Blockchain, Quality Metrics, ISO/IEC 25010, Systematic Mapping Study

1 Introduction

In recent years, blockchain technology has emerged as a promising innovation, potentially contributing to overcoming different challenges, drawing comparisons to the initial impact of the internet and web in their early days [30]. According to Tapscott and Tapscott [36], the ability of blockchain to ensure the integrity and transparency of transactions through a decentralized ledger has attracted attention from academia and industry. Given this high potential, the need to evaluate and ensure the quality of these solutions has proved imperative.

In general, software quality is intrinsically related to the software's ability to be maintained, updated, and adapted to new requirements and environments, which is fundamental for continuous evolution [44]. However, quality assessment in blockchain systems presents multiple challenges due to their distributed architecture and the complexity of cryptographic operations involved [42]. Such systems interact with the distributed environment, introducing constraints and demands on developers, emphasizing the need for quality assessment approaches and tools [27].

In the Information Systems (IS) context, quality assessment ensures that systems effectively support business processes, data management, and information flow, promoting strategic, and operational alignment [26]. The ISO/IEC 25010 standard is a widely recognized quality model that specifies characteristics for assessing and improving complex systems [13, 16]. We chose this standard due to its well-established structure and international recognition in defining software quality. According to Soundararajan and Shenbagaraman [35], ISO/IEC 25010 has been utilized in various domains, including emerging technologies such as blockchain. Their study highlights how this standard provides a structured approach for assessing quality attributes like security, reliability, transparency, and governance in technological systems.

In blockchain systems, where decentralized environments introduce specific challenges, the proper definition and application of metrics have demonstrated critical. Thus, it becomes necessary to explore and understand metrics that evaluate not only traditional aspects but also elements such as complexity, communication capability, resource consumption (e.g., the gas in the Ethereum platform), and overall blockchain system performance [27, 30].

Despite the substantial body of literature on software metrics in centralized systems [17, 24], a research gap persists regarding the systematic understanding of metrics adoption for quality assessment in blockchain-based systems. This gap is particularly relevant in IS because blockchain systems fundamentally challenge traditional quality assessment approaches due to their decentralized architecture, immutable data structures, and reliance on cryptographic mechanisms. This multifaceted perspective aligns with the challenges of systems-of-systems and complexity in IS discussed at the I Grand Research Challenges in Information Systems in Brazil

2016-2026 (GranDSI-BR) [6], which emphasizes the need for robust methodologies to address interoperability, scalability, and security in integrated and distributed information systems.

Given the previously discussed motivation, this study aims to identify, analyze, and categorize software metrics used for quality assessment in blockchain-based systems, focusing on quality characteristics aligned with the ISO/IEC 25010 standard. In order to achieve this objective, we conducted a Systematic Mapping Study (SMS), considering the need to identify knowledge gaps and research trends on the studied topic [19, 28]. Our systematic analysis identified key findings including predominant metrics used in blockchain systems, most frequently evaluated quality characteristics, commonly employed evaluation methods and tools, including research gaps and technical challenges.

Studies have explored blockchain metrics but often focus on isolated aspects without a comprehensive view based on ISO/IEC 25010 [2, 9]. For example, Yli-Huumo et al. [41] and Zhou et al. [43] examined technical challenges and consensus mechanisms, respectively, yet neither structured a systematic classification of quality metrics. Similarly, Qammar et al. [32] addressed security in blockchain-based federated learning without a standardized framework. Unlike these works, our study integrates blockchain quality metrics into a structured model aligned with ISO/IEC 25010, providing a broader vision.

Therefore, this research contributes to academia and industry in the IS context, encompassing aspects related to people, technology, and processes. From a people perspective, it presents researchers and practitioners with a mapping of quality metrics adoption in blockchain systems. In terms of technology, it details the most used software metrics used in blockchain systems, highlighting limitations and opportunities for new research. Regarding processes, it emphasizes the importance of continuous evaluation and improvement of blockchain systems. Finally, for industry, this study provides evidence about which metrics have been most used in blockchain systems, including the characteristics of such metrics and how they have been approached.

This paper is structured as follows: section 2 presents the theoretical background, while section 3 details the method explored for conducting the SMS. Section 4 presents the achieved results. Section 4 discusses the research's findings and implications. Section 6 analyzes the threats to the study's validity; and finally, section 7 brings the conclusions.

2 Background

This section presents the theoretical foundations that underpin this research. Section 2.1 explores blockchain's fundamental concepts, while section 2.2 addresses quality metrics applicable to these systems, focusing on characteristics defined by the ISO/IEC 25010 standard.

2.1 Introducing the Blockchain Technology

Blockchain has emerged as a technology with significant impacts on improving processes and data security in various domains, such as healthcare and finance [10, 18]. Its decentralized and secure architecture is based on a distributed ledger managed by peer-to-peer (P2P) networks, which enables append-only operations that prevent modifications or deletions of stored data [40]. This architecture fundamentally eliminates the need for a centralized authority by relying on decentralized consensus mechanisms, such as Proof of Work or Proof of Stake, to validate transactions and maintain the system's integrity. Security and transparency are achieved through consistently replicating the ledger across all nodes, where each node independently maintains a synchronized copy of all transactions and states [8].

The decentralized nature of blockchain enhances the system's resilience against single points and promotes trust among participants by ensuring that no single entity can unilaterally alter the data. Furthermore, blockchain's immutable and transparent characteristics provide robust safeguards for data integrity and auditability, making it particularly suited for applications requiring high levels of security and accountability. This distributed model also facilitates interoperability across diverse information systems in areas such as finance, supply chain management, and healthcare.

In the blockchain, each block is represented by a hash (a fixed-length bit sequence that maps its content to ensure data integrity). Each block's hash includes the previous block's hash, forming an interlinked chain that traces each block back to the genesis block, the network's first block (see Figure 1). Thus, this structure enables the identification of any modification attempts, reinforcing blockchain's security and integrity [4].

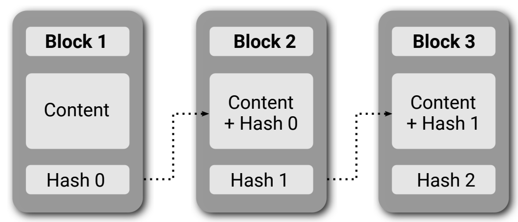

Figure 1: Representation of blocks interlinked by previous block hash references.

With the emergence of second-generation blockchain technology, driven by the Ethereum platform [7], the adoption of smart contracts has also expanded considerably. According to Xu et al. [40], smart contracts are programs deployed and executed directly on the blockchain during transactions. These contracts can store and transfer digital assets managed by the blockchain and interact with other smart contracts stored in the network. Smart contracts also paved the way for developing decentralized applications known as dApps. Built on blockchain platforms, dApps use smart contracts to operate autonomously and without intermediaries, ensuring transparency and security in their operations [33, 38].

2.2 Software Metrics and Blockchain

Using metrics is important to understand, control, and improve software quality throughout its lifecycle [12]. Thus, metrics can contribute to continuously improving information systems [1]. To ensure consistency and comparability in assessments, recent software quality models have undergone standardization, establishing a standard set of metrics aligned with accepted criteria [23]. In this context, the ISO/IEC 25010 standard provides a comprehensive framework for quality assessment, defining characteristics such as performance, security, functionality, compatibility, reliability, usability, maintainability, and portability [13, 16]. This standard is widely adopted in industry to ensure that software products meet rigorous quality criteria, with each characteristic further subdivided into sub-characteristics (e.g., resource utilization and Time Behavior for performance) [25].

In the context of blockchain, **performance** metrics such as latency and throughput are particularly useful in measuring the system's efficiency. Latency reflects the time taken for transaction confirmations, while throughput measures the number of transactions processed per second, together providing a picture of the network's capacity and responsiveness [11, 42]. Regarding **security**, blockchain requires metrics that measure safeguards against specific vulnerabilities, including 51% attacks, cryptographic key protection, and secure data storage within smart contracts. These security metrics are critical to addressing blockchain systems' specific threats in decentralized environments [31].

The characteristic of **functionality** relates to the system's ability to maintain data accuracy and integrity and fulfill transaction completeness requirements, ensuring, for example, the immutability and transparency of information within a blockchain context [39]. Assessing **compatibility** may involve blockchains interacting with other networks, ensuring consistent and secure information exchange. According to Belchior et al. [5], compatibility is a foundational characteristic for fostering collaboration across blockchain platforms, enabling diverse applications and secure cross-chain transactions.

Regarding **reliability**, the distributed nature of blockchain systems requires each network node to maintain an accurate, redundant transaction history, preserving the system's functionality and data integrity even amidst failures or temporary node disconnections as Lin and Liao [21] highlights, reliability in blockchain systems is supported by the redundancy of data and resilience of the nodes within the network. **Usability** also becomes increasingly important as blockchain technologies expand to end-user applications; intuitive interfaces and user-friendly experiences are relevant to facilitate the broader adoption of blockchain solutions [15].

Maintainability in blockchain systems is a complex issue, as it requires that updates, corrections, and adaptations be made without compromising the stability of the decentralized network. This quality characteristic is relevant for supporting blockchain systems as they evolve to meet new demands and integrate with emerging technologies [14]. Finally, **portability** metrics are important to ensure that blockchain solutions operate consistently across various environments and platforms, offering flexibility and broad applicability.

3 Research Method

We employed a Systematic Mapping Study (SMS) as the research method for this work, considering its alignment with our objective. Through the SMS, we identified predominant metrics in blockchain systems, frequently evaluated quality attributes, commonly used evaluation methods, tools, research gaps, and technical challenges. This approach allowed for the systematic identification, categorization, and analysis of these elements, offering a comprehensive understanding of practices and trends in the field. The subsequent sections detail the research questions that guided this study and the protocol steps followed to implement the SMS [19, 28].

3.1 Research Questions

To achieve the defined objective for this study, we formulated the following Research Questions (RQs):

- **RQ1) What are the most commonly used metrics for measuring software quality in blockchain systems?**
  Rationale: To identify the most used metrics in assessing blockchain systems to reveal trends and recurring practices in the community.
- **RQ2) What quality characteristics are frequently assessed in blockchain systems?**
  Rationale: To understand which quality aspects according to ISO/IEC 25010 are being considered in assessing blockchain systems.
- **RQ3) How are blockchain studies' metrics applied and assessed?**
  Rationale: To explore the main methods and tools used to apply and assess metrics in blockchain studies.

3.2 Study Selection Process

Figure 2 illustrates the process for selecting primary studies. The selection process was carried out in three main stages.

3.2.1 Search Strategy. The papers were collected using a predefined search string across selected databases. We performed the systematic search in March 2024, focusing on three main themes: blockchain, software measurement, and software quality characteristics as defined in ISO/IEC 25010. We initiated our study with a pilot search, combining the primary keywords blockchain, measurement, and ISO 25010 and testing them individually. Based on the preliminary results, we identified relevant correlates and synonyms, allowing us to refine and expand our search string as shown below:

("blockchain") AND ("measurement" OR "software measurement" OR "metrics" OR "software metrics") AND ("functionality" OR "performance" OR "compatibility" OR "usability" OR "reliability" OR "security" OR "maintainability" OR "portability")

We defined the search string to encompass a wide range of quality aspects as defined by the ISO/IEC 25010 standard, which provides a detailed model for evaluating system and software quality. We chose this standard as a reference for this study due to its comprehensiveness and international recognition in software quality assessment. Furthermore, we included keywords related to software metrics in the search string.

We searched academic databases, including Scopus, IEEE Xplore, Science Direct, and ACM Digital Library, ensuring extensive and relevant coverage of available scientific literature. These databases are recommended by Keele et al. [19] for conducting secondary studies since they offer comprehensive coverage of relevant topics. Initially, we identified 1,031 papers distributed across the four databases as presented in Figure 2.

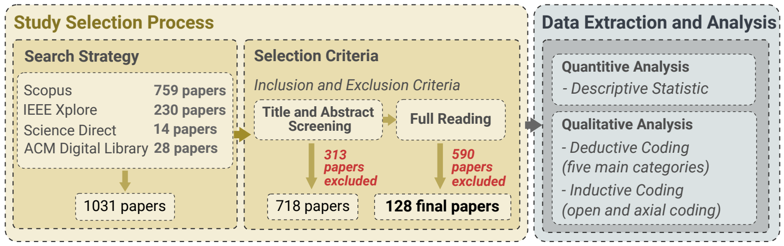

Figure 2: Steps for primary studies selection

3.2.2 Selection Criteria. We conducted this phase in two main stages, each designed to progressively refine our final set of studies. For this purpose, we guided the screening process by applying Inclusion Criteria (IC) and Exclusion Criteria (EC) initially determined to ensure alignment with our research questions while maintaining the quality and relevance of included studies. We defined the following ICs and ECs:

Inclusion Criteria (IC):

- **IC1:** Papers that apply software metrics in the evaluation of blockchain-based systems;
- **IC2:** Papers published in journals and conferences;
- **IC3:** Papers written in English.

Exclusion Criteria (EC):

- **EC1:** Papers without full text availability;
- **EC2:** Secondary or tertiary papers (systematic reviews, surveys, etc.);
- **EC3:** Papers in editorial form, etc., as they do not provide sufficient information;
- **EC4:** Papers that do not evaluate metrics in the specific context of blockchain.

The first stage of criteria application involved *Title and Abstract Screening*, where two authors independently reviewed each retrieved paper. During this phase, we removed 232 duplicate papers and 81 non-scientific papers. After this initial screening, 718 relevant papers were considered for the second stage.

In a second moment, we conducted the *Full Reading*, where the remaining papers were divided between two authors for detailed review. This thorough examination removed 590 papers that did not meet our ICs or fell under our ECs. Consequently, we reached a final list of **128 primary studies**. For both stages, when reviewers disagreed about a paper's relevance, they held meetings to discuss and resolve these differences through consensus.

3.3 Data Extraction and Analysis

In this phase, we conducted a systematic data extraction process for the 128 primary studies by two reviewers who evaluated the primary studies. To ensure consistency and reliability of the assessments, we adopted a cross-checking approach where each reviewer was responsible for evaluating half of the paper; subsequently, the other reviewer verified their assessment. This cross-checking process aimed to reach a consensus and reduce possible individual biases regarding the metrics analyzed in each study.

We employed both quantitative and qualitative approaches for the data analysis. To present an overview and characterization of the studies, we explored descriptive statistics to account for frequency over the years, type, and location of publications. To answer the defined RQs, we employed a combined deductive and inductive coding approach [22]. Initially, using deductive coding, we classified the information into predefined categories:

1. **Names of metrics used in the paper:** Identification of metrics applied in each study to measure software quality.
2. **Classification and subclassification of metrics according to ISO 25010:** Organization of metrics according to categories and subcategories defined by the ISO/IEC 25010 standard, allowing for structured and comparative analysis.
3. **Tools used in metric evaluation:** Identification of tools and/or methods employed to evaluate the metrics, such as specific software or scripts developed by the authors.
4. **Evaluation Layer Level:** Classification of metrics according to the application level, including Blockchain Network, Smart Contract, and System Integration Evaluation.
5. **Process for obtaining metrics used in the study:** Procedure adopted to collect and analyze the metrics, including experimental methods, simulations, or empirical analyses.

Regarding the classification and subclassification of metrics, we classified the codes based on the ISO/IEC 25010 main standard categories, which included performance, security, functionality, compatibility, reliability, usability, maintainability, and portability. Similarly, we identified that the evaluation level usually clustered into three main application domains: a) blockchain network evaluation (studies focusing on assessing new blockchains, protocols, and consensus mechanisms); b) smart contract evaluation (research primarily centered on smart contract analysis and metrics), and c) system integration evaluation (studies examining decentralized applications and systems that leverage blockchain technology).

Subsequently, we performed an inductive analysis through iterative "open" and "axial" coding of the content of each predefined category, allowing us to discover emerging patterns beyond our initial classification scheme. In cases of disagreement between the points mentioned above, a joint evaluation was performed by both reviewers to reach a consensus on this point, thus seeking to minimize biases and ensure accuracy. Relevant data extracted from selected studies were organized in a data extraction table containing all necessary information to answer the research questions. Additionally, other information was captured, such as ID, title, year, and publication venue, which can be found in our open repository [34]. Due to space limitations, we have included the complete list of all 128 primary studies and their respective extracted data in our repository.

4 Results

This section presents the SMS results, addressing the research questions, extracted data, and their analyses. Initially, we provide a general characterization of the selected primary studies (see Section 4.1). Subsequently, the answers to each research question are addressed based on the analysis of these studies (see Sections 4.2, 4.3, and 4.4).

4.1 Characterization of Primary Studies

Figure 3 shows the evolution of the primary studies over the years, starting from 2016. We can notice that almost all studies have been published in the last eight years, with no publications before 2016, indicating that we are covering a recent topic. In particular, 2023 stood out with the highest number of published papers, totaling 36, followed by 2021 with 34 papers. Despite a slight decrease to 29 papers in 2022, this number still exceeded the years prior to 2021. It is worth noting that for 2024, the search only considered the first quarter, resulting in just two papers.

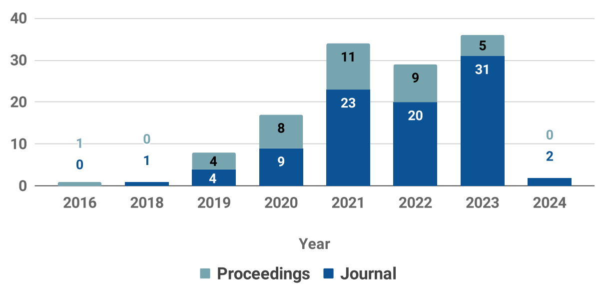

Figure 3: Evolution of primary studies over the years.

When analyzing the publication venues of the primary studies, we found that only 38 papers (29.7%) were published in conferences, while 90 articles (70.3%) were published in journals. Furthermore, the studies published in conferences were distributed across 29 different venues, with only three conferences having more than one published article: IEEE International Conference on Blockchain (4), IEEE International Conference on Communications (3), and International Conference on Distributed Computing Systems (2). Among the journals that stand out for publishing most of the articles found are the IEEE Internet of Things Journal and IEEE Access, with 7 and 5 articles, respectively, and Applied Sciences, which also published five articles. The highest concentration of publications is from IEEE Xplore, representing 32.6% of primary studies. It is followed by publications from MDPI with 21.3%, ScienceDirect with 19.1%, Springer with 11.2%, and Wiley with 4.5%. The remaining journals (11.2%) correspond to articles with a single publication in publishers such as Taylor and Francis, PeerJ, and AGSER.

4.2 RQ1) What are the most commonly used metrics for measuring software quality in blockchain systems?

Initially, to analyze the frequency and application of metrics, we used deductive open coding to group similar metrics into representative conceptual categories. For example, although presented in various forms — such as Request Latency and Transaction Latency — the Latency metric essentially represents the same response time evaluation. Similarly, Delay was mentioned under different terms, such as Delay in Block Propagation, Consensus Delay, Network Delay, Maximum Delay, and Average Delay, all consolidated under the term Delay. This grouping allowed metrics with common concepts to be analyzed, avoiding duplications and facilitating a more coherent evaluation of their presence and relevance in the studies.

As illustrated in Figure 4, when examining the metrics used in the primary studies, 160 metrics were identified across the 128 reviewed articles, of which 27 appeared in more than one study. This consolidation revealed that the Latency metric was applied in 27% of the primary studies, closely followed by the Throughput metric, present in 26.1% of the articles. These metrics, Latency and Throughput, stood out in more than half of the studies, with similar frequencies. Other metrics, such as Delay (5.7%) and Transaction Cost (4.7%), ranked third and fourth respectively. Additionally, a group labeled Others appears in the graph, representing 4.7% of the identified metrics; this group comprises a set of 10 metrics that were cited only twice in the reviewed articles.

Analyzing the most commonly used metrics in Figure 4, we can see that the results demonstrate the prioritization of Throughput and Latency metrics in blockchain systems evaluation, reflecting their relevance to system performance. For example, studies such as [S7, S101] validated Hyperledger-based architectures for health records, demonstrating that high Throughput and reduced Latency are fundamental to ensuring speed and reliability in critical environments. Similarly, [S116] highlighted the importance of these metrics in digital health solutions, where agility and efficiency are indispensable. Furthermore, [S54, S87] explored their application in communications between satellites and cognitive radio networks, while [S71, S88] emphasized the balance between these metrics as determinants for efficiency and security in access control systems.

Regarding the Delay metric, works such as [S4, S37, S79] showed that minimization increases efficiency and improves security and scalability. For example, [S4, S37] analyzed the impacts of Delay on Bitcoin's efficiency and security, showing that reducing block propagation time mitigates forks and improves network stability. Meanwhile, [S79] demonstrated how separating consensus into continuous leaders improves latency in high-demand scenarios. In turn, [S50] highlighted optimizing propagation protocols for greater efficiency and reliability.

The Transaction Cost metric's relevance was evidenced across different domains. In healthcare, [S22, S127] analyzed the impact of smart contract execution costs in medical record management, highlighting the importance of maintaining low costs to enable adoption. In the energy sector, [S23, S94] presented energy trading schemes, showing the role of transaction cost in blockchain adoption for energy monitoring and transactions. In the IoT context, [S106, S126] evaluated how reduced costs improve scalability and efficiency in distributed networks.

Execution Time and Energy Consumption both occupy the fifth position with 3.8% of the papers. The importance of execution time is demonstrated in studies such as [S11], where execution time is analyzed using Node.js to measure the performance of blockchain-based authentication systems under different scenarios. Likewise, in [S55], execution time is evaluated alongside data consumption to benchmark the efficiency of a blockchain-based certificate revocation system, emphasizing its role in ensuring high performance. Regarding energy consumption, [S30] explored strategies to optimize efficiency in blockchain systems applied to healthcare IoT. Their approach utilized smart contracts to identify and select processing nodes that minimized energy usage.

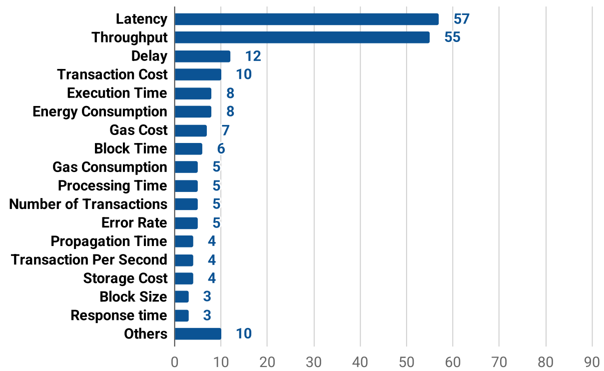

Figure 4: Distribution of metrics evaluated in blockchain studies.

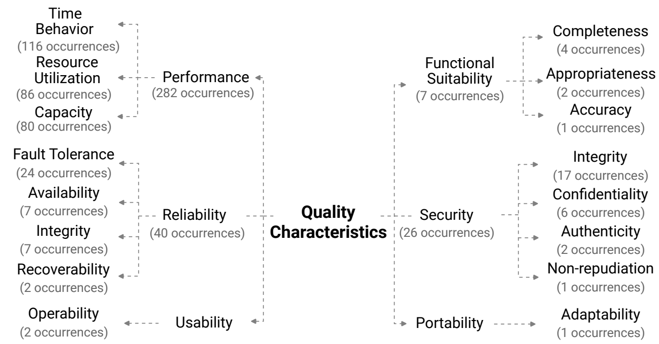

Figure 5: Quality characteristics according to ISO/IEC 25010.

4.3 RQ2) What quality characteristics are frequently assessed in blockchain systems?

Figure 5 highlights the occurrence distribution among the identified characteristics within the ISO/IEC 25010 standard according to the mapped metrics. It is important to note that the presented values do not represent the number of papers analyzed, as a single paper may contain multiple distinct occurrences encompassing different quality characteristics and subcharacteristics.

In particular, Performance, Reliability, and Security were the most frequently identified characteristics. Performance was the most evaluated characteristic, with a total of 282 occurrences, predominantly in the subcharacteristics Time Behavior (116 occurrences), Resource Utilization (86 occurrences), and Capacity (80 occurrences). Reliability followed with 40 occurrences, distributed among the subcharacteristics Fault Tolerance (24 occurrences), Availability (7 occurrences), Integrity (7 occurrences), and Recoverability (2 occurrences). For the Security characteristic, there were 26 occurrences, with Integrity being the most frequent subcharacteristic (17 occurrences), followed by Confidentiality (6 occurrences), Authenticity (2 occurrences), and Non-repudiation (1 occurrence). In contrast, characteristics such as Functional Suitability (7 occurrences), Usability (2 occurrences), and Portability (1 occurrence) had the lowest number of occurrences.

As we can see, the primary studies attributed greater importance to the Performance characteristic, focusing on efficient resource utilization and temporal behavior of blockchain systems. Examples include [S26], which investigated transaction cost and integrity in healthcare systems, and [S44], which analyzed efficiency and communication latency in wireless PBFT networks, highlighting the relevance of metrics such as throughput and latency. Additionally, the capacity subcharacteristic is addressed only in [S93], which introduced the ET-DeaL scheme, demonstrating how integrating technologies, such as IPFS, can increase scalability and the ability to handle large transaction volumes.

Reliability is a quality characteristic that encompasses Fault Tolerance, Recoverability, and Availability, ensuring blockchain systems' continuous and consistent operation even in the face of faults or attacks. In this regard, [S27] proposed the BlockCtrl architecture for Software Defined Optical Networking (SDON, which utilizes blockchain's distributed ledger and consensus mechanisms to improve Fault Tolerance. This design enables SDON switching nodes to recover to an optimal slave controller during failures quickly. Likewise, [S44] examined the Practical Byzantine Fault Tolerance (PBFT) networks over the IEEE 802.11 protocol in wireless environments, highlighting PBFT's ability to sustain reliable consensus even under challenging conditions.

Regarding Security, the subcharacteristics of Integrity and Confidentiality are important in protecting sensitive data, ensuring transaction accuracy, and maintaining continuous system access. [S57] proposed a large-volume data trading system that uses smart contracts and proxy re-encryption, ensuring confidentiality by restricting access to authorized users and protecting transaction integrity throughout the entire lifecycle. Furthermore, [S63] presented a risk classification framework in Ethereum, based on "suspiciousness" metrics, which identifies illicit activities and reinforces transaction integrity, promoting greater trust and security in the ecosystem. Moreover, [S74] introduced the MCS (Multiple Coordinators Selection) algorithm in the IOTA context. This approach distributes the consensus process among multiple coordinator nodes, eliminating single points of failure and strengthening system availability.

Finally, Functional Suitability, which includes subcharacteristics such as Functional Completeness, Appropriateness, and Correctness, appears only 7 times in the studies, reflecting its limited attention despite its importance in ensuring blockchain systems effectively meet user requirements. For example, [S57] explores this characteristic by proposing a blockchain-based data trading system, evaluating functionality to ensure that operations meet specific user needs and deliver reliable outcomes. Similarly, Usability and Portability were among the least addressed characteristics, with only 2 and 1 occurrences, respectively. [S1, S25] highlighted the importance of Usability in blockchain systems, [S1] focused on user-centric shard allocation to improve accessibility and operational efficiency, while [S25] emphasized the need for intuitive access control mechanisms to simplify user interaction. Moreover, [S2] discussed Portability challenges, particularly the difficulties in transferring assets and functionalities across blockchain platforms.

4.4 RQ3) How are blockchain studies' metrics applied and assessed?

To answer RQ3, metrics were categorized in blockchain studies across three main levels, as illustrated in Table 6: Blockchain Network, System Integration, and Smart Contract. This categorization was based on the application of metrics to evaluate different aspects of blockchain systems, reflecting the various layers at which blockchain technology can operate.

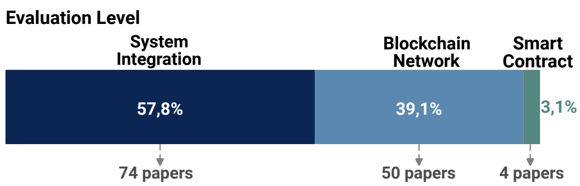

Figure 6: Metric evaluation level.

The system integration level, encompassing 74 papers, stood out as the most addressed. Metrics at this level are used to evaluate blockchain integration with other systems and technologies, providing aspects such as security, performance, and scalability. For example, [S6] presents SafeMetaNet, which integrates proximity authentication, encryption, and blockchain to protect data exchange between IoT devices and the Metaverse, ensuring integrity and confidentiality. In the healthcare sector, [S7] proposed an architecture combining blockchain and machine learning to predict fraud and manage health data, using the Hyperledger Fabric platform to create smart contracts and ensure immutable data storage. Similarly, [S40] introduced MedBlock, an AI-enabled blockchain system to maintain the privacy of COVID-19 patient health records, employing the IPFS protocol to reduce storage costs and improve data access efficiency.

At the Blockchain Network level, covering 50 papers, metrics are used to evaluate elements within the blockchain network, such as resource utilization and attack resistance. Papers such as [S1] and [S46] explored metrics related to network efficiency: [S1] evaluated load balancing in sharded blockchains, contributing to the reduction of vulnerabilities such as shard-targeted attacks. In contrast, [S46] used the M/D/1 model to measure throughput and latency in Hyperledger Fabric, optimizing parameters to improve performance. Additionally, some works like [S92] highlighted costs-related metrics, analyzing Ethereum's gas mechanism and the impact of transaction fees on blockchain network efficiency and decentralization.

The Smart Contract level, addressed in only four papers, evaluates metrics specific to smart contracts. These metrics examine the security, resilience, and efficiency of smart contracts. Papers [S2, S74, S127] focused on smart contract security and resilience, with [S2] analyzing vulnerabilities such as anonymity loss and key leakage, [S74] proposing metrics like transaction distribution to prevent single points of failure in decentralized systems, and [S127] exploring information traceability and security through smart contracts, ensuring transparency and audibility for a measurement control system. Meanwhile, [S107] addressed efficiency, applying transaction cost, block time, and implementation cost metrics to evaluate contract scalability in medical record management.

Regarding the tools used for obtaining and evaluating metrics, about 48.9% of studies reported using specific tools. Among these, 15.6% reported using Hyperledger Caliper [S7, S13, S14, S86], a widely recognized tool for measuring blockchain performance. Additionally, about 8.1% of studies developed scripts in languages such as Python [S11, S12, S23, S28, S31, S42, S47, S55], JAVA [S57], Go [S19], Node.js [S29] for metric evaluation.

Furthermore, there were two cases where the authors developed custom tools for evaluation, highlighting the importance of specialized tools in blockchain metric analysis. The first case [S24] was specifically developed to measure the bandwidth and latency of Bitcoin network nodes. This tool was deployed in seven geographically distributed locations to record data rates and block propagation times, ensuring bandwidth saturation and using tools like libpcap for download traffic analysis. In the second case [S15], UBETA (Unified Permissioned Blockchain-based P2P-ET Architecture) was proposed, a permissioned blockchain architecture based on Hyperledger Besu with the Istanbul Byzantine Fault Tolerance (IBFT) consensus algorithm. UBETA allows performance metrics such as latency, throughput of read and write transactions, and failure rate to be measured using a Java-developed tool for mass transaction generation and performance monitoring.

Hyperledger Caliper remains the most widely used tool for assessing blockchain networks and system integration, offering standardized support, comprehensive documentation, and robust capabilities for measuring latency and throughput [S7, S13, S14]. Calyper can also be used to evaluate resource consumption [S86], but requires integration with external monitoring tools, like Wireshark for network analysis and potentially Prometheus or Grafana for system-level monitoring. While Caliper excels in structured evaluations, it lacks flexibility for defining custom metrics. In contrast, Python, Java, and Go scripts [S11, S19, S29] may offer greater flexibility, enabling researchers to gather transaction and block propagation data, analyze network discrepancies, and monitor mining behavior more effectively. However, these scripts require technical expertise, high implementation effort, and lack external support. Meanwhile, there are tailored tools [S15, S24], developed specifically for blockchain evaluations, addressing specific performance aspects that standardized tools and custom scripts may not fully capture.

Finally, obtaining and analyzing metrics in blockchain-related studies demonstrates a diverse and detailed approach, combining practical and theoretical methods. Using computational simulations is predominant, with specialized tools mentioned earlier, such as Hyperledger Caliper and scripts, standing out in the metric collection. However, there were cases of applying mathematical models, such as Markov chains [S11, S214], software instrumentation [S211] and statistical analyses [S175], and the use of specific protocols, such as Wormhole implemented in Rust [S1].

5 Discussion

This study conducted an SMS to identify, analyze, and categorize software metrics used for quality assessment in blockchain-based systems, guided by three main research questions. Below, we discuss how the results obtained engage with the existing literature, highlighting answers to the established research questions.

RQ1) What are the most commonly used metrics for measuring software quality in blockchain systems? The results pointed out that throughput and latency, as widely used metrics, are supported by several identified studies. Croman et al. [11] highlights that these metrics are critical for evaluating the operational efficiency of blockchains, particularly in the context of distributed networks requiring high processing speeds and low latency for reliable transactions. Similarly, Zheng et al. [42] emphasizes that latency is critical to ensuring fast response times, while throughput is a primary indicator of network scalability. Additionally, though less frequent, metrics such as transaction and gas cost prove relevant in specific contexts and represent areas that could be further explored in future studies. In this regard, Alharby [2] suggests that analyzing additional metrics may provide a more thorough understanding of blockchain system performance.

Key takeaway: *Throughput and latency were the most commonly used metrics for measuring software quality in blockchain systems, mainly in distributed networks requiring high speed and low latency for efficient operation. Though less frequently, transaction and gas costs were also considered and proved relevant in specific scenarios. Expanding the focus to include additional metrics could have enhanced the comprehensive understanding of blockchain system quality.*

RQ2) What quality characteristics are frequently assessed in blockchain systems? In particular, Performance emerged as the most frequently assessed quality characteristic in the primary studies, appearing in 282 occurrences, followed by Reliability in 40 occurrences. Meanwhile, Security, despite being a fundamental aspect of blockchain systems [20, 42], was addressed in only 26 occurrences. This findings highlights a gap between its theoretical importance and empirical evaluation, suggesting that security aspects such as degree of anonymity [S56], Data Confidentiality [S6], and structural weaknesses like Private Key Leakage and Mutable Metadata [S2] remain underexplored. Similarly, Usability was scarcely considered, appearing only twice [S1, S25], despite being critical for adoption beyond technical users. The lack of studies addressing usability aspects, such as ease of interaction and private key management, suggests that research primarily focuses on technical performance rather than user experience, for example. Additionally, Functional Suitability (7 occurrences) and Portability (1 occurrence) received little attention, while Compatibility, a key factor for interoperability, was not mentioned at all. Given that interoperability is directly associated with compatibility and is fundamental to enabling integrated and cross-chain solutions [5]. This finding reinforces the necessity of expanding research efforts beyond performance to include security, usability, and interoperability for a more comprehensive evaluation of blockchain systems.

Key takeaway: *Performance, Reliability, and Security were the primary characteristics assessed in the studies, with Performance appearing in 282, 40, and 26 occurrences, respectively. However, Functional Suitability, Usability, and Portability were rarely addressed, while Compatibility was not mentioned. This finding highlights a research gap and the need to broaden evaluations to include these underexplored characteristics for a more robust quality assessment of blockchain systems.*

RQ3) How are blockchain studies' metrics applied and assessed? The Hyperledger Caliper tool, predominantly using computational simulations, was extensively utilized in the analyzed studies to evaluate performance metrics across three key scenarios: system integration, blockchain networks, and smart contracts. However, despite its central importance in various blockchain applications, the smart contract scenario was significantly underutilized. This lower frequency of studies in this context aligns with the observations of Tonelli et al. [37], who points out that research on smart contracts is still in its early stages, with gaps in analyzing and developing metrics to ensure reliability and efficiency. A relevant contribution is presented by Pierro et al. [29], who proposes organized repositories to collect and classify smart contracts. These repositories provide a structured foundation, facilitating empirical analysis and security assessment and advancing research related to developing specific metrics for smart contracts.

Key takeaway: *According to the primary studies, metrics are predominantly applied and assessed through computational simulations, with Hyperledger Caliper being the most widely used tool. In our study, we identified its application in three main scenarios: system integration, blockchain networks, and smart contracts. We primarily focused on the first two, while the smart contract scenario remains overlooked.*

5.1 Implications for IS Research

This systematic mapping provides the Information Systems (IS) academic community with a broad and detailed overview of current research on quality metrics applied to blockchain systems. By identifying and organizing these metrics according to the ISO/IEC 25010 model, the study contributes to theoretical advancements, expanding knowledge in the field and offering a deeper understanding of how blockchain systems are evaluated in terms of quality, encompassing aspects such as performance and security. Additionally, it provides a systematic reference for future studies aiming to develop and refine metrics specific to the blockchain context. Finally, the gaps identified in the literature demonstrate opportunities for new IS studies, encouraging the exploration of metrics not yet widely investigated, such as those focused on usability or compatibility, which are characteristics considerably approached by the IS community.

About the I Grand Research Challenges in Information Systems in Brazil 2016-2026 (GranDSI-BR), this study aligns closely with its emphasis on addressing the complexity of distributed and integrated architectures for information systems. By contributing to the understanding and assessing quality in blockchain systems, this research directly supports challenges identified by the GranDSI-BR, including the need for robust methodologies to ensure interoperability, scalability, and security in modern IS infrastructures. This alignment reinforces the study's relevance to the broader IS research agenda, providing actionable insights into how quality metrics can shape the future of blockchain-based systems in both academic and practical contexts.

5.2 Implications for Practitioners

This study provides practitioners with findings that can inform both the enhancement of existing blockchain systems and the development of new solutions. The key implications are as follows:

- We consolidate quality metrics explored to blockchain systems, offering a foundation for evaluating critical dimensions such as security, performance, and scalability. These metrics enable practitioners to perform more targeted assessments, addressing the specific challenges inherent to decentralized architectures;
- We assist stakeholders in making informed decisions regarding the selection of tools, frameworks, and development methodologies suited to the quality assessment of blockchain-based systems;
- We introduce an initial reference for continuous quality evaluation, supporting identifying and mitigating risks in blockchain projects, especially in high-impact areas like system reliability, interoperability, and data protection.

Furthermore, this research highlights gaps in the current understanding and application of quality metrics, encouraging practitioners to explore less-developed areas, such as usability and compatibility, which are important for user adoption and system integration. By aligning academic research with industry needs, this study offers practitioners evidence-based strategies for addressing the challenges of blockchain quality management, thus contributing to creating robust, scalable, and user-centric blockchain systems that meet real-world demands.

6 Threats to Validity

This section discusses the potential threats to validity identified throughout this study and the measures taken to mitigate them, based on Ampatzoglou et al. [3]'s checklist.

Study Inclusion/Exclusion Bias: A threat to be considered is the bias in study inclusion/exclusion. The search string may not have captured all relevant articles due to term limitations or because some papers were outside the consulted databases. Although a protocol was adopted for study selection, there is a risk of excluding important papers dealing with blockchain quality metrics without explicitly using search string terms. Some studies may have also been excluded for not meeting the inclusion criteria. The search string was formulated based on the ISO/IEC 25010 standard to mitigate this threat and applied across multiple academic databases. The inclusion of additional terms, along with the use of multiple databases, aimed to minimize the exclusion of relevant studies.

Researcher Bias and Repeatability: Two independent reviewers performed data extraction, which may introduce variability in interpretations and selections. Each reviewer may have interpreted the information extracted from papers differently, which could affect the consistency of collected data. To mitigate this risk, the evaluations made by each reviewer were verified by others seeking consensus. Regular meetings to discuss interpretations and resolve discrepancies also helped reduce this threat.

Limited Generalization of Results: The generalization of this study's results to other contexts is a valid concern. The results are based on papers published up to the first quarter of 2024, meaning that very recent results were found. Furthermore, the identified metrics may be specific to certain blockchain systems, limiting applicability in other scenarios or technologies. The constant evolution of blockchain technology suggests that this study's results should be interpreted cautiously, recognizing the possibility of new metrics and methodologies emerging.

Construct Validity: The definition and classification of metrics based on the ISO/IEC 25010 standard threatens construct validity. Although this standard is widely recognized in the software industry, it may not capture all the nuances of blockchain systems, which have unique technical characteristics such as decentralization and cryptographic mechanisms. To mitigate this threat, the study constantly sought to adapt and expand classification categories as new types of metrics emerged in the primary studies.

Conclusion Validity: The conclusions of this study may be limited by the availability and nature of data provided by the reviewed papers. The subjectivity in metric evaluation is an important limitation, as different reviewers may interpret the data differently. Additionally, the lack of detailed data or specific evaluation tools for blockchain in some reviewed studies may have affected the accuracy of the drawn conclusions.

7 Conclusion

This study conducted a Systematic Mapping Study to identify, analyze, and categorize software metrics used for quality assessment in blockchain-based systems, focusing on quality characteristics aligned with the ISO/IEC 25010 standard. Our comprehensive review of 128 primary studies identified the most commonly used metrics, the most frequently evaluated quality characteristics, and the evaluation methods and tools employed in the literature.

The results indicated that performance metrics, such as throughput and latency, are the most prevalent in blockchain studies, highlighting the importance of efficiency and speed in blockchain network transactions. The most frequently evaluated quality characteristics were performance and security, reflecting the critical need for blockchain systems that can handle large transaction volumes securely and efficiently. Furthermore, integration with other systems and interoperability were also prominent focus areas.

The findings of this study have important implications for research and practice in the context of Information Systems. Identifying key metrics and quality characteristics can guide developers and researchers in selecting appropriate tools and methods to evaluate and improve blockchain systems. The emphasis on performance and security underscores the need for robust and reliable solutions for the widespread adoption of blockchain technology across different industries. However, this study also revealed gaps and challenges in metric evaluation for blockchain systems. The variation in metric application and definition and the constant evolution of technology suggest the need for more research to develop and standardize metrics that can better capture the specificities of blockchain systems. Therefore, this research contributed to literature and practice by providing a detailed view of currently used quality metrics.

Future work could explore developing and validating new quality metrics explicitly tailored to the characteristics of blockchain systems, such as decentralization, immutability, and cryptographic security. Additionally, longitudinal studies could assess the practical impact of these metrics on real-world blockchain projects. Investigating underexplored dimensions, such as usability and interoperability, could further enhance the applicability of quality metrics to support blockchain adoption in diverse contexts. Finally, integrating these metrics into automated tools and frameworks could streamline quality assessment processes and promote consistency across blockchain implementations.

References

[1] Rafa E Al-Qutaish. 2010. Quality models in software engineering literature: an analytical and comparative study. *Journal of American Science* 6, 3 (2010), 166–175.

[2] Maher Alharby. 2023. Transaction latency within permissionless blockchains: analysis, improvement, and security considerations. *Journal of Network and Systems Management* 31, 1 (2023), 22.

[3] Apostolos Ampatzoglou, Stamatia Bibi, Paris Avgeriou, Marijn Verbeek, and Alexander Chatzigeorgiou. 2019. Identifying, categorizing and mitigating threats to validity in software engineering secondary studies. *Information and Software Technology* 106 (2019), 201–230.

[4] Andreas M Antonopoulos and Gavin Wood. 2004. *Mastering Ethereum: Building Smart Contracts and Dapps.* O'Reilly Media, Inc., USA. 415 pages.

[5] Rafael Belchior, André Vasconcelos, Sérgio Guerreiro, and Miguel Correia. 2021. A survey on blockchain interoperability: Past, present, and future trends. *Acm Computing Surveys (CSUR)* 54, 8 (2021), 1–41.

[6] Clodis Boscarioli, Renata Mendes de Araujo, Rita Suzana Maciel, Valdemar Vicente Graciano Neto, Flavio Oquendo, Elisa Yumi Nakagawa, Flavia Cristina Bernardini, José Viterbo, Dalessandro Vianna, Carlos Bazilio Martins, et al. 2017. I GranDSI-BR: Grand Research Challenges in Information Systems in Brazil 2016-2026. (2017).

[7] Vitalik Buterin et al. 2013. Ethereum white paper. *GitHub repository* 1 (2013), 22–23.

[8] Sara Cohen and Aviv Zohar. 2018. Database perspectives on blockchains. *arXiv preprint arXiv:1803.06015* (2018).

[9] Marco Conoscenti, Antonio Vetro, and Juan Carlos De Martin. 2016. Blockchain for the Internet of Things: A systematic literature review. In *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*. IEEE, 1–6.

[10] Emanuel F Coutinho, Wagner L Braga Bezerra, and Delano Maia. 2021. Uma Análise Inicial sobre a Aplicaçao de Blockchain na Sociedade. In *Anais do II Workshop sobre as Implicações da Computação na Sociedade*. SBC, 45–56.

[11] Kyle Croman, Christian Decker, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Elaine Shi, Emin Gün Sirer, et al. 2016. On Scaling Decentralized Blockchains: (A Position Paper). In *International conference on financial cryptography and data security*. Springer, 106–125.

[12] Florian Deissenboeck, Elmar Juergens, Klaus Lochmann, and Stefan Wagner. 2009. Software quality models: Purposes, usage scenarios and requirements. In *2009 ICSE workshop on software quality*. IEEE, 9–14.

[13] John Estdale and Elli Georgiadou. 2018. Applying the ISO/IEC 25010 quality models to software product. In *Systems, Software and Services Process Improvement: 25th European Conference, EuroSPI 2018, Bilbao, Spain, September 5-7, 2018, Proceedings 25*. Springer, 492–503.

[14] Markus Fynn, Alysson Bessani, and Fernando Pedone. 2020. Towards Blockchain Interoperability. *Proceedings of the IEEE Symposium on Reliable Distributed Systems* (2020). https://arxiv.org/abs/2004.05933

[15] Xianyi Gao, Gradeigh D Clark, and Janne Lindqvist. 2016. Of two minds, multiple addresses, and one ledger: characterizing opinions, knowledge, and perceptions of Bitcoin across users and non-users. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1656–1668.

[16] ISO. 2023. ISO 25010:2023 - Blockchain and distributed ledger technologies — Reference architecture. https://www.iso.org/standard/78176.html Accessed: 2024-11-12.

[17] Simran Jaitly, Anil Kumar Mishra, and Latika Singh. 2014. A systematic review on the impact of metrics in software process improvement. *Compusoft* 3, 3 (2014), 624.

[18] Mohamad Kassab, Joanna DeFranco, Tarek Malas, Valdemar Vicente Graciano Neto, and Giuseppe Destefanis. 2019. Blockchain: A panacea for electronic health records?. In *2019 IEEE/ACM 1st International Workshop on Software Engineering for Healthcare (SEH)*. IEEE, 21–24.

[19] Staffs Keele et al. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. Technical report, ver. 2.3 ebse technical report. ebse.

[20] Jiewu Leng, Man Zhou, J Leon Zhao, Yongfeng Huang, and Yiyang Bian. 2020. Blockchain security: A survey of techniques and research directions. *IEEE Transactions on Services Computing* 15, 4 (2020), 2490–2510.

[21] Iuon-Chang Lin and Tzu-Chun Liao. 2017. A survey of blockchain security issues and challenges. *Int. J. Netw. Secur.* 19, 5 (2017), 653–659.

[22] Matthew B Miles. 1994. Qualitative data analysis: An expanded sourcebook. *Thousand Oaks* (1994).

[23] Arthur-Jozsef Molnar, Alexandra Neamţu, and Simona Motogna. 2020. Evaluation of software product quality metrics. In *Evaluation of Novel Approaches to Software Engineering: 14th International Conference, ENASE 2019, Heraklion, Crete, Greece, May 4–5, 2019, Revised Selected Papers 14*. Springer, 163–187.

[24] Amrita Raj Mukker, Latika Singh, and Anil Kumar Mishra. 2014. Systematic review of metrics in software agile projects. *Compusoft* 3, 2 (2014), 533–539.

[25] Międzynarodowa Organizacja Normalizacyjna. 2011. *Systems and Software Engineering-Systems and Software Quality Requirements and Evaluation (SQuaRE)-System and Software Quality Models*. ISO.

[26] James A O'Brien and GM Marakas. 2011. Developing business/IT solutions. *Management information systems* 488489 (2011), 74–89.

[27] Marco Ortu, Matteo Orrú, and Giuseppe Destefanis. 2019. On comparing software quality metrics of traditional vs blockchain-oriented software: An empirical study. In *2019 IEEE International Workshop on Blockchain Oriented Software Engineering (IWBOSE)*. IEEE, 32–37.

[28] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic mapping studies in software engineering. In *12th international conference on evaluation and assessment in software engineering (EASE)*. BCS Learning & Development.

[29] Giuseppe Antonio Pierro, Roberto Tonelli, and Michele Marchesi. 2020. Smartcorpus: an organized repository of ethereum smart contracts source code and metrics. *arXiv preprint arXiv:2011.01723* (2020).

[30] Simone Porru, Andrea Pinna, Michele Marchesi, and Roberto Tonelli. 2017. Blockchain-oriented software engineering: challenges and new directions. In *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. IEEE, 169–171.

[31] Benedikt Putz and Günther Pernul. 2020. Detecting blockchain security threats. In *2020 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 313–320.

[32] Attia Qammar, Ahmad Karim, Huansheng Ning, and Jianguo Ding. 2023. Securing federated learning with blockchain: a systematic literature review. *Artificial Intelligence Review* 56, 5 (2023), 3951–3985.

[33] Siraj Raval. 2016. *Decentralized applications: harnessing Bitcoin's blockchain technology.* " O'Reilly Media, Inc.".

[34] R. Saraiva, A. A. Araújo, P. Soares, Pontes J. C., and J. Souza. 2024. Supplementary material of Metrics for Quality Assessment in Blockchain-based Systems: A Systematic Mapping Study. https://zenodo.org/records/14198721.

[35] Rajesh Soundararajan and VM Shenbagaraman. 2023. Unlocking the Potential of Blockchain Through Multi-Criteria Decision Making in Platform Selection. *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.* 8, 4 (2023), 49.

[36] Don Tapscott and Alex Tapscott. 2016. *Blockchain revolution: how the technology behind bitcoin is changing money, business, and the world.* Penguin.

[37] Roberto Tonelli, Giuseppe Destefanis, Michele Marchesi, and Marco Ortu. 2018. Smart contracts software metrics: a first study. *arXiv preprint arXiv:1802.01517* (2018).

[38] Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper* 151, 2014 (2014), 1–32.

[39] Xiwei Xu, Ingo Weber, and Mark Staples. 2019. *Architecture for blockchain applications.* Springer.

[40] Xiwei Xu, Ingo Weber, Mark Staples, Xiwei Xu, Ingo Weber, and Mark Staples. 2019. Blockchain patterns. *Architecture for Blockchain Applications* (2019), 113–148.

[41] Jesse Yli-Huumo, Deokyoon Ko, Sujin Choi, Sooyong Park, and Kari Smolander. 2016. Where is current research on blockchain technology?—a systematic review. *PloS one* 11, 10 (2016), e0163477.

[42] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. 2017. An overview of blockchain technology: Architecture, consensus, and future trends. In *2017 IEEE international congress on big data (BigData congress)*. Ieee, 557–564.

[43] Sisi Zhou, Kuanching Li, Lijun Xiao, Jiahong Cai, Wei Liang, and Arcangelo Castiglione. 2023. A systematic review of consensus mechanisms in blockchain. *Mathematics* 11, 10 (2023), 2248.

[44] Mengmeng Zhu and Hoang Pham. 2022. Software reliability modeling and methods: A state of the art review. *Optimization Models in Software Reliability* (2022), 1–29.