Metadata models for Quantum Circuits Experiments: A Systematic Mapping

Samuel Sepúlveda[1], Nawel Huenchuleo[1], and Ania Cravero[1]

Depto. Ciencias de la Computación e Informática, QuantumLab-DCI
Universidad de La Frontera, Temuco, Chile
{samuel.sepulveda,ania.cravero}@ufrontera.cl; n.huenchuleo01@ufromail.cl

Abstract. Statistical reproducibility in Noisy Intermediate-Scale Quantum Computing (NISQ) is hindered by hardware variability and by the lack of metadata management practices able to capture calibration, compilation decisions, and data provenance. Although these elements are essential for experimental traceability, the evidence on how they are modeled and operationalized remains scattered. This paper characterizes the state of the art in metadata management for quantum software, focusing on gaps that constrain interoperability, auditability, and statistical reproducibility. We conducted a systematic mapping study that included 40 primary studies. The results reveal a fragmented landscape with limited adoption of formal standards. Existing approaches are mostly provider-specific, conceptually incomplete, or weakly integrated with compilation and execution workflows. Overall, current approaches remain insufficient to support auditable and reproducible quantum experimentation in NISQ environments.

Keywords: systematic mapping study, quantum computing, quantum circuit, quantum metadata

Modelos de metadatos para experimentos en circuitos cuánticos: Un Mapeo Sistemático

Abstract. La reproducibilidad estadística en la era de la Computación Cuántica de Escala Intermedia Ruidosa (NISQ, Noisy Intermediate-Scale Quantum) se ve limitada por la variabilidad del hardware y por la ausencia de prácticas de gestión de metadatos capaces de capturar calibración, decisiones de compilación y procedencia de los datos. Aunque estos elementos son fundamentales para la trazabilidad experimental, la evidencia sobre cómo se modelan y operacionalizan sigue siendo dispersa. Este trabajo caracteriza el estado del arte en gestión de metadatos para software cuántico, enfocándose en brechas que restringen la interoperabilidad, auditabilidad y reproducibilidad estadística. Se realizó un mapeo sistemático de literatura con 40 estudios primarios. Los resultados evidencian un panorama fragmentado, con baja adopción de estándares formales. Los enfoques existentes son, en su mayoría, específicos de proveedor, conceptualmente incompletos o débilmente integrados con los flujos de compilación y ejecución. En conjunto, los enfoques actuales siguen siendo insuficientes para sostener experimentación cuántica auditable y reproducible en entornos NISQ.

Keywords: mapeo sistemático, computación cuántica, circuito cuántico, metadatos cuánticos

1 Introduction

Quantum computing (QC) represents a computational paradigm that is radically different from classical computing. While classical systems process information through bits in discrete states, QC employs qubits that exploit quantum-mechanical phenomena such as superposition, entanglement, and interference to perform computations (Preskill, 2018). These properties allow a system of n qubits to explore 2^n states simultaneously, enabling exponential computational advantages for specific classes of problems (Nielsen & Chuang, 2010).

Current QC developments take place in the Noisy Intermediate-Scale Quantum (NISQ) era, characterized by intermediate-scale devices that remain highly sensitive to noise and still lack robust error-correction capabilities (Preskill, 2018). As a result, the execution of quantum algorithms is strongly conditioned by time-dependent hardware factors such as decoherence and imperfect gate fidelity (IBM, 2023; Krantz et al., 2019). Under these conditions, nominally identical executions can produce significantly different outcomes due to undocumented fluctuations in hardware properties, making statistical reproducibility a central challenge in quantum experimentation (LaRose, 2019; Tannu & Qureshi, 2019).

Against this backdrop, metadata management has become a key concern for hybrid quantum-classical architectures, as structured records of contextual information, compilation decisions, and execution conditions are needed to support traceability, auditability, and statistical reproducibility (Senapati et al., 2023; Weder et al., 2022). Robust metadata models also affect the technical performance of such systems. The availability of detailed contextual information enables compilers to adopt hardware-aware optimization strategies, which have been shown to improve the fidelity of executed circuits (Murali et al., 2019; Sivarajah et al., 2020). Therefore, metadata should be regarded as an active enabler of performance improvement in QC.

This paper reports a Systematic Mapping Study (SMS) aimed at characterizing the state of the art in metadata models for quantum software, identifying the prevailing approaches and technological gaps that restrict interoperability, auditability, and statistical reproducibility in NISQ systems.

The remainder of the paper is organized as follows. Section 2 presents the background. Section 3 presents the methodology. The results and a discussion are given in Sections 4 and 5. Section 6 presents the conclusions and future work.

2 Background

Metadata in QC scenarios require us to understand fundamental quantum principles, the limitations of current hardware, and circuit transformation processes.

2.1 Quantum Computing and the NISQ Era

Unlike classical computing, QC relies on qubits and quantum-mechanical phenomena such as superposition, entanglement, and interference, which allows state vectors to be processed in 2^n-dimensional Hilbert spaces, providing exponential advantages over classical computing (Nielsen & Chuang, 2010).

In practice, however, current quantum devices operate in the NISQ era, where limited coherence and noise make experimental outcomes strongly dependent on fluctuating hardware conditions (Krantz et al., 2019; Preskill, 2018). Parameters such as relaxation time (T_1) and decoherence time (T_2) define the 'lifetime' of quantum information, but their temporal variability directly affects the statistical reproducibility of experiments (Dasgupta et al., 2022).

2.2 Transpilation and Hybrid Architectures

To execute a quantum algorithm on real hardware, it must undergo a transpilation process, adapting the circuit to the physical and topological constraints of the hardware. First, circuit decomposition to translate it into native base gates (Aleksandrowicz et al., 2019). Second, mapping to assign logical qubits to physical qubits (Sivarajah et al., 2020). Third, routing SWAP gates to satisfy restricted connectivity (Li et al., 2019). Finally, optimization to reduce depth and noise (Duncan et al., 2020). Given coherence limitations, current approaches focus on hybrid architectures, where the QPU works in synergy with classical processors (Sivarajah et al., 2020).

2.3 Metadata, Traceability and statistical reproducibility

Metadata is essential to capture the context, provenance, and quality of data (Gilliland, 2008). In complex systems, provenance documents the history of data transformations, with the W3C-PROV standard serving as the reference for modeling these relationships through entities, activities, and agents (Moreau & Missier, 2013). In the NISQ era, traceability is vital for linking results to specific calibration configurations (Weder et al., 2022). Scientific reproducibility requires obtaining consistent results under similar conditions (Wohlin et al., 2012). However, the lack of recording of environmental variables (e.g. T_1 and T_2)

creates a contextual validity gap (Dasgupta et al., 2022). Since quantum measurement is probabilistic and requires multiple runs, the results must be treated as probability distributions. Therefore, statistical reproducibility is proposed, by which metadata must capture not only the output value, but also histograms and sampling configurations to ensure traceability (Senapati et al., 2023).

3 Methodology

This study follows an SMS methodology according to (Petersen et al., 2015). Next, we detail the *planning and conducting* phases.

3.1 Planning phase

Aim and need. The purpose of this SMS is to characterize the state of the art in metadata models applied to QC architectures, with the view to identify existing approaches, technical limitations, and research gaps in the context of NISQ systems. This study aims to structure and classify scientific contributions related to metadata management, provenance, quantum compilation, and experimental statistical reproducibility, providing a comprehensive overview of the field.

Motivation comes from the need to address the statistical reproducibility crisis in the NISQ era, where hardware variability and the opacity of compilation processes generate uncertainty in experimental results. In this context, metadata plays a fundamental role by enabling the capture of the execution context, circuit transformation decisions, and the physical conditions of the device. These issues suggest potential limitations in the practical adoption and effectiveness of current metadata approaches, which are examined in the present SMS.

Consequently, this SMS lays the foundations for the development of more robust and adaptive metadata models, designed to improve statistical reproducibility, interoperability, and auditability in hybrid quantum systems.

Research Questions (RQs). This SMS seeks to answer RQs aimed at systematizing available knowledge, identifying approaches to metadata modeling, and analyzing the limitations in their application. The RQs are detailed in Table 1.

Search String. The construction of the search string was based on Kitchenham's (2015) guidelines. First, we extracted key terms from the RQ and the study context. Second, we identified synonyms and alternative terms to broaden the scope of the search. Third, we applied the PICOC criterion (Petticrew & Roberts, 2008) to structure the search strategy.

The search was executed in Semantic Scholar as the primary engine, using its ability to index multiple relevant sources. Five specific search strings were defined to capture different dimensions of the problem (see details in Table 2).

Methodology validation. This SMS was conducted following the empirical software engineering guidelines (Petersen et al., 2015). The protocol defined the search strategy, inclusion and exclusion criteria, and the data extraction and analysis procedure. Semantic Scholar was used as a meta-search engine to

Table 1: Research questions (RQs).

#	RQ	Objective	Classification Scheme
RQ1	What metadata models and approaches are used in quantum computing?	Identify existing metadata modeling proposals.	Model types, formalization level, and application domain.
RQ2	How is traceability and data lineage addressed in quantum systems?	Analyze provenance and auditing mechanisms.	Use of standards, record granularity level, representation of processes and relationships.
RQ3	How are metadata managed in quantum compilation and execution processes?	Evaluate integration with the hybrid workflow.	Compiler decisions, recording of metrics.
RQ4	What limitations exist in the statistical reproducibility of quantum experiments?	Identify gaps in experimental context management.	Lack of calibration versioning, temporal variability, statistical reproducibility.
RQ5	What technological gaps exist in metadata management for NISQ systems?	Detect research opportunities.	Format fragmentation, lack of interoperability, low adoption of standards.

Table 2: Query strings specification.

ID	Query String	Technical Goal
Q1	"quantum computing" AND "provenance" AND "NISQ"	Identify papers on traceability in noisy environments.
Q2	"quantum" AND "metadata model"	Capture metadata model proposals.
Q3	"quantum circuit compilation" AND "metadata management"	Analyze the use of metadata in compilers.
Q4	"quantum circuit" AND "compilation" AND "metadata"	Broaden coverage in software engineering.
Q5	"NISQ" AND "calibration" AND "data management"	Identify calibration management systems.

broaden coverage and reduce selection bias. The selection of the paper was performed through the screening (title and abstract), the evaluation of the eligibility of the full-text, and the consolidation of the final corpus. Data extraction relied on a standardized template, reducing subjectivity and maintaining consistency.

Although no formal external review was performed, the principles of transparency and traceability were prioritized, documenting each stage of the SMS process. This approach ensures the methodological consistency of the study and provides a solid foundation for the systematization of knowledge in the field of metadata management for hybrid quantum-classical architectures. For traceability and reliability purposes, the full protocol is available in (Sepúlveda & Huenchuleo, 2026).

3.2 Conducting phase

First, we remove duplicates. Subsequently, a three-stage filtering procedure (metadata filtering, title and abstract screening, and full-text eligibility assessment) based on predefined IC and EC, resulting in the final corpus of papers. Data extraction was performed using a standardized template to ensure consistency and reduce interpretation bias, followed by thematic categorization to identify trends and research gaps (Petersen et al., 2015).

Search strategy. The search strategy was designed to ensure broad and representative coverage of the relevant literature, minimize selection bias, and follow best practices for secondary studies in software engineering. Given the interdisciplinary nature of the problem, which encompasses QC, metadata management, and compilation, a high coverage meta-search engine was selected.

The Semantic Scholar Search Engine[1] was used as an information retrieval strategy. This tool was selected for its ability to index and retrieve relevant literature from multiple primary sources (e.g. IEEE Xplore, Springer, and arXiv), using artificial intelligence algorithms (Asta Ai2[2]), allowing comprehensive coverage of the domain. In this context, the queries were executed on October 21 and 22, 2025, with the aim of capturing the most up-to-date state of the art in metadata management and traceability in quantum systems.

The search process consisted of an automated phase in which five search strings based on the RQs were applied to the selected source.

Inclusion and exclusion criteria. To ensure the relevance and quality of the selected papers, explicit inclusion (IC) and exclusion criteria (EC) were defined. These criteria allowed for the filtering of literature according to the SMS's objective and ensured the pertinence of the final corpus. See details in Table 3.

Table 3: Inclusion and exclusion criteria.

Inclusion Criteria (IC)	
IC1	**Temporality:** Papers published between 2017 and 2025. For RQ1 only, seminal pre-2017 studies were exceptionally retained when they introduced foundational metadata models or specifications necessary to contextualize later NISQ-oriented proposals.
IC2	**Language:** Papers written in English.
IC3	**Publication Type:** Scientific journal articles, peer-reviewed conference proceedings, and arXiv preprints with verifiable citations.
IC4	**Subject Matter:** Papers addressing metadata, provenance (lineage), compilation, calibration, or statistical reproducibility in QC.
IC5	**Relevance:** Papers whose contribution is directly or complementarily applicable to the data architecture problem.
Exclusion Criteria (EC)	
EC1	**Obsolescence:** Papers published before 2017 (except seminal studies retained under the RQ1 exception described in IC1).
EC2	**Out of Scope:** Specific applications that do not contribute to metadata management or software engineering.
EC3	**Grey Literature:** Tutorials, news, posters, or papers without full text.
EC4	**Redundancy:** Duplicates or preliminary versions of works.

Selection process. The study selection process was conducted in three main stages, following the methodological guidelines proposed by Petersen et al. (2015), to ensure a rigorous, reproducible and transparent mapping process. From the automated search, an initial set of 181 records was obtained.

[1] https://www.semanticscholar.org/
[2] https://asta.allen.ai/

- ***Stage 1:*** *Initial filtering (title, abstract, and keywords).* In this stage, the articles retrieved through automated search were first filtered according to their titles, abstracts, and keywords. We applied the IC1, IC2, IC3, IC5 and EC1 criteria. We aim to eliminate irrelevant records and ensure alignment with the scope. Most articles published before 2017 were excluded under IC1 and EC1; however, three seminal studies before 2017 were exceptionally retained for RQ1 because they provide the historical baseline for the analysis of formal metadata models.
- ***Stage 2:*** *Intermediate filtering (scope refinement).* This stage allowed for a more specific selection of papers focused on quantum-classical architectural approaches or related systems integration strategies, while excluding papers not aligned with the research topic (EC2) or not suitable in format (EC4).
- ***Stage 3:*** *Final filtering and eligibility.* This stage was conducted to ensure the completeness and uniqueness of the papers. Only the most recent version of each study was kept (IC4), and duplicate or inaccessible articles were removed (EC3 and EC4). Articles that passed this stage were submitted to full-text evaluation to confirm their contribution to at least one RQ. Each selected article was assigned a unique identifier for tracking during the analysis.

Classification and categorization. Following the recommendations of (Petersen et al., 2015), a thematic categorization of the selected articles was carried out to structure the findings. The emerging categories (Formal Models, Contextual Compilation, Statistical reproducibility, Traceability) allowed for the identification of patterns and the gaps that this work aims to address.

4 Results

This section presents the SMS findings, detailing the selection process and the identification of technological gaps in quantum metadata management.

4.1 Overview of study selection

The execution of the five search queries (see Table 2) yielded a total corpus of 181 candidate papers (Q1: 40; Q2: 30; Q3: 38; Q4: 36; and Q5: 37 papers).

Given that the search strings were designed to cover complementary aspects of the problem, a non-empty intersection of results between the different queries was anticipated. Of the 181 initial papers identified, 43 redundant entries were detected. After this cleaning, a single corpus of 138 papers was obtained. Subsequently, this refined set was subjected to the refinement process, rigorously applying the (IC) and (EC). Finally, 40 primary studies were selected. Figure 1 illustrates the document reduction workflow.

This final corpus provides the basis for the thematic analysis reported in RQ1–RQ4. The full list of selected studies is available in the public study protocol and dataset (Sepúlveda & Huenchuleo, 2026).

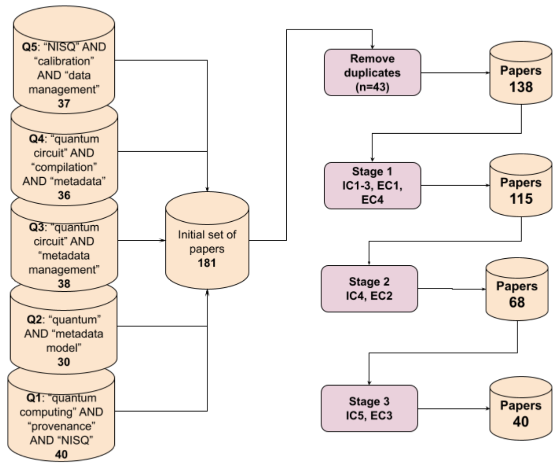

Fig. 1: Results of the selection process (PRISMA diagram).

4.2 Formal Metadata Models (RQ1)

A key finding of this study is the scarcity of hardware-agnostic metadata models, understood as those capable of representing quantum experiments independently of the underlying technology or provider, thus enabling interoperability without structural refactoring. Of the 40 fully reviewed papers, only 8 (20%) propose structured metadata schemes. The analysis shows that, although some approaches are theoretically agnostic, they lack the technical depth required to capture critical aspects of NISQ such as noise and calibration.

In general, the state of the art reveals a clear structural fragmentation. Early standards are now obsolete, as they focus on logical circuit representation and ignore the physical layer. In contrast, recent proposals address calibration and execution but are tightly coupled to specific ecosystems, limiting interoperability. Other proposals remain conceptual, lack implementation, or focus on provenance or data storage, rather than the full experimental context.

This dichotomy reveals *Gap 1 (Fragmentation).* No existing metadata standard achieves hardware independence, specification of dynamic physical calibration, and compatibility with compilation workflows. For details, see Table 4.

This gap was identified by comparing each proposal in four dimensions—formalization, provider agnosticism, support for dynamic calibration, and integration with compilation workflows—and observing that no study covered these dimensions simultaneously.

Consequently, RQ1 includes a small number of foundational studies before 2017 as historical baselines, since the objective of this RQ is to characterize the evolution of metadata modeling proposals rather than only their NISQ-era instantiations.

Table 4: Analysis of existing metadata models and standards (RQ1).

ID	Proposal	Main Focus	Identified Limitation (Gap)
6	QIS-XML	First XML-based standard to describe logic circuits.	**Obsolescence:** Lacks structure for calibration or noise (Pre NISQ).
9	Quantum UML	UML extension for modeling quantum systems.	**Abstraction Level:** Not capture physical execution or calibration.
36	Qibocal	For QPU characterization and calibration.	**Coupling:** Designed for the Qibo / Rigetti ecosystem, difficult to port.
25	Quantum Data Mgmt	Data management for scientific experiments.	**Theoretical:** Robust model, lacks implementation in popular SDKs.
40	Ecosystem Agnostic	Metadata for cloud orchestration.	Focused on deployment not the physical statistical reproducibility of the experiment.
2	Quantum Provenance	Provenance model for workflows.	Focuses on process lineage, not detail hardware structure (device metadata).
7	QIS-SQL	Relational database extension for QC.	Focuses on classical storage, not on quantum experiment semantics.
8	OpenFlow Metadata	Metadata for quantum network control.	Specific to communications, not applicable to gate-based computing.

4.3 Traceability and Provenance (RQ2)

A total of 7 papers (17.5%) address the traceability and provenance of the data. The analysis reveals a domain characterized by strong theoretical development but limited practical adoption. Although some approaches adapt established standards to the quantum domain, most proposals remain at an abstract or infrastructural level. These solutions either rely on complex middleware architectures or focus on high-level workflow orchestration, without integrating into low-level experimental processes. In particular, no lightweight libraries or tools were identified for integration into existing development environments.

From the results between theory and practice arise *Gap 4 (Lineage Disconnection)*. Although formal provenance models exist, their adoption in experimental workflows is minimal due to the lack of accessible tools that automate fine-grained data lineage capture. For details, see Table 5.

The gap was derived from the recurrent contrast between semantically rich provenance proposals and the absence of lightweight workflow-level mechanisms that can be readily adopted in everyday experimental settings.

4.4 Context-Aware Compilers (RQ3)

A total of 15 papers (37.5%) address circuit transformation, confirming noise-aware compilation as one of the most active areas in quantum software engineering. The analysis shows that, although multiple strategies exist to adapt logical circuits to hardware topology and noise, most compilers operate under a black-box paradigm. Advanced tools leverage calibration data to optimize circuits, but rarely expose structured metadata that describe key decisions such as qubit mapping, SWAP insertion, or optimization passes.

Across the reviewed approaches, transparency is consistently limited: some provide partial insights, others generate non-standardized or non-persistent data,

Table 5: Proposals on Provenance and Traceability (RQ2).

ID	Proposal	Maturity Level (Gap)
2	QProv (W3C PROV Ontology)	**Theoretical:** Robust semantic model, but lacks accessible implementation libraries for the end user.
3	Quantum Lifecycle	**High-level:** Based on BPMN; useful for process orchestration, but does not descend to the level of shots or physical calibration.
4	Analysis & Provenance	**Service-oriented:** Focused on the orchestration of web services and containers, ignoring the device's internal physics.
5	Qunicorn	**Complex Infrastructure:** Heavy microservices solution, difficult to adopt in simple experimental scripts.
24	Overview QC	**Diagnostic:** Identifies the lack of standardization as a risk, without proposing a concrete technical solution.
31	Calibration Use	**Pragmatic (Non-standardized):** Connects calibration with results through ad-hoc solutions with no formal reusable format.
34	Integrated Tools	**Development Environment:** Proposes a graphical interface (IDE), but does not define an underlying data exchange model.

and several (particularly those based on machine learning or real-time optimization) are inherently opaque. As a result, critical transformation steps cannot be reconstructed, preventing full traceability of the experimental workflow.

This systemic lack of visibility arises from *Gap 2 (Transpilation Opacity).* Current compilers optimize execution performance but do not provide auditable and structured metadata about the transformation process. For details, see Table 6.

This gap was inferred from the repeated absence of persistent structured records describing mapping, routing, optimization passes, and other transformation decisions in otherwise advanced compilation approaches.

4.5 Statistical reproducibility and Variability in NISQ (RQ4)

A total of 10 papers (25%) address statistical reproducibility and experimental stability, providing strong evidence of intrinsic variability of the NISQ hardware. Longitudinal analysis shows that key hardware properties, such as coherence times (T1, T2) and gate error rates, exhibit significant fluctuations over short time frames (hours or days). Under these conditions, exact reproducibility is unattainable, and recent work proposes a shift towards statistical reproducibility, which requires access to historical calibration data.

The evidence consistently shows that experimental results cannot be reliably interpreted without incorporating the hardware state at execution time. However, current approaches rely on static metadata and lack mechanisms to represent this evolving context, limiting their explanatory power.

This limitation arises in *Gap 3 (Unmanaged Volatility).* Without metadata that captures time-dependent calibration and hardware conditions, it is impossible to determine whether variations in results are caused by algorithmic behavior or hardware drift. For details, see Table 7.

This gap emerges from the convergence of empirical studies showing time-dependent hardware variation together with the lack of metadata mechanisms that can preserve the calibration state associated with a given execution.

Table 6: Analysis of Compilation Strategies and Transparency Level (RQ3).

ID	Strategy / Tool	Transparency Level (Gap)
1	Not All SWAPS (SWAP Optimization)	**Partial:** Correctly optimizes, but does not document the final state of the physical links used.
10	tket (Retargetable Compiler)	**None (Black Box):** Internal optimization process; user receives the final circuit without a trace of the changes.
11	Full stack Optimization (Vertical)	**Low:** Generates high data granularity that is lost due to the lack of a standardized output format.
12	Software Mitigation (Crosstalk)	**Partial:** Uses calibration data internally but does not export them for external validation.
13	Noise-Adaptive (Error Mapping)	**None (Black Box):** No record is provided to justify the mapping selection.
14	Time-Stitched (Distributed)	**Low:** Introduces temporal complexity (network latencies) not modeled in current metadata.
15	QContext (Context aware)	**None:** Prioritizes real-time efficiency, sacrificing data auditing.
16	Modular Compilation	**Partial:** Proposes modularity but does not define the resulting data structure.
17	Relaxed Peephole	**Low:** Aggressive local transformations make it difficult to reconstruct the original logical traceability.
19	MQT Predictor (Machine Learning)	**Opaque (AI):** Uses metadata as input, but the model acts as a black box.
29	Calibration aware (Freshness)	**Partial:** Uses recent data but does not link the calibration timestamp to the result.
30	Pulse-level Control	**Low:** Operates at a deep physical level without standardization to represent those data.
26	Just in Time (JIT)	**Ephemeral:** The volatile nature of JIT makes it difficult to persist the final version of the executed circuit.
37	Noise Mitigation (Rewriting)	**Opaque:** Alters the logical structure without explicitly reporting the transformations.

Table 7: Studies on Statistical Reproducibility and Hardware Variability (RQ4).

ID	Study Focus	Main Finding	Implication for Metadata
20	Verifying Results	Verification of results in quantum workflows.	Requires knowing the exact circuit executed post-compilation.
21	SupermarQ	Benchmarking suite for QC.	Benchmarks lose validity without the calibration context of the day.
22	Quantum Cam	Visual inspection of quantum processes.	Hardware focus, complementary to software metadata.
23	Quantum Hardware	Analysis of hardware constraints.	Defines the static attributes that DeviceMetadata must have.
27	Redefining Repro.	Proposes “statistical reproducibility”.	Reproducibility requires capturing distributions, not just values.
28	Efficient Calib.	Efficient calibration methods.	Calibration is costly; its data must be stored and reused.
32	Stability Analysis	Temporal stability of noisy processors.	Evidence that hardware changes drastically in hours (T1 and T2).
33	Long-Term Analysis	Longitudinal analysis of IBM devices.	Historical data are key to predicting failures.
37	Noise Analysis	Noise characterization for compilers.	Noise profiles must be explicit inputs of the model.
38	Qubit Health	Qubit health monitoring.	Proposes health metrics that the model must be able to store.

5 Discussion

This section integrates the findings from RQ1–RQ4 and discusses the cross-cutting gaps that answer RQ5.

5.1 Research Gaps and Opportunities

The evidence synthesized across RQ1–RQ4 indicates that current limitations in quantum metadata management are not isolated deficiencies, but manifestations of a broader infrastructural problem. Existing approaches remain fragmented across modeling, provenance, compilation, and statistical reproducibility concerns, which prevents metadata from functioning as an integrated software artifact throughout the experimental lifecycle. In practice, this affects three properties that are central to NISQ experimentation: interoperability across tools and providers, auditability of circuit transformations and execution context, and statistical reproducibility under time-varying hardware conditions. From this perspective, the four gaps identified in this study—fragmentation, transpilation opacity, unmanaged hardware volatility, and weak provenance integration—should be understood as interdependent obstacles to the maturation of quantum software engineering rather than as isolated technical issues.

Gap 1 *(Fragmentation & lack of agnosticism).* Current models are obsolete, purely conceptual, or closely coupled with specific vendors, limiting interoperability and portability. This highlights the need for unified and agnostic models.
Gap 2 *(Transpilation opacity).* The quantum compiler operates largely as a black box, hiding metadata about transformation decisions, hindering traceability, and statistical reproducibility. This opens a research line on explainable compilation.
Gap 3 *(Unmanaged hardware volatility).* The temporal variability of the hardware is not captured, preventing a reliable interpretation of the results. This requires incorporating calibration versioning and time-dependent hardware states.
Gap 4 *(Lineage disconnection).* Although provenance models exist, their lack of practical implementation limits adoption. Lightweight and integrated solutions are needed to enable automated data lineage capture in real workflows.

Collectively, these gaps answer RQ5 and reveal the absence of a robust metadata infrastructure across the quantum computing stack as a systematic issue.

Recent work has begun to operationalize part of this agenda through concrete metadata and schema proposals for traceable quantum-circuit experimentation. In particular, QC-MM (Huenchuleo et al., 2026) illustrates how some of the requirements identified in this SMS—such as calibration binding, compilation traceability, provenance links, and schema-based validation—can be integrated into a concrete platform-agnostic metadata model.

5.2 Limitations and Threats to Validity

This study presents typical limitations of secondary research (Petersen et al., 2015). First, we mitigated *selection bias* by using Semantic Scholar to cover multiple types of indexed databases, thus reducing the risk of omitting relevant literature not indexed in a single source. In addition, RQ1 retained a small number of foundational studies before 2017 as an explicit exception to the main temporal window. This exception may introduce a limited historical bias in RQ1; however, it was intentionally bounded to a small set of seminal studies required to contextualize the evolution of metadata models. Then, we mitigated interpretation by controlling subjectivity in data classification and extraction, and we used a standardized template, focusing on binary technical characteristics instead of qualitative descriptions. The interpretation was based on explicit evidence; however, subjectivity may arise due to the multidisciplinary nature of the field (interpretive validity). Finally, the results reflect recent NISQ-focused literature and may not fully generalize to emerging developments (limited generalization).

6 Conclusions

This paper presented a Systematic Mapping Study on metadata management in quantum software. The analysis was structured around four analytical dimensions—metadata modeling, provenance/traceability, compilation, and statistical reproducibility—and showed that current approaches remain insufficient to support interoperability, auditability, and statistical reproducibility in hybrid quantum-classical experimentation.

Future work should address the development of a unified and hardware-agnostic metadata model; enable transparent and traceable compilation processes; incorporate calibration-aware, time-dependent metadata to support statistical reproducibility; and bridge the gap between theoretical provenance models and practical implementations through lightweight, integrable tools. Advancing these directions is key to transitioning to a mature discipline with reproducible and auditable experimental practices.

Acknowledgments. Samuel Sepúlveda is funded by ANID, Initiation Research Project (grant number 11240702). Thanks to RIPAISC (CyTED 525RT0174). The authors also thank to Fernanda Gutíerrez for her useful technical support.

Disclosure of Interests. The authors have no competing interests to declare.

References

Aleksandrowicz, G., et al. (2019). Qiskit: An open-source framework for quantum computing. https://doi.org/10.5281/zenodo.2562110

Dasgupta et al., S. (2022). Characterizing and mitigating noise in nisq devices. *IEEE Int. Conference on Quantum Computing and Engineering (QCE).*

Duncan, R., Kissinger, A., Perdrix, S., & Van De Wetering, J. (2020). Graph-theoretic simplification of quantum circuits with the zx-calculus. *Quantum*, *4*, 279.

Gilliland, A. J. (2008). *Setting the stage* (3rd ed.). Getty Research Institute.

Huenchuleo, N., Sepúlveda, S., & Fernández, A. (2026). QC-MM: A metadata and schema model for traceable quantum-circuit experiments. *Applied Sciences*, *16*(13). https://doi.org/10.3390/app16136346

IBM. (2023). IBM debuts next-generation quantum processor & ibm quantum system two [Accessed: 2025].

Krantz, P., Kjeldsen, M., Lin, F., Bylander, J., & Oliver, W. D. (2019). A quantum engineer's guide to superconducting qubits. *Applied Physics Reviews*, *6*(2).

LaRose, R. (2019). Overview and comparison of gate level quantum software platforms. *Quantum*, *3*, 130.

Li, G., Ding, Y., & Xie, Y. (2019). Tackling the qubit mapping problem for nisq-era quantum devices. *24th Int. Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 1001–1014.

Moreau, L., & Missier, P. (2013, April). *PROV-DM: The PROV Data Model* (W3C Recommendation). W3C. https://www.w3.org/TR/prov-dm/

Murali, P., Linke, N. M., Martonosi, M., Abhari, A. J., Nguyen, N. H., & Adler, C. H. (2019). Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. *24th Int. Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 1015–1029.

Nielsen, M. A., & Chuang, I. L. (2010). *Quantum computation and quantum information: 10th anniversary edition*. Cambridge University Press.

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1–18.

Preskill, J. (2018). Quantum computing in the nisq era and beyond. *Quantum*, *2*, 79.

Senapati, P., Wang, Z., Jiang, W., Humble, T. S., Fang, B., Xu, S., & Guan, Q. (2023). Towards redefining the reproducibility in quantum computing: A data analysis approach on nisq devices. *IEEE Int. Conf. on Quantum Computing and Engineering (QCE)*, *1*, 468–474.

Sepúlveda, S., & Huenchuleo, N. (2026, April). Metadata models for Hybrid Quantum-Classical Systems: A systematic mapping study protocol. https://doi.org/10.17605/OSF.IO/C3SDH

Sivarajah, S., Dilkes, S., Cowtan, A., Simmons, W., Edgington, A., & Duncan, R. (2020). T|ket>: A retargetable compiler for nisq devices. *Quantum Science and Technology*, *6*(1), 014003.

Tannu, S. S., & Qureshi, M. K. (2019). Not all qubits are created equal: A case for variability-aware policies for nisq grids. *24th Int. Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 987–999.

Weder, B., Barzen, J., Leymann, F., & Vietz, D. (2022). Quantum software development lifecycle. In *Quantum software engineering* (pp. 61–83). Springer.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer.