A Data Architecture for Building Health Data Governance and Its Logical Integrity Validation

Xiaoying Li
International College of Digital Innovation (ICDI), Chiang Mai University
Chiang Mai, Thailand

Nopasit Chakpitak
International College of Digital Innovation (ICDI), Chiang Mai University
Chiang Mai, Thailand

Fang Miao
Big Data Research Institute, Chengdu University
Chengdu, China

Piyachat Udomwong
International College of Digital Innovation (ICDI), Chiang Mai University
Chiang Mai, Thailand

ABSTRACT
To address the dispersed sources, inconsistent representations, insufficient contextual associations, and non-uniform authorization mechanisms of building health data, this study proposes a Building Health Data Architecture (BHDA). The architecture enables the unified organization and controlled provision of raw and derived data through data preprocessing, standardized coding, unique identification, metadata registration, rights-attribution recording, and authorization rules. In the case validation, 300 data objects were registered, the derived-data association rate reached 100%, and all eight logical integrity checks were passed. The architecture also supported unauthorized-access rejection, multi-criteria queries, and standardized data output. The results demonstrate that BHDA provides a feasible framework for the governance and sharing of building health data.

General Terms
Data Management; Information Architecture; Security; Standardization.

Keywords
Building health data; Data governance; Data architecture; Metadata registration; Access control.

1. INTRODUCTION
The widespread adoption of the Internet of Things (IoT) and intelligent sensing technologies has gradually shifted building operation and environmental monitoring from periodic inspection toward continuous data acquisition [1]. The information required for building health analysis includes not only environmental parameters, such as temperature, relative humidity, illuminance, and $PM_{2.5}$, but also engineering context involving projects, buildings, floors, rooms, sensors, acquisition times, and monitoring batches. Meanwhile, the increasing incorporation of Building Information Modeling (BIM), Geographic Information Systems (GIS), meteorological data, energy consumption data, and external analysis programs into building data systems has introduced substantial differences in data sources, formats, temporal granularities, and spatial scales [2]. Although the integration of BIM and IoT provides the technical foundation for real-time building sensing and operational analysis, existing applications are generally developed around specific devices, platforms, or business tasks, and cross-system data exchange and continuous reuse remain limited [3]. From the perspective of data processing, Huang et al. identified differences in data volume, technological type, data schema, and level of abstraction as major barriers to the deep integration of BIM and IoT [4].

Building health data can support analyses across spaces, devices, and time periods only when explicitly connected to engineering objects and data acquisition contexts[5]. Semantic Web technologies can represent building objects and their relationships, providing a foundation for data interoperability in the architecture, engineering, and construction (AEC) domain; however, semantic boundaries remain between different models and applications [6]. Brick improves the queryability of building operational data by providing a unified representation of building spaces, equipment, sensing points, and measurement relationships[7]. Regarding the integration of BIM with dynamic monitoring data, Wang et al. extended building models using SensorML, enabling spatial objects, sensor metadata, and time-series observations to be linked within a digital twin environment [8]. Moretti et al. adopted an openBIM approach to integrate static models with IoT data and discussed data access under conditions of incomplete information for existing buildings [9]. These studies have strengthened the semantic representation and object association of building data. Nevertheless, they mainly focus on ontology modeling, digital twins, or asset management, with limited attention to the coding, identification, registration, and authorization processes occurring before and after data access.

Open standards and service-oriented architectures provide another technical pathway for the cross-system application of building data. Dave et al. developed an open-standard-based BIM–IoT integration framework to connect building models, sensing devices, and information services [10]. Eneyew et al. proposed a multilayer data integration architecture for smart building digital twins to improve semantic interoperability between BIM and IoT data [11]. Based on the requirements of IoT data management and real-time analytics, Bashir et al. established a reference architecture for smart buildings [12]. Chamari et al. adopted a service-oriented approach to support building data acquisition, semantic queries, and application invocation, reducing the coupling between data sources and upper-layer services [13]. These studies have addressed some challenges associated with data transmission, system connectivity, and application invocation. However, an integrated specification tailored to building health scenarios is still lacking to define how an individual building health data record should receive a persistent identifier, how its associated project, space, device, time, batch, and source should be registered, and how traceable relationships between raw and derived data should be maintained.

From a data governance perspective, data that are findable, accessible, interoperable, and reusable require explicit identifiers, sufficient metadata, and enforceable access conditions [14]. Building projects typically involve multiple participants engaged in construction, monitoring, inspection, management, and analysis. Therefore, data provenance, responsible parties, and usage boundaries must be retained throughout data circulation. Celik et al. applied blockchain technology to BIM data provenance to record information sources, versions, and participating entities, thereby improving traceability during collaboration [15]. Zheng et al. proposed a context-aware access-control model for BIM, demonstrating that user identity, location, time, and request conditions can affect data access permissions [16]. Practical studies on integrating BIM with IoT sensor data have also shown that data access, field mapping, and continuous updating need to be coordinated within a unified framework [17]. Nevertheless, existing provenance and access-control methods are generally tied to blockchain systems [18], cloud platforms, or specific BIM environments and have not yet formed a continuous governance chain encompassing the coding, registration, association, and provision of building health data.

To address these issues, this study proposes the Building Health Data Architecture (BHDA). BHDA uses information on projects, buildings, floors, spaces, devices, and acquisition times to describe the engineering and spatiotemporal context of data. It preprocesses, standardizes, encodes, and registers metadata for both raw monitoring data and derived data, while generating deterministic unique identifiers from normalized content. On this basis, BHDA records data sources, providers, rights holders, and usage restrictions and determines access according to user identity, authorization scope, declared purpose, and validity period. A reproducible logical validation program is further developed to examine unique identifier generation, contextual association, data registration, authorization decisions, multi-criteria queries, and standardized output. This work provides a methodological foundation for the structured organization, controlled sharing, and subsequent application of building health data.

2. DESIGN OF BHDA

2.1 Overall Architecture Model of BHDA

BHDA is designed to standardize the complete process of multisource building health data, from data access, processing, and registration to authorized use and service output. Its overall structure is shown in Figure 1. The architecture comprises five components: data sources, data governance, data description and registration, the logical data center, and security and data services. Data sources include environmental monitoring data, such as temperature, relative humidity, illuminance, and $PM_{2.5}$, as well as data concerning building spaces, sensing devices, external environments, and derived analyses. The data governance component performs field identification, format conversion, unit standardization, temporal alignment, spatial matching, and quality labeling to resolve differences in naming, structure, and representation among data sources. The data description and registration component assigns standardized codes and unique identifiers to normalized data and registers the engineering context, data source, provider, rights holder, usage restrictions, and storage location. The logical data center comprises raw data, standardized data, derived data, metadata catalogs, object associations, and authorization records, enabling the unified location and organization of distributed data. The security and data services component processes access requests according to user identity, authorization scope, intended purpose, and validity period, and converts query results into standard formats for subsequent building health applications. BHDA follows the principles of unified description, spatiotemporal association, authorized use, process traceability, and technological neutrality. It specifies only the data objects and their governance logic, without restricting the selection of database products or analytical models.

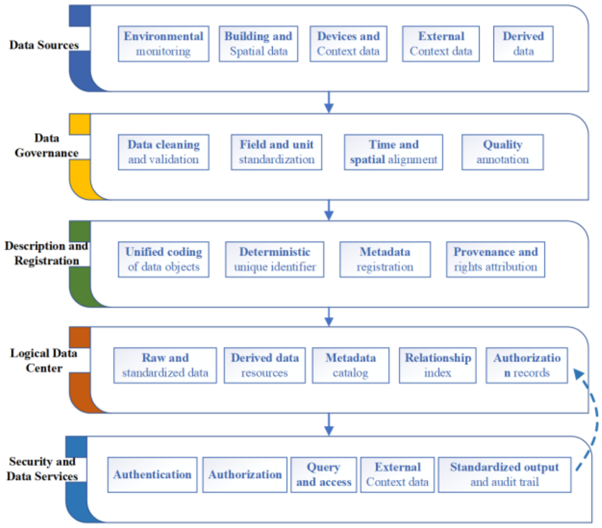

Fig 1. Overall Architecture Model of BHDA

2.2 Data Objects, Unified Coding, and Unique Identifier Generation

The core data objects in BHDA include Project, Building, Floor, Space, Device, and Observation, and their relationships are illustrated in Figure 2. A project serves as the starting point for organizing engineering data and may contain one or more buildings. Each building is further divided into floors and spaces, within which sensing devices are deployed to generate environmental observations at specific times. Observation is the core object connecting engineering context with monitoring data. Each observation is associated with a parameter type, observed value, measurement unit, acquisition time, and monitoring batch, thereby specifying the engineering location, sensing device, and temporal source of the data. A Derived Result generated through cleaning, aggregation, analysis, or evaluation can be linked to the corresponding observation through common engineering and spatiotemporal attributes. In addition, Provider, Rights Holder, User, and Authorization describe the data source, management responsibility, accessing entity, and the corresponding authorization scope, purpose, and validity period, respectively. Together, these objects constitute the governance context required for observation registration and controlled use.

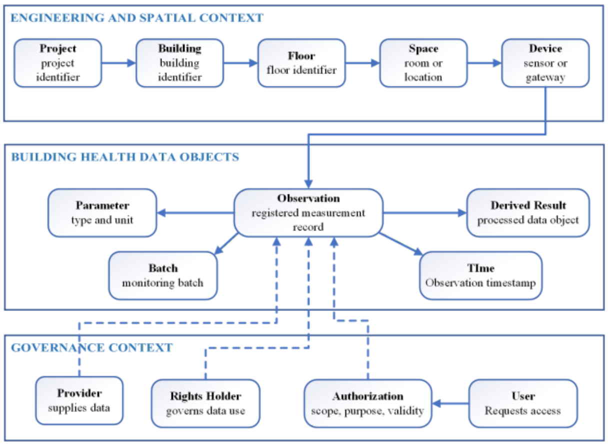

Fig 2. Building Health Data Object Model

To eliminate naming inconsistencies among different data sources, BHDA establishes unified coding rules for projects, buildings, floors, spaces, devices, parameters, and data batches. Each type of object is represented by a fixed prefix, specified length, and unique sequence number, with format normalization performed during data ingestion. For example, the prefixes "P," "B," "F," and "S" are used for projects, buildings, floors, and devices, respectively. The primary coding fields, formats, and examples are presented in Table 1. Unified coding ensures that the same engineering object is represented consistently across different files and systems, thereby providing a foundation for unique identifier generation and cross-source data association.

Table 1. Unified Coding Rules for BHDA Data Objects

Object	Symbol	Coding Example	Description
Project	P	P03	The third engineering project
Building	B	B02	Building No. 2 within the project
Floor	F	F02	The second floor of the building
Space	Sp	LR01	Living Room No. 1
Device	D	S02	Sensor No. 2
Parameter	Pa	PM25	$PM_{2.5}$ parameter
Batch	Ba	B20251105	Data batch dated November 5, 2025

The context of the i-th data record can be expressed as:

$$C_i = (P_i, B_i, F_i, Sp_i, D_i, T_i, Pa_i, Ba_i) \quad (1)$$

Where P_i,B_i,F_i,Sp_i,D_i,T_i,Pa_i, and Ba_i represent the project, building, floor, space, device, acquisition time, parameter, and data batch, respectively.

After coding and time standardization, the system concatenates the data source and contextual fields in a fixed order and generates a unique data identifier using a deterministic hash function:

$$ID_i = H(Src_i \parallel P_i \parallel B_i \parallel F_i \parallel Sp_i \parallel D_i \parallel T_i \parallel Pa_i \parallel Ba_i) \quad (2)$$

Where ID_i denotes the unique data identifier, $H(\cdot)$ is the hash function, Src_i represents the data source type, and $\parallel$ denotes field concatenation. Data records with the same source and context generate the same identifier, whereas changes in the parameter, time, batch, or source produce a different identifier. This method provides a stable index for data registration, retrieval, association, and traceability.

2.3 Data Registration, Rights Attribution, and Service Process

The registration process begins by identifying the data source and data type. Data generated directly by sensing devices are registered as raw monitoring data, whereas data produced through cleaning, aggregation, analysis, or evaluation are registered as derived data. The system then standardizes object codes, time formats, and parameter names and records the measurement unit, monitoring batch, data provider, rights holder, usage restrictions, storage location, and registration time. Raw data may remain in project files, on-site servers, or other storage environments, while the logical data center locates them through unique identifiers and storage references. BHDA therefore supports both centralized storage and the unified organization of distributed data. Derived data are not required to have exactly the same field structure as raw monitoring data; they can be associated with the raw data provided that they contain the necessary engineering and spatiotemporal attributes.

Data registration establishes connections among raw data, the logical data center, and upper-layer applications. Rather than simply writing data into a database, it creates a descriptive record for each data item so that the item can be identified, located, and traced. After preprocessing, unified coding, and unique identifier generation, the registration object for the i-th data record is expressed as:

$$R_i = (ID_i, C_i, V_i, Src_i, Q_i, O_i, A_i, U_i) \quad (3)$$

Where R_i denotes the registration object; ID_i is the unique identifier; C_i represents the engineering and spatiotemporal context; V_i is the observed value or derived data content; Src_i indicates the data source; Q_i represents data quality information; O_i contains provider and rights attribution information; A_i specifies the authorization conditions; and U_i denotes the data storage location or access reference.

For a raw record d_i and a derived record d_j, their association rule is defined as:

$$\text{Link}(d_i, d_j) = \begin{cases} 1, & \text{if } \text{Match}(P, B, F, \text{Sp}, D, T), \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Where $\text{Link}(d_i, d_j) = 1$ indicates that the two records have been successfully associated, whereas $\text{Link}(d_i, d_j) = 0$ indicates that no association has been established. $Match(\cdot)$ indicates that the project, building, floor, space, device, and time satisfy the predefined matching rules. This mechanism enables derived data to be traced to the corresponding raw observations, engineering locations, and acquisition times while preserving the independence of BHDA from specific analytical methods.

Rights attribution is included in the registration information to specify who provided the data, who is responsible for their management, and the purposes for which they may be used. Each record contains at least the data provider, rights holder, source type, usage restrictions, and registration time. For derived data, the generation source and associations with the raw data must also be recorded. Rights attribution in this study is intended to document data provenance, responsibility relationships, and the basis for authorization; it does not constitute a legal determination of data ownership.

When requesting data, a user must submit their identity, query scope, intended purpose, and access time. The system then checks the authorization records to determine whether the user is permitted to access the specified project, building, floor, parameter, time, and data batch. The authorization decision is expressed as:

$$Pe(u,q,t) = Ie(u) \wedge Sc(u,q) \wedge Pu(u,q) \wedge Va(u,t) \quad (5)$$

Where u denotes the user, q represents the query request, and t is the request time. $Ie(u)$ indicates that the user identity is valid; $Sc(u,q)$ indicates that the query scope complies with the authorization; $Pu(u,q)$ indicates that the intended purpose satisfies the specified requirements; and $Va(u,t)$ indicates that the authorization remains valid at the time of the request.

Access is granted only when all four conditions are satisfied; otherwise, the request is denied. Once authorization has been approved, the data service module searches the registration catalog according to conditions such as floor, parameter, time, and batch and uses storage references to locate the target data. The query results are organized into a standard format containing fields such as project, building, floor, space, device, time, parameter, and value. For applications requiring multiple environmental parameters simultaneously, parameter-level data in long format can also be converted into wide format using common spatial and temporal indices.

3. CASE STUDY AND LOGICAL INTEGRITY VALIDATION

To verify whether BHDA can perform data identification, registration, association, authorization, and service output in accordance with the methods described in Section 2, a reproducible Python validation program was developed using actual building environmental monitoring data. The validation focuses on the logical integrity of the architecture and does not evaluate the storage performance, query efficiency, or concurrent processing capability of a specific database.

3.1 Case Data and Validation Environment

The case data were collected from an environmental monitoring node located in the living room on the second floor of a residential building. Monitoring was conducted from 10:30 to 11:29 on November 5, 2025, at a sampling interval of 1 min, producing 60 raw records. Each record contains contextual information, including the project, building, floor, space, sensor, and acquisition time, together with four environmental parameters: illuminance, temperature, relative humidity, and $PM_{2.5}$.

To verify the capability of BHDA to ingest different types of data, the case also incorporates 60 derived records generated by an external analysis program. Although the derived data have a field structure different from that of the raw monitoring data, they contain common attributes, including project, building, space, device, and time, which enable the contextual association between raw and derived data to be examined. In this study, these records are treated solely as an external data source; their generation method and result accuracy are not discussed.

The validation program was implemented in Python within Jupyter Notebook. Pandas was primarily used for data reading, transformation, association, and querying, while the SHA-256 hash function was applied to generate deterministic unique identifiers. The program supports both CSV and XLSX files and automatically produces a data registration catalog, authorization records, standardized query results, and a logical integrity report. The case data and validation environment are summarized in Table 2.

Table 2. Case Data and Validation Environment

Item	Raw Monitoring Data	Derived Data
Number of records	60	60
Time range	10:30–11:29	10:30–11:29
Time interval	1 min	1 min
Main fields	L, T, RH, and $PM_{2.5}$	Derived states and related results
Contextual fields	Project, building, floor, space, device, and time	Project, building, space, device, and time
Data purpose	Environmental parameter registration and querying	Validation of multi-type data ingestion and association
Validation tools	Python 3.10, Jupyter Notebook, Pandas, and SHA-256	Same as the left column

3.2 Data Preprocessing, Coding, and Registration

Each row of the raw file contains four parameters: illuminance, temperature, relative humidity, and $PM_{2.5}$. To allow each parameter to be registered, queried, and authorized independently, the program first converts the data from wide format into parameter-level long format. Consequently, the 60 raw records are transformed into 240 standardized environmental parameter records. Each record corresponds to a single parameter value while retaining information such as the project, building, floor, space, device, acquisition time, and batch.

During format conversion, the program standardizes field names, time formats, and object codes. For example, P3 and P03 from different data sources are uniformly represented as P03, while B2, F2, and S2 are normalized as B02, F02, and S02, respectively. This process prevents the same engineering object from being identified as multiple objects because of differences in character length or naming conventions. When the floor field is not directly provided in the derived data, the program supplements the floor attribute according to the registered case context and file information. However, the supplementation rule must be explicitly documented in the registration information.

After coding standardization, the program generates a unique identifier based on the data source, project, building, floor, space, device, acquisition time, parameter, and batch. Identifier uniqueness is verified using the following condition:

$$N_{\mathrm{ID}} = N_{\mathrm{record}} \quad \wedge \quad N_{\mathrm{duplicate}} = 0 \quad (6)$$

The validation results show that 240 unique identifiers were generated for the 240 parameter-level environmental data records, while 60 unique identifiers were generated for the 60 derived data records, with no duplicate identifiers detected. Together, the two data types produced 300 registered objects. Each registration record contains the data identifier, engineering and spatiotemporal context, parameter type, data batch, source type, provider, rights holder, usage restrictions, storage reference, and registration time. The raw data may remain in their original files, while the registration catalog locates them through unique identifiers and storage references. This verifies that BHDA can organize distributed data without changing their original storage arrangements. The complete case-processing workflow, including data reading, preprocessing, coding, unique identifier generation, registration, association, authorization, and service output, is illustrated in Figure 3.

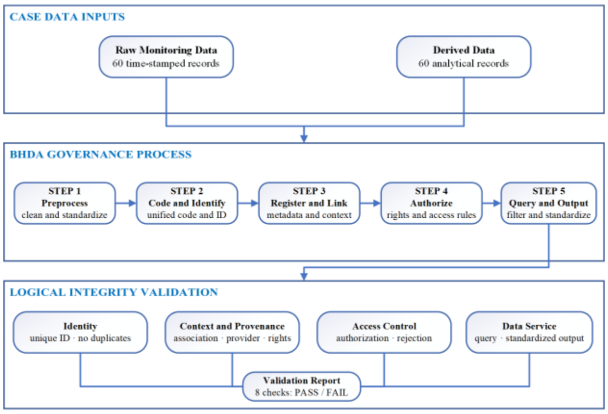

Figure 3. BHDA Case-Based Data Governance and Logical Validation Process

3.3 Data Association and Rights Attribution Registration

BHDA uses project, building, floor, space, device, and acquisition time as the common context for data association. For any raw record d_i and derived record d_j, the system establishes an association when their core contextual information is consistent:

$$\text{Link}(d_i, d_j) = \text{Match}(P, B, F, Sp, D, T) \tag{7}$$

Where P, B, F, Sp, D, and T represent the project, building, floor, space, device, and acquisition time, respectively. This method does not require the two data types to have identical parameters or field structures; they need only provide the necessary contextual information.

The association results show that all 60 derived records were successfully matched with the corresponding raw environmental data through their common context. The association rate in the case study was calculated as:

$$R_{link} = \frac{N_{linked}}{N_{derived}} \times 100\% = \frac{60}{60} \times 100\% = 100\% \tag{8}$$

During data registration, the data provider, rights holder, source type, usage restrictions, and registration time were recorded for both data types. The provider of the raw monitoring data was registered as the monitoring data provider, whereas the provider of the derived data was registered as the external analytical data provider. The inspection results show that all 300 registered objects contained complete provider and rights attribution information.

3.4 Authorization and Data Query Validation

To verify the controlled-access capability of BHDA, the program defined both authorized and unauthorized users. Each authorization record contains the user identity, project, building, permitted floors, parameters, data batches, intended purpose, and authorization start and end times. After a user submits a query request, the system sequentially checks the user's identity, access scope, declared purpose, and authorization validity period.

First, an unregistered user attempted to access temperature data from the second floor. Because the system found no valid authorization record matching the user's identity and intended purpose, the request was denied. This result demonstrates that possession of the data query conditions alone does not grant access; the user must hold authorization corresponding to the requested scope.

The program then used an authorized user to query data under the following conditions: project, P03; building, B02; floor, F02; parameter, temperature; time, 10:30–11:29 on November 5, 2025; data batch, raw monitoring batch; and purpose, building health research. After verifying the user's identity and authorization conditions, the system located the corresponding data through the registration catalog and returned 60 temperature records that satisfied the query. The project, building, floor, parameter, time range, and data batch in the returned results were consistent with the request, demonstrating that BHDA supports multi-criteria data filtering and authorized access.

To examine its data service capability, the program separately queried illuminance, temperature, relative humidity, and $PM_{2.5}$ data and combined them using common engineering, spatial, device, and temporal indices. The parameter-level records in long format were subsequently converted into standardized wide-format data containing 60 time points. The output fields include the project, building, floor, space, device, acquisition time, and four environmental parameters and can be used directly as input for subsequent building health analysis programs. Eight validation items were established across five aspects: data identity, object association, rights attribution records, authorized access, and service output. The validation conditions and results are presented in Table 3.

Table 3. Logical Integrity Validation Conditions and Results

No.	Validation Item	Acceptance Criterion	Result
1	Unique data identification	The number of identifiers equals the number of records, with no duplicate identifiers	PASS
2	Contextual integrity	Project, building, floor, space, device, and time information are complete	PASS
3	Provider and rights attribution registration	No required source or rights attribution fields are missing	PASS
4	Derived data ingestion and association	Derived data are registered and matched with the context of the corresponding raw data	PASS
5	Authorization scope registration	User, scope, purpose, and validity-period information are complete	PASS
6	Unauthorized user rejection	Access is denied when no valid authorization	PASS

		exists	
7	Authorized conditional query	Correct data are returned according to floor, parameter, time, and batch	PASS
8	Standardized service output	Multi-parameter data suitable for subsequent applications are generated	PASS

3.5 Comprehensive Evaluation and Visualization of Results

To further evaluate the logical integrity of BHDA during case-study data processing, this section summarizes the validation results obtained in Sections 3.2–3.4 and compares the expected and observed results in terms of data registration, unique identification, rights attribution, data association, access authorization, conditional querying, and standardized output. This comprehensive evaluation provides a unified summary and visual presentation of the preceding case validation results.

As shown in Table 4, eight items were evaluated to validate logical integrity. Regarding data registration, all 300 data objects in the case study were successfully registered and assigned 300 unique identifiers, with no duplicate identifiers detected. Regarding rights attribution, all 300 registered data objects contained the corresponding rights-attribution information, resulting in a rights-record completeness rate of 100%. These results indicate that BHDA maintained consistent correspondence among data objects, unique identifiers, and rights-attribution information under the defined coding and registration rules.

Table 4. Comprehensive Evaluation Results of BHDA Logical Integrity

Evaluation Item	Expected Result	Observed Result	Evaluation
Registered data objects	300	300	PASS
Unique identifiers	300	300	PASS
Duplicate identifiers	0	0	PASS
Complete rights records	300	300	PASS
Linked derived records	60	60	PASS
Unauthorized query	Rejected	Rejected	PASS
Records returned by authorized query	60	60	PASS
Standardized output time points	60	60	PASS

Regarding data association, all 60 derived records were successfully linked to their corresponding source observations, resulting in an association rate of 100%. This result indicates that BHDA can establish relationships between derived results and original observation data based on shared engineering, spatial, and temporal attributes. Regarding access control, the unauthorized query was correctly rejected, demonstrating that the defined authorization rules could restrict requests that did not satisfy the access conditions. When the authorization conditions were satisfied, the authorized query returned the expected 60 records and generated 60 standardized multi-parameter time points. Therefore, within the authorization scope and query conditions defined in this case study, BHDA successfully implemented a complete processing workflow encompassing data registration, rights attribution, data association, controlled querying, and standardized output.

To present the quantitative evaluation results more intuitively, Figure 4 compares the expected and observed values of six quantitative indicators. Figure 4(a) presents the data registration and rights-attribution results, including registered data objects, unique identifiers, and complete rights records. Figure 4(b) presents the data association, authorized querying, and standardized output results, including linked derived records, records returned by the authorized query, and standardized output time points. Duplicate identifiers and the unauthorized query represent a zero-value indicator and a categorical result, respectively; therefore, they are not included in the bar charts but are retained in Table 4 for evaluation.

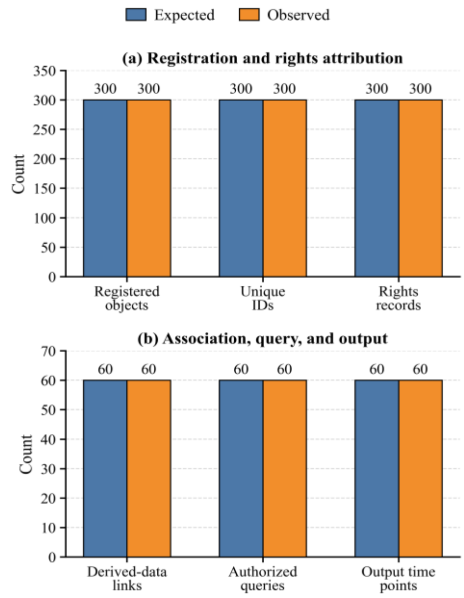

Figure 4. Comparison of Expected and Observed Results for BHDA Logical Integrity Validation

As shown in Figure 4, the observed values of all six quantitative indicators were consistent with their expected values. The expected and observed values were both 300 for data registration, unique identification, and rights attribution, whereas they were both 60 for derived-data association, authorized querying, and standardized output. Together with the results presented in Table 4—zero duplicate identifiers and the correct rejection of the unauthorized query—all eight evaluation items satisfied the predefined logical validation requirements. These findings demonstrate that, within the scope of this case study, BHDA maintained logical consistency across data-object coding, unique-identifier generation, rights attribution, data association, authorization determination, data querying, and result output.

4. CONCLUSIONS AND FUTURE WORK

This study proposed BHDA to address the dispersed sources, inconsistent representations, insufficient engineering context associations, and non-uniform sharing and authorization mechanisms of building health data. Through data preprocessing, unified coding, deterministic unique identification, and metadata registration, BHDA associates raw monitoring data and derived data with the corresponding project, building, floor, space, device, and acquisition time. It also incorporates data providers, rights holders, usage restrictions, and authorization rules into the data service process.

Reproducible validation using actual building environmental data showed that 60 raw records were transformed into 240 parameter-level environmental data records, which, together with 60 derived records, formed 300 registered objects. Unique identifiers were successfully generated for all objects, and the contextual association rate for the derived data reached 100%. The system successfully rejected unauthorized requests and enabled authorized users to query data by floor, parameter, time, and batch. All eight logical integrity checks were passed. These results demonstrate that BHDA establishes a continuous governance process encompassing data coding, registration, association, rights attribution, authorization decisions, and standardized output, thereby providing a data foundation for the unified organization, controlled sharing, and subsequent analysis of building health data.

The present validation focused primarily on the logical integrity of BHDA and did not evaluate its operational performance under large-scale data processing, concurrent access, or cross-organizational collaboration. Future research will implement a BHDA prototype using appropriate database and service technologies and conduct tests on data scalability, query efficiency, concurrent access, and cross-system interoperability. The applicability and scalability of the coding rules, data association methods, and authorization mechanisms will also be further examined across different building projects and data sources.

5. REFERENCES

[1] K. P. Seng, L. M. Ang, and E. Ngharamike, "Artificial intelligence Internet of Things: A new paradigm of distributed sensor networks," *Int. J. Distrib. Sens. Netw.*, vol. 18, no. 3, p. 15501477211062835, Mar. 2022, doi: 10.1177/15501477211062835.

[2] D. Andritsou, C. Alexiou, and C. Potsiou, "BIM, 3D Cadastral Data and AI for Weather Conditions Simulation and Energy Consumption Monitoring," *Land*, vol. 13, no. 6, p. 880, Jun. 2024, doi: 10.3390/land13060880.

[3] S. Tang, D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, and X. Gao, "A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends," *Autom. Constr.*, vol. 101, pp. 127–139, May 2019, doi: 10.1016/j.autcon.2019.01.020.

[4] X. Huang, Y. Liu, L. Huang, E. Onstein, and C. Merschbrock, "BIM and IoT data fusion: The data process model perspective," *Autom. Constr.*, vol. 149, p. 104792, May 2023, doi: 10.1016/j.autcon.2023.104792.

[5] S. Dong, H. Li, and Q. Yin, "Building information modeling in combination with real time location systems and sensors for safety performance enhancement," *Saf. Sci.*, vol. 102, pp. 226–237, Feb. 2018, doi: 10.1016/j.ssci.2017.10.011.

[6] P. Pauwels, S. Zhang, and Y.-C. Lee, "Semantic web technologies in AEC industry: A literature overview," *Autom. Constr.*, vol. 73, pp. 145–165, Jan. 2017, doi: 10.1016/j.autcon.2016.10.003.

[7] B. Balaji *et al.*, "Brick: Metadata schema for portable smart building applications," *Appl. Energy*, vol. 226, pp. 1273–1292, Sep. 2018, doi: 10.1016/j.apenergy.2018.02.091.

[8] T. Wang, V. J. L. Gan, D. Hu, and H. Liu, "Digital twin-enabled built environment sensing and monitoring through semantic enrichment of BIM with SensorML," *Autom. Constr.*, vol. 144, p. 104625, Dec. 2022, doi: 10.1016/j.autcon.2022.104625.

[9] N. Moretti, X. Xie, J. Merino, J. Brazauskas, and A. K. Parlikad, "An openBIM Approach to IoT Integration with Incomplete As-Built Data," *Appl. Sci.*, vol. 10, no. 22, p. 8287, Nov. 2020, doi: 10.3390/app10228287.

[10] B. Dave, A. Buda, A. Nurminen, and K. Främling, "A framework for integrating BIM and IoT through open standards," *Autom. Constr.*, vol. 95, pp. 35–45, Nov. 2018, doi: 10.1016/j.autcon.2018.07.022.

[11] D. D. Eneyew, M. A. M. Capretz, and G. T. Bitsuamlak, "Toward Smart-Building Digital Twins: BIM and IoT Data Integration," *IEEE Access*, vol. 10, pp. 130487–130506, 2022, doi: 10.1109/ACCESS.2022.3229370.

[12] M. R. Bashir, A. Q. Gill, and G. Beydoun, "A Reference Architecture for IoT-Enabled Smart Buildings," *SN Comput. Sci.*, vol. 3, no. 6, p. 493, Sep. 2022, doi: 10.1007/s42979-022-01401-9.

[13] L. Chamari, E. Petrova, and P. Pauwels, "An end-to-end implementation of a service-oriented architecture for data-driven smart buildings," IEEE Access, vol. 11, pp. 117261–117281, 2023, doi: 10.1109/ACCESS.2023.3325767./

[14] H. M. Wong and S. F. Norbaini, "The Data Governance: A Comprehensive Literature Review from Professional Viewpoints," in *Lect. Notes Networks Syst.*, Springer Science and Business Media Deutschland GmbH, 2025, pp. 491–510. doi: 10.1007/978-3-031-96380-3_42.

[15] Y. Celik, I. Petri, and M. Barati, "Blockchain supported BIM data provenance for construction projects," *Comput. Ind.*, vol. 144, p. 103768, Jan. 2023, doi: 10.1016/j.compind.2022.103768.

[16] R. Zheng, J. Jiang, X. Hao, W. Ren, F. Xiong, and T. Zhu, "CaACBIM: A Context-aware Access Control Model for BIM," *Information*, vol. 10, no. 2, p. 47, Feb. 2019, doi: 10.3390/info10020047.

[17] G. Desogus, E. Quaquero, G. Rubiu, G. Gatto, and C. Perra, "BIM and IoT Sensors Integration: A Framework for Consumption and Indoor Conditions Data Monitoring of Existing Buildings," *Sustainability*, vol. 13, no. 8, p. 4496, Apr. 2021, doi: 10.3390/su13084496.

[18] Ö. Karaduman and G. Gülhas, "Blockchain-Enabled Supply Chain Management: A Review of Security, Traceability, and Data Integrity Amid the Evolving Systemic Demand," *Appl. Sci.*, vol. 15, no. 9, p. 5168, May. 2025, doi: 10.3390/app15095168.