

# From Digital Field Survey to Geospatial Decision Support: A Nationwide Data-Quality and GIS-Integration Framework for Saudi Public-School Facilities

---

Mustafa Yousef AbuMarzouq<sup>1,3\*</sup>, Antar A. Aboukorin<sup>1</sup>, Umar Lawal Dano<sup>1</sup>, Fahad S. Almutlaq<sup>1</sup>, Abdulaziz Altuwalah<sup>2</sup>, and Rabea Damlakhy<sup>3</sup>

<sup>1</sup> Department of Urban and Regional Planning, College of Architecture and Planning, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia

<sup>2</sup> Tatweer Buildings Company (TBC), Riyadh, Saudi Arabia

<sup>3</sup> Ayen Company, Riyadh, Saudi Arabia

\* Corresponding author: Mustafa Yousef AbuMarzouq (email to be supplied before submission)

[dr.mustafa.abumarzouq@gmail.com](mailto:dr.mustafa.abumarzouq@gmail.com)

[aaaboukorin@iau.edu.sa](mailto:aaaboukorin@iau.edu.sa)

[uldano@iau.edu.sa](mailto:uldano@iau.edu.sa)

[phdfahad2024@gmail.com](mailto:phdfahad2024@gmail.com)

[Ske770@gmail.com](mailto:Ske770@gmail.com)

[rabe3da@gmail.com](mailto:rabe3da@gmail.com)

## Abstract

This study aimed to assess the quality and GIS readiness of a nationwide Saudi public-school facility dataset and identify quality-assurance controls required to support reliable GIS-based decision-making. A cross-sectional secondary data-quality audit was conducted using 17,008 records and 44 variables. The study population comprised 13,048 eligible public elementary, intermediate, and secondary school-facility assessment records covering all 13 Saudi administrative regions and 149 governorates. Data were analyzed using deterministic counts and percentages to assess identifier completeness and conformance, field completeness, administrative consistency, coordinate plausibility, thematic plausibility, and data lineage. The findings showed that all 13,048 eligible records had complete and unique Unit Codes, inspection-report identifiers, and primary ministerial numbers. Moreover, 13,045 records contained building-quality values, and 13,042 records (99.954%) met the predefined core GIS-readiness criteria. However, important data gaps were identified: evacuation status and sprinkler availability were complete for only 4.139% and 20.317% of records, respectively, while inspection dates, inspector identifiers, instrument versions, and edit/quality-assurance timestamps were unavailable. The study therefore concludes that the dataset demonstrates high structural readiness for linkage and broad GIS representation but does not, by itself, establish positional or thematic accuracy. It recommends implementing version-controlled quality assurance, persistent asset identifiers, complete inspection lineage, explicit missing-data classifications, documented exception resolution, and traceable versioned GIS releases to strengthen the dataset's reliability and long-term decision-support value.

**Keywords.** digital field survey; geospatial data quality; GIS integration; school facilities; quality assurance; asset management; decision support; Saudi Arabia.

## 1. Introduction

Public infrastructure decisions depend on reliable records that link physical assets, field observations, locations, and responsible actions. Weaknesses in any of these links may result in unreliable maps, duplicated work, inefficient resource allocation, or unsupported assessments of facility condition. This challenge is particularly important for school facilities because they are geographically dispersed, serve different educational stages, and require decisions related to maintenance, safety, and long-term planning. Therefore, reliable and traceable evidence is essential at both facility and aggregate levels.

Digital field-data collection can strengthen this evidence base by capturing structured identifiers, standardized fields, geographic coordinates, and consistent survey information. Mobile systems can also support validation rules, required fields, and rapid transfer to centralized databases. Previous research has reported potential benefits of digital data collection for availability, efficiency, and error control, although these benefits depend on system design and implementation (Hartung et al., 2010; Thriemer et al., 2012; Taber et al., 2020). Importantly, digital collection does not automatically guarantee data accuracy. Digitally generated records may still contain incorrect coordinates, inconsistent administrative classifications, incomplete fields, or undocumented scoring procedures. Data quality should therefore be understood as multidimensional and dependent on the intended use of the information (Wang & Strong, 1996; Batini & Scannapieco, 2016).

Geospatial data require additional quality dimensions, including completeness, logical consistency, positional quality, thematic quality, temporal quality, and lineage (International Organization for Standardization, 2023). Thus, a dataset may be adequate for broad mapping while remaining insufficient for facility-level operational decisions. Similarly, coordinates that fall within the geographical boundaries of Saudi Arabia demonstrate only general plausibility and do not establish positional accuracy. Transparent documentation of record selection, data processing, missing information, and transformations is also essential when administrative data are reused for research and decision-making (Benchimol et al., 2015; Kahn et al., 2016; Wilkinson et al., 2016).

In school-facility management, reliable asset identification and traceable condition information are particularly important. Recent research emphasizes the need for transparent building-condition assessment models and systematic links between condition assessments and asset databases (Begić & Krstić, 2024; Begić Juričić & Krstić, 2025). Saudi national guidance similarly emphasizes asset identification, standardized assessment procedures, traceable field evidence, and linked records (Government Expenditure & Projects Efficiency Authority, 2022). The integration of GIS and building information can further support facility management, but its reliability depends on the quality of the underlying identifiers, spatial information, and data lineage (Congiu et al., 2024).

Saudi Arabia provides an important national context for examining these issues because the present program covers public elementary, intermediate, and secondary facilities across all 13 administrative regions and 149 governorates. Although previous GIS research has demonstrated the value of school-location data for distribution and accessibility analysis (Abulibdeh et al., 2024), and geospatial technologies can support evidence-based governance (Dano et al., 2020), limited methodological attention has been given to transparent national audits of the quality and decision-readiness of large school-facility datasets. In particular, there is a need to establish how records are retained, whether

identifiers are stable, how spatial and administrative inconsistencies are handled, which fields remain incomplete, and whether sufficient lineage exists for subsequent GIS applications.

Accordingly, this study conducts a methods-oriented audit of a nationwide Saudi public-school facility dataset. It evaluates identifier completeness and uniqueness, field completeness and conformance, administrative consistency, coordinate plausibility, thematic plausibility, and data lineage, and identifies the quality-assurance and governance requirements needed to transform the dataset into a traceable GIS and decision-support resource. The study focuses specifically on the evidence infrastructure underlying future school-facility analyses, rather than estimating school accessibility, population pressure, Quality of Life, regional condition outcomes, or regulatory compliance.

### **Study significance**

- The study distinguishes data completeness and plausibility from positional accuracy, thematic validity, and longitudinal traceability.
- Its national coverage provides a basis for assessing whether a large administrative and field-survey extract is sufficiently structured and transparent for GIS integration and decision support.
- This study is important because national school-facility decisions depend on reliable records linking physical assets, observations, locations, and responsible actions.

### **Study objectives**

- To propose a traceable quality-assurance and governance architecture linking field capture, exception resolution, GIS publication, and decision use.
- To identify missing metadata that constrain validation, longitudinal analysis, and accountability.
- To evaluate administrative consistency, coordinate plausibility, thematic plausibility, and field-level completeness.
- To assess the completeness, uniqueness, and conformance of identifiers in the national public-school facility dataset.

### **Study Hypotheses**

- Because the study is descriptive and diagnostic and applies no inferential hypothesis test, these are analytical expectations rather than statistically tested causal hypotheses.
- H2: Despite high structural readiness, substantial field-specific completeness and lineage deficiencies remain in selected decision-relevant variables.
- H1: The national school-facility dataset demonstrates high structural readiness for record linkage and broad GIS integration.

### **Study Problem**

The research problem is that a digitally structured national dataset may appear complete while still containing unresolved spatial, administrative, thematic, and lineage weaknesses that can affect downstream GIS and asset-management decisions.

### Study Questions

- What quality-assurance and governance architecture can convert the national field-survey extract into a traceable GIS and decision-support resource?
- Which missing metadata limit validation, longitudinal use, and accountability?
- To what extent are the records complete, numerically conformant, geographically plausible, and administratively consistent?
- What proportion of eligible records contain complete and unique identifiers required for linkage across systems?

### Study Population

- The records cover all 13 administrative regions and 149 governorates. The analytical unit is the facility-assessment record identified by Unit Code rather than the literal school name or an inferred count of unique physical buildings.
- The study population comprises the 13,048 eligible public elementary, intermediate, and secondary facility-assessment records retained after the prespecified scope and school-stage criteria were applied.

### Study Delimitations

- Subject delimitation: data quality, GIS readiness, administrative and spatial consistency, field completeness, metadata lineage, and governance requirements for school-facility decision support. The study does not estimate accessibility, population-to-school ratios, regional quality rankings, or Quality-of-Life outcomes.
- Spatial delimitation: public elementary, intermediate, and secondary school-facility assessments across the 13 administrative regions and 149 governorates of Saudi Arabia.
- Temporal delimitation: the analysis is limited to the nationwide data extract available for the present audit (Draft R0, 8 August 2026). Because inspection dates were unavailable, the field-collection period cannot be established from the research data.

### Previous Studies

Dano et al. (2020) showed how geospatial technologies can connect spatial evidence with urban governance and decision-making. The present study extends this governance perspective by focusing specifically on the quality-assurance controls needed to make national school-facility evidence sufficiently traceable for GIS-based decision support.

Congiu et al. (2024) proposed an integrated Building Information Modeling and GIS framework for facility management. Their study emphasizes linking building information with geographic context. The present study addresses a prerequisite for such integration by examining identifiers, coordinates, administrative relationships, missingness, and data lineage in a national school-facility extract.

Begić Juričić and Krstić (2025) developed a framework for school building-condition assessment in Osijek-Baranja County. Their work highlights the value of systematic condition assessment linked to asset information. The present

study complements this approach by auditing the data structure, completeness, spatial plausibility, and lineage needed to support school-facility asset management at national scale.

Abulibdeh et al. (2024) examined the spatial distribution and accessibility of public and private schools in Qatar using GIS. The study demonstrates the value of geospatial analysis for school distribution and accessibility. The present study is related through its use of school-location evidence, but focuses on data quality, record readiness, and governance before downstream spatial decisions.

## **2. Conceptual framework**

### **2.1 Digital capture as a control environment—not a validity certificate**

A digital field system can place quality controls close to the moment of observation. Stable task identifiers can prevent unlinked forms; schema constraints can reject impossible types; conditional rules can require supporting fields; location capture can support routing and plausibility review; and synchronized records can make exceptions visible before a campaign ends. These features are control opportunities. Their actual effect depends on the configured instrument, training, supervision, device behaviour, offline workflow, correction permissions, and quality-review process. Because the research extract did not contain application logs or the instrument configuration, this study infers only that the resulting records are digitally structured; it does not reconstruct or certify the historical application workflow.

This distinction also avoids an unsupported technology-effect claim. The source material contains no randomized or parallel paper comparison, no time-on-task logs, no error counts before and after digitization, and no cost ledger. Accordingly, the audit does not estimate time saved, cost reduced, or accuracy improved by digital collection. Evidence from other settings shows that electronic collection may reduce delayed entry and some error classes, but also that per-observation productivity and implementation costs vary with context (Thriemer et al., 2012; Taber et al., 2020). The appropriate conclusion for the Saudi programme must be derived from the structure and exceptions visible in its own extract.

### **2.2 A multidimensional data-quality profile**

The audit combines geospatial quality concepts from ISO 19157-1 with the practical conformance–completeness–plausibility logic of Kahn et al. (2016). The latter originated in health data and is not treated as a school-facility standard; it supplies a compact secondary-use taxonomy. Six dimensions are evaluated:

- Identity and conformance: whether stable identifiers are populated and unique, whether required numeric fields can be parsed as numeric, and whether exact duplicate rows exist.
- Completeness: the proportion of eligible records containing a value for each decision-relevant field, reported field by field rather than averaged.
- Logical and administrative consistency: whether territorial labels form a consistent region–governorate hierarchy and whether repeated labels require review.
- Spatial plausibility: whether numeric coordinate pairs are complete, globally valid, and located within a broad national envelope; repeated coordinate pairs are flagged for review.

- Thematic plausibility: whether selected numeric values fall within deliberately broad screening ranges that identify likely entry or transformation errors without defining regulatory acceptability.
- Temporal quality and lineage: whether the record retains the inspection date, inspector or team identifier, instrument version, and edit or QA timestamps needed to reconstruct how an observation was produced and changed.

No composite data-quality index is produced. Composite scores would require defensible weights and compensation rules: complete coordinates could offset missing life-safety data, for example, even though the two deficiencies have different consequences. A dimensional profile preserves the link between each finding, the decision it affects, and the corrective action required. It also reflects ISO 19157-1, which defines quality components and evaluation procedures but does not prescribe a universal minimum threshold for all uses.

### **2.3 From a field record to a governed decision product**

A decision-support architecture must separate the source record from later interpretations. The source layer should remain immutable and retain the values received from the field programme. A curated release may standardize encodings, resolve confirmed administrative assignments, and derive analytical variables, but every transformation should be scripted, versioned, and accompanied by an issue register. The GIS layer should then reference the curated release and its metadata rather than a manually edited copy. This structure makes corrections possible without erasing what was originally recorded.

The architecture must also distinguish entities that spreadsheets often collapse. A physical building is an asset; a school is an organizational service entity; an inspection report is an event; and an observed condition item belongs to that event. One building may host multiple schools, and one building may be inspected repeatedly over time. A durable relational model therefore requires a persistent asset identifier, a school identifier, a dated occupancy relationship, an inspection-event identifier, and item-level observations linked to the applicable instrument version. This prevents a school name from serving as a surrogate building key and makes historical comparisons interpretable.

Finally, decision products should be purpose-specific. Facility-level coordinates and exception details can remain in a restricted operational view for authorized asset and field teams. Aggregate coverage, completeness, or planning indicators can be published at region or governorate level where appropriate. The same governed data release can support both products, but permissions, disclosure risk, and decision thresholds differ. A closed loop records whether an exception was verified, corrected, accepted as legitimate, or escalated to reinspection, then publishes a new traceable release rather than silently modifying the old one.

## **3. Materials and methods**

### **3.1 Study design and reporting approach**

We conducted a cross-sectional secondary data-quality study of a national administrative and facility field-survey extract. The design is descriptive and diagnostic: it quantifies record flow, uniqueness, completeness, type conformance, logical conflicts, coordinate plausibility, thematic plausibility, and lineage availability. It does not test associations between facility characteristics and condition outcomes. Reporting was informed by ISO 19157-1 and adapted transparency elements from RECORD, particularly explicit source selection, cleaning, linkage, limitations, and access

conditions (Benchimol et al., 2015; International Organization for Standardization, 2023). Because RECORD is health specific, the manuscript does not claim formal compliance with that guideline.

### **3.2 Setting, scope, and unit of observation**

The study setting is the Kingdom of Saudi Arabia. The source workbook contained administrative school attributes, geographical identifiers, coordinates, facility characteristics, and structured condition and safety fields. The study scope was defined before the present audit by the source variable Scope and the school-stage field. Records coded In and classified as elementary, intermediate, or secondary were retained. Records coded Out and one In record without a target-stage classification were excluded from the eligible set.

The analytical unit is the facility-assessment record identified by a Unit Code, not the literal school name and not an inferred count of unique physical buildings. This terminology is necessary because the extract includes a variable recording how many schools operate in the building. A multi-school record represents a shared or co-located operating arrangement as captured by the source; it does not authorize deduplication of records on the basis of name or coordinate alone. All national and governorate counts in this paper refer to eligible facility assessments unless stated otherwise.

### **3.3 Data provenance and institutional roles**

The nationwide school-facilities dataset used in this study is owned by Tatweer Buildings Company (TBC), which authorized its use for academic research and facilitated access to documentation needed for responsible interpretation. Ayen Company acted as principal consultant for the field-survey programme, financing and managing expert field teams responsible for nationwide data collection, verification, analysis, and GIS representation on behalf of TBC. Institutional project documentation supplied with the research materials corroborated this provenance and national operating scale; it was not used to infer undocumented application functions or record-level validation steps.

The present research audit was performed on the extract as received. The dataset owner, commissioning context, and field-programme consultant are reported so readers can distinguish data custody, programme delivery, and academic analysis. Neither institutional involvement nor digital structure is treated as a substitute for transparent checks. Findings that may represent errors are therefore described as review flags unless the extract itself provides an authoritative basis for correction.

### **3.4 Source variables and prespecified audit domains**

The source sheet contained 17,008 rows and 44 columns. The audit emphasized variables required to connect a facility assessment to administrative, geographical, physical-asset, and condition information. Governance and classification fields included authority, scope, region, governorate, stage, sex, and urban–rural classification. Identity and linkage fields included Unit Code, inspection-report identifier, primary ministerial number, school name, and Plus Code. Geospatial fields included longitude, latitude, Plus Code, city, governorate, and location text. Asset fields included building age, establishment year, schools in the building, built area, floors, and classrooms. Condition and safety fields included the supplied building-quality percentage and component domains such as electrical/mechanical condition, health and safety, fire extinguishers, emergency exits, sprinklers, fire pumps, perimeter fence, floors, evacuation status, and external areas. Supplementary Table S1 reports field-level completeness for the audited variables.

### 3.5 Record selection and normalization

Whitespace was trimmed for key text comparisons. Arabic alef variants and tatweel were normalized only to classify school stages consistently; the original source values were retained. Eligible records satisfied both Scope = In and a normalized stage of elementary, intermediate, or secondary. The primary scored subset additionally required a nonmissing numeric value in the supplied building-quality field. This score was not recalculated because the full instrument, formula, weights, and version were not present in the research extract.

Identifier uniqueness was evaluated after trimming repeated whitespace. School-name uniqueness was also examined after simple Arabic orthographic normalization to demonstrate why a name is unsuitable as a relational key. No fuzzy name matching or spatial deduplication was used, because similar names and shared coordinates can be legitimate. Exact duplicate rows were counted across all source columns in the eligible subset.

### 3.6 Audit rules

For field  $j$ , completeness was calculated as  $C_j = 100 \times n_j/N$ , where  $n_j$  is the number of eligible records containing a nonmissing source value and  $N = 13,048$ . Numeric conformance was evaluated separately by coercing selected fields to numeric values and counting nonmissing source entries that could not be parsed. This separation prevents a nonempty text string from being mistaken for an analytically usable number.

- Core readiness: a record required a nonmissing Unit Code and inspection-report identifier, a numeric coordinate pair within the broad national envelope, nonmissing study region and governorate, and a nonmissing numeric building-quality value.
- Coordinate completeness and plausibility: longitude and latitude were parsed independently. Complete pairs were checked against global numeric limits and a deliberately broad Saudi screening envelope of 34°–56° east and 15°–33° north. The envelope is not an official boundary and does not measure positional accuracy. Repeated exact numeric pairs were flagged but not removed.
- Administrative consistency: the number of region labels associated with each governorate name was counted. For diagnostic purposes only, the modal region within each governorate was compared with the observed label. Nonmodal assignments were flagged for authoritative review; the mode was not silently imposed on the source.
- Thematic plausibility: exploratory ranges were 0–100 for the supplied percentage score, 0–100 years for building age, 50–20,000 m<sup>2</sup> for built area, 1–10 floors, 1–100 classrooms, and 1–10 schools per building. These intentionally broad ranges identify probable entry or transformation anomalies; they are not engineering, code-compliance, or investment thresholds.
- Lineage: the schema was searched for inspection date, inspector or team identifier, instrument version, and edit or QA timestamps. Their absence was recorded as a limitation rather than treated as missing record-level values.

### 3.7 Analysis, visualization, and reproducibility

The audit used deterministic counts and percentages. No sampling uncertainty or hypothesis test was applied because the purpose was to characterize the supplied national extract, not infer from a sample to an unobserved population. The reproducible workflow reads the unchanged Schools worksheet, applies explicit eligibility rules, writes machine-

readable audit tables, and produces publication figures. Coordinate coverage is visualized through hexagonal aggregation; no facility-level point symbols or identifiers are published.

All review flags remain separable from the original values. The analysis does not automatically delete coordinate duplicates, overwrite administrative labels, repair implausible values, or impute incomplete condition fields. This choice preserves auditability and supports later adjudication by the data owner and authorized technical teams. The analysis code can be supplied with aggregate outputs subject to TBC and institutional approval; the facility-identifying workbook is not redistributed.

### 3.8 Ethics, confidentiality, and disclosure control

The extract contains facility and administrative records and no individual student, staff, household, patient, or survey-participant data. Nevertheless, exact school-facility locations and operational exception details can be sensitive. The manuscript therefore reports only aggregate coordinate density and aggregate data-quality results. Any institution-specific non-human-subjects determination or ethics waiver required by the target journal must be confirmed before submission.

## 4. Results

### 4.1 Record flow and analytical coverage

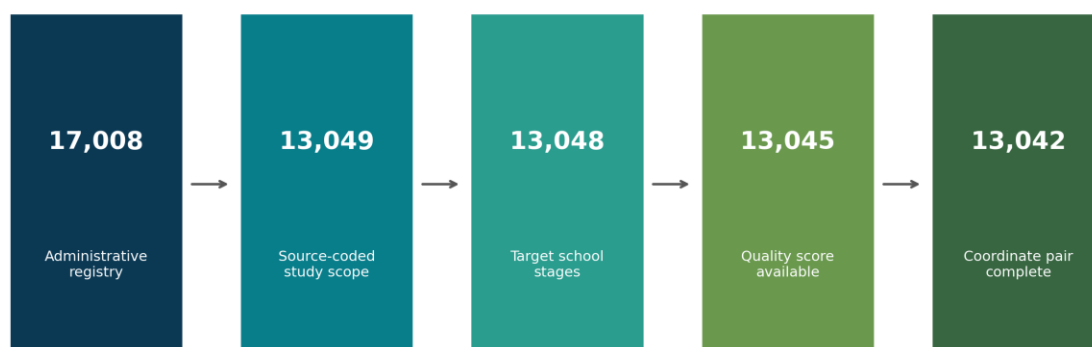
The source registry contained 17,008 records. The pre-existing Scope variable retained 13,049 records and excluded 3,959. One retained record lacked a target-stage classification, leaving 13,048 eligible public elementary, intermediate, or secondary facility assessments. The supplied building-quality field was available for 13,045 eligible records (99.977%). Within that scored subset, 13,042 records had a complete numeric coordinate pair. The primary flow therefore retains 99.954% of the eligible set at the joint score-and-coordinate stage (Table 1; Figure 1). All 13 regions and 149 governorate names were represented in the eligible data.

**Table 1. Sequential construction of the eligible and GIS-ready analytical extract**

| Sequential step                                       | Records retained | Excluded at step | % of eligible* |
|-------------------------------------------------------|------------------|------------------|----------------|
| Administrative source registry                        | 17,008           | —                | —              |
| Source-coded study scope retained                     | 13,049           | 3,959            | —              |
| Elementary, intermediate, or secondary stage retained | 13,048           | 1                | 100.000        |
| Survey-derived building-quality value present         | 13,045           | 3                | 99.977         |
| Complete numeric coordinate pair within scored subset | 13,042           | 3                | 99.954         |

**Note.** The denominator for the final column is the 13,048-record eligible set; the 13,049 scope-retained count precedes exclusion of one record without a target-stage classification.

#### Auditable construction of the GIS-ready analytical extract



Counts are sequential. Coordinate completeness is assessed within records with a supplied building-quality score.

**Figure 1.** Auditable construction of the GIS-ready analytical extract. Counts are sequential; coordinate completeness is assessed within records containing the supplied building-quality value.

#### 4.2 Identifier integrity and record identity

All 13,048 eligible records contained a Unit Code, inspection-report identifier, and primary ministerial number, and each of these fields was unique. No exact duplicate rows were identified. This provides a strong record-level linkage foundation for the supplied extract. Plus Code was complete for all records and contained 13,047 distinct values; one repeated-value group involved two records. Repetition in a secondary locator is a review signal, not proof of duplication.

School names behaved differently. After simple orthographic normalization and whitespace trimming, 11,500 distinct names remained; 955 repeated-name groups involved 2,503 records. The repeats may reflect common naming, branches, stages, or shared facilities. They demonstrate that the human-readable name cannot safely replace a stable identifier. The result is operationally important because 7,548 records (57.848%) reported more than one school in the building, compared with 5,500 records reporting one school; the recorded maximum was 10. A system that conflates the organizational school with the physical building would therefore create systematic ambiguity across more than half of the eligible records.

**Table 2.** Completeness and uniqueness of principal linkage fields

| Field     | Complete n (%)   | Unique values | Repeated groups / records | Interpretation                          |
|-----------|------------------|---------------|---------------------------|-----------------------------------------|
| Unit Code | 13,048 (100.000) | 13,048        | 0 / 0                     | Suitable record key in supplied extract |

| Field                      | Complete n (%)   | Unique values | Repeated groups / records | Interpretation                              |
|----------------------------|------------------|---------------|---------------------------|---------------------------------------------|
| Inspection-report ID       | 13,048 (100.000) | 13,048        | 0 / 0                     | Unique event/report reference               |
| Primary ministerial number | 13,048 (100.000) | 13,048        | 0 / 0                     | Strong cross-system link                    |
| Normalized school name     | 13,048 (100.000) | 11,500        | 955 / 2,503               | Not suitable as a unique key                |
| Plus Code                  | 13,048 (100.000) | 13,047        | 1 / 2                     | Secondary location reference; review repeat |

**Note.** Name uniqueness uses trimmed and simply normalized Arabic text. Repeated values were not removed because co-location and repeated names may be legitimate.

### 4.3 Geospatial readiness and coordinate review

Longitude was nonmissing in 13,047 eligible records and latitude in 13,045. Strict numeric parsing identified three nonnumeric longitude entries and no nonnumeric latitude entries. Among the 13,045 records with the supplied building-quality value, 13,042 contained a complete numeric pair. Every complete pair satisfied global coordinate limits and fell within the broad Saudi plausibility envelope. This establishes national-scale locational plausibility for the analytical extract; it does not establish the exact coordinate reference system or accuracy relative to a building footprint, cadastral parcel, or ground-survey control.

The 13,042 complete pairs represented 13,040 distinct coordinate pairs. Two exact repeated-coordinate groups involved four records. Given the documented frequency of multi-school buildings and the possibility of co-located facilities, these records cannot be classified as duplicates from coordinates alone. The appropriate action is to cross-check the persistent asset identifier, building occupancy relationship, Plus Code, imagery or supporting evidence, and inspection report before deciding whether a correction is required.

Figure 2 presents aggregate spatial density rather than facility points. The spatial footprint is compatible with nationwide coverage, but the source extract did not include an explicit coordinate reference system identifier. Any operational GIS release should therefore document the source and target CRS, transformation history, coordinate precision, capture method, and date of positional verification.

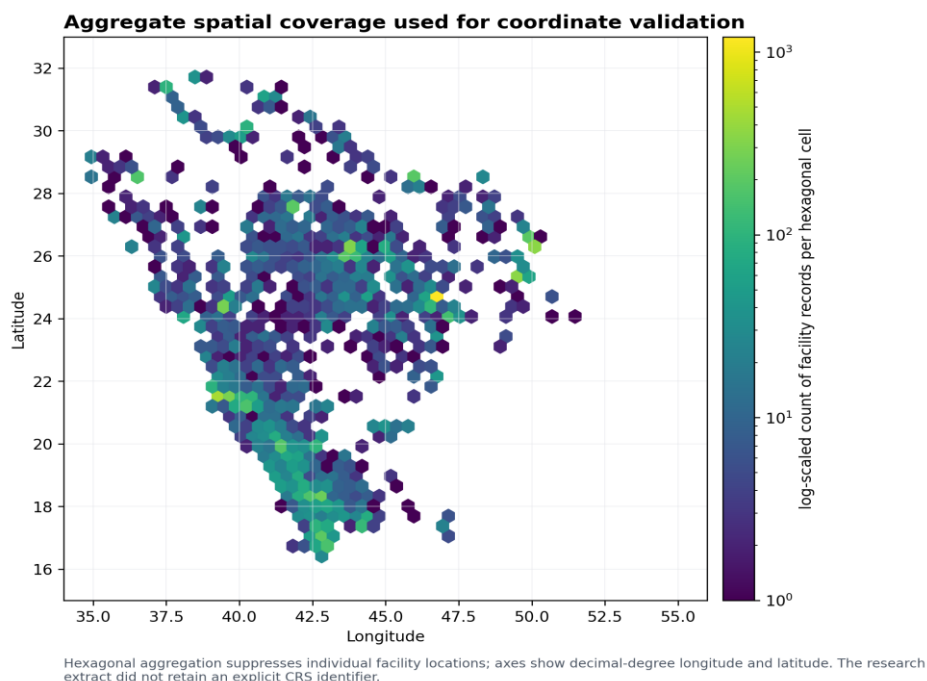


**Figure 2.** Aggregate spatial coverage used for coordinate validation. Hexagonal aggregation suppresses individual facility locations. Axes show decimal-degree longitude and latitude; the research extract did not retain an explicit CRS identifier.

#### 4.4 Administrative hierarchy consistency

The eligible extract contained 149 unique governorate names but 157 observed region–governorate pairs. Seven governorate names were associated with more than one region label. A modal-region screen flagged 11 records whose region differed from the dominant region recorded for the same governorate. The flags were concentrated in eight observed-to-modal combinations and involved Al-Dair, Zahran al-Janub, Jazan, Al-Ardiyat, Balqarn, Al-Ula, and Buraydah (Table 3).

The discrepancy rate is small—0.084% of eligible records—but analytically consequential. A single incorrect hierarchy label can move a facility between regional summaries, routes, dashboards, or accountability units. The mode is useful for detecting unusual assignments, not for authoritatively resolving them. Resolution requires a controlled region–governorate master table, effective dates for boundary changes where relevant, and documented adjudication by the data custodian.

**Table 3.** Region labels that differed from the modal label for the same governorate

| Governorate     | Observed region | Modal region | Flagged records |
|-----------------|-----------------|--------------|-----------------|
| Al-Dair         | Asir            | Jizan        | 2               |
| Zahran al-Janub | Najran          | Asir         | 2               |
| Jazan           | Asir            | Jizan        | 2               |

| Governorate | Observed region | Modal region | Flagged records |
|-------------|-----------------|--------------|-----------------|
| Al-Ardiyat  | Asir            | Makkah       | 1               |
| Balqarn     | Al Baha         | Asir         | 1               |
| Al-Ula      | Tabuk           | Madinah      | 1               |
| Buraydah    | Hail            | Qassim       | 1               |
| Balqarn     | Makkah          | Asir         | 1               |

**Note.** The modal-region comparison is a diagnostic rule derived from this extract. It is not an official correction and should be resolved against an authoritative administrative master.

#### 4.5 Field completeness

Governance and classification fields were complete for all eligible records. Core identity fields and Plus Code were also complete. Source-level longitude and latitude completeness exceeded 99.97%; the difference between source nonmissingness and strict numeric readiness is reported separately. The building-quality field and most component condition fields were 99.977% complete. Asset characteristics were generally high, although building age and establishment year were only 95.241% and 95.256% complete, respectively.

Two decision-relevant fields had distinctly different completeness profiles. Evacuation status contained values for 540 records (4.139%), leaving 12,508 missing. Sprinkler availability contained values for 2,651 records (20.317%), leaving 10,397 missing. Fire-extinguisher information was 97.969% complete. The large field-specific gaps cannot be interpreted as absence, noncompliance, or not-applicable status because the research extract does not contain the survey logic or instrument versions needed to distinguish structural missingness from omitted observation. They should therefore remain an explicit unknown category until the source instrument and conditional rules are reviewed.

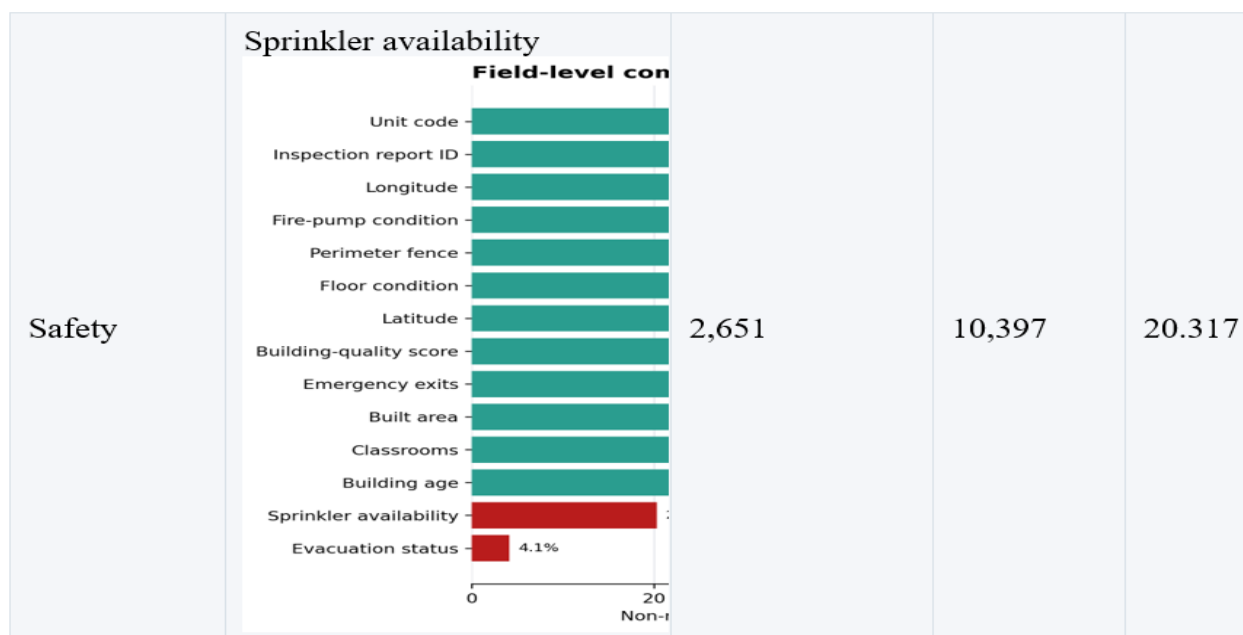
The completeness profile illustrates why a single average data-quality percentage would be misleading. The extract is nearly complete for mapping and record linkage, yet unsuitable for nationwide claims based on the two low-completeness fields. Figure 3 and Table 4 preserve those differences.

**Table 4. Selected field-level completeness among 13,048 eligible facility assessments**

| Domain     | Field                  | Nonmissing n | Missing n | Completeness (%) |
|------------|------------------------|--------------|-----------|------------------|
| Identity   | Unit Code              | 13,048       | 0         | 100.000          |
| Geospatial | Longitude              | 13,047       | 1         | 99.992           |
| Geospatial | Latitude               | 13,045       | 3         | 99.977           |
| Asset      | Building age           | 12,427       | 621       | 95.241           |
| Asset      | Built area             | 13,044       | 4         | 99.969           |
| Asset      | Floors                 | 12,979       | 69        | 99.471           |
| Asset      | Classrooms             | 12,990       | 58        | 99.555           |
| Condition  | Building-quality value | 13,045       | 3         | 99.977           |
| Safety     | Fire extinguishers     | 12,783       | 265       | 97.969           |
| Safety     | Emergency exits        | 13,045       | 3         | 99.977           |
| Safety     | Evacuation status      | 540          | 12,508    | 4.139            |

**Note.** Completeness is source-level nonmissingness, not proof of validity. Strict numeric parsing is reported separately for numeric fields.

**Figure 3. Field-level completeness relevant to GIS integration and decision support. Reporting by field avoids compensation between complete identifiers and incomplete safety variables.**



#### 4.6 Numeric conformance and plausibility flags

Strict parsing identified three nonnumeric values in a nonmissing longitude field and 15 nonnumeric values in the nonmissing built-area field. These entries illustrate a basic conformance distinction: a cell may contain text yet remain unusable for numeric analysis. The supplied building-quality value, building age, floor count, classroom count, and schools-per-building field showed no nonnumeric nonmissing values.

Broad plausibility screening flagged 61 of 12,427 numeric building-age values (0.491%), including impossible negative values and a maximum of 405 years. Two of 13,029 numeric built-area values fell outside 50–20,000 m<sup>2</sup>; 13 of 12,979 floor counts fell outside 1–10; and five of 12,990 classroom counts fell outside 1–100. No supplied building-quality values fell outside 0–100, and no schools-per-building values fell outside 1–10. The extreme age minimum indicates at least one probable transformation or entry error; it should not be winsorized or recoded without reviewing the source record and associated establishment year.

Together with coordinate repetitions and administrative conflicts, these flags define a small but auditable exception queue. None is automatically deleted, because some apparent anomalies may represent unusual but legitimate facilities, co-location, or source formatting. Table 5 separates conformance, plausibility, and consistency so that each flag is routed to the appropriate reviewer.

**Table 5. Machine-detectable exceptions requiring review**

| Quality dimension     | Check                                       | Records / groups flagged | % Of relevant denominator | Required interpretation                     |
|-----------------------|---------------------------------------------|--------------------------|---------------------------|---------------------------------------------|
| Numeric conformance   | Nonnumeric longitude                        | 3 records                | 0.023% of eligible        | Repair only from source evidence            |
| Numeric conformance   | Nonnumeric built area                       | 15 records               | 0.115% of eligible        | Standardize units/format after verification |
| Spatial plausibility  | Repeated coordinate pair                    | 2 groups / 4 records     | 0.031% of complete pairs  | Check co-location and asset identity        |
| Administrative logic  | Nonmodal region label                       | 11 records               | 0.084% of eligible        | Check authoritative hierarchy               |
| Thematic plausibility | Building age outside 0–100                  | 61 records               | 0.491% of numeric         | Check entry and establishment year          |
| Thematic plausibility | Built area outside 50–20,000 m <sup>2</sup> | 2 records                | 0.015% of numeric         | Check unit and source document              |
| Thematic plausibility | Floors outside 1–10                         | 13 records               | 0.100% of numeric         | Check unusual facility vs entry error       |

| Quality dimension     | Check                    | Records / groups flagged | % Of relevant denominator | Required interpretation           |
|-----------------------|--------------------------|--------------------------|---------------------------|-----------------------------------|
| Thematic plausibility | Classrooms outside 1–100 | 5 records                | 0.038% of numeric         | Check zero/outlier against source |

**Note.** Ranges are intentionally broad research-screening rules, not regulatory thresholds. Percentages use the relevant numeric or coordinate denominator shown in Methods.

#### 4.7 Core readiness and the lineage boundary

A total of 13,042 eligible records simultaneously contained the two principal record identifiers, a numeric coordinate pair within the broad national envelope, region and governorate labels, and the supplied building-quality value. Core readiness was therefore 99.954%. This is a strong basis for record-level GIS integration and aggregate descriptive analysis after documented exception handling. It is not a certificate of positional accuracy, instrument validity, or current facility condition.

The extract did not contain an inspection date, inspector or team identifier, instrument version, or edit/QA timestamp. The full survey instrument and score formula was also unavailable. These omissions prevent several important evaluations: temporal currency, change over time, repeat-inspection agreement, inspector effects, version effects, adjudication timing, and reconstruction of how the supplied overall value was produced. A complete lineage package is therefore the principal remaining requirement for converting snapshot readiness into a durable national asset-information system.

#### 4.8 Proposed field-to-decision QA architecture

The empirical audit supports a six-stage architecture (Figure 4; Table 6). Stages 1–4 make data auditable: an authoritative registry and survey specification feed a structured field event; explicit QA gates generate a documented exception queue; and approved corrections are published as a versioned relational/GIS release while the source remains immutable. Stages 5–6 make data actionable: controlled products support asset and planning decisions, and verification or reinspection returns evidence to the next release. The architecture is a proposed control model informed by the audited extract; it should not be read as a retrospective claim that every control was present in the original programme.

**Table 6. Proposed quality gates for a national school-facility field-to-decision system**

| Gate               | Minimum control                                                                                                         | Evidence retained                                                  | Release or escalation rule                                        |
|--------------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|-------------------------------------------------------------------|
| G0 · Data contract | Define asset, school, inspection event, required fields, code lists, CRS, conditional logic, and acceptance thresholds. | Approved schema, data dictionary, instrument version, role matrix. | No field deployment until definitions and ownership are approved. |

| Gate                           | Minimum control                                                                                                                         | Evidence retained                                                          | Release or escalation rule                                                     |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| G1 · Field capture             | Validate task ID, numeric type/range, mandatory and conditional fields, coordinate capture, and evidence attachment.                    | Device/app version, user/team ID, timestamp, raw location, form payload.   | Block invalid types; allow justified exception with reason code.               |
| G2 · Intake                    | Check duplicates, missing identifiers, coordinate envelope, administrative crosswalk, and synchronization completeness.                 | Automated validation log and immutable source snapshot.                    | Quarantine failed records; do not overwrite source.                            |
| G3 · Technical QA              | Review plausibility flags, severe categories, co-location, and supporting evidence using risk-based sampling and targeted verification. | Issue register, reviewer, disposition, evidence, closure date.             | Approve, correct with provenance, accept as legitimate, or reinspect.          |
| G4 · Versioned release         | Apply scripted transformations and authoritative crosswalks; publish metadata and quality statement.                                    | Release ID, code commit, change log, field metrics, unresolved exceptions. | Publish only when agreed fitness-for-purpose criteria are met.                 |
| G5 · GIS and decision products | Maintain restricted facility layer and disclosure-controlled aggregate products; preserve source-to-product lineage.                    | Layer metadata, access log, dashboard/version reference, decision rule.    | Use only for decisions matched to the validated quality profile.               |
| G6 · Feedback and renewal      | Record action, verification, reinspection, repair closure, and asset/school relationship changes.                                       | Work order/action ID, new inspection event, status and effective date.     | Create a new traceable release; never silently revise historical observations. |

**Note.** Recommended governance logic: TBC retains data ownership and acceptance authority; Ayen manages field delivery and programme-level verification within the commissioned scope; release criteria and exception closure should be jointly documented, with independent technical review where risk warrants it.

#### Auditable field-to-decision architecture for school-facility data

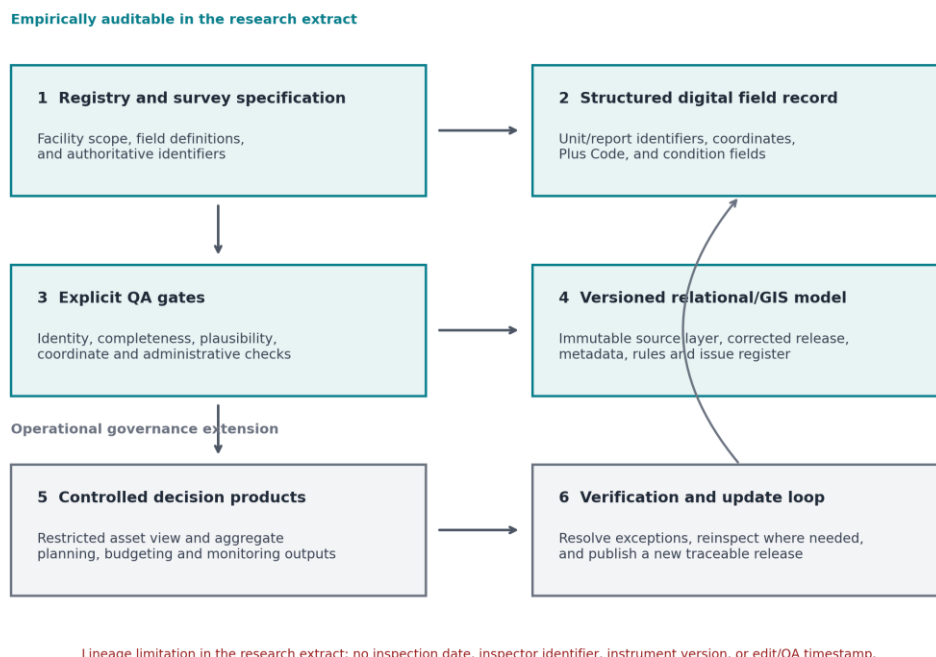


**Figure 4.** Auditable field-to-decision architecture for school-facility data. The upper stages are directly assessable in the research extract; the lower stages are the proposed operational governance extension. The red note identifies lineage fields absent from the research extract.

## 5. Discussion

### 5.1 Principal findings

This national audit reveals a data resource with two simultaneous characteristics. First, its structural foundation is strong: all eligible records have unique Unit Codes, inspection-report identifiers, and ministerial numbers; no exact duplicate rows were found; almost every scored record can be positioned with a numeric coordinate pair; and 99.954% of eligible records meet the prespecified core readiness rule. Second, the extract contains targeted but consequential weaknesses: nonnumeric values in nominally numeric fields, two repeated coordinate groups, 11 administrative hierarchy flags, 81 thematic range flags across four asset variables, two severely incomplete safety fields, and no event-level temporal or audit lineage.

The value of the audit lies in holding those statements together. Reporting only the 99.954% readiness rate would imply a degree of completeness that does not apply to evacuation or sprinkler information. Reporting only the anomaly counts would understate the unusual national coverage and stable identifier structure. A dimensional quality profile supports a more precise conclusion: the extract is highly ready for linkage and broad GIS representation, conditionally ready for

selected descriptive uses after exception review, and not ready for temporal, inspector-reliability, score-validation, or complete evacuation/sprinkler analyses.

## **5.2 Digital structure is necessary but not sufficient**

The audited records demonstrate several advantages compatible with digitally structured capture: stable identifiers, standardized categorical domains, geospatial fields, and near-complete core variables. However, those features cannot show which app controls were active, whether a value was observed or defaulted, how exceptions were supervised, or whether coordinates were captured at the correct asset. Digitalization should therefore be evaluated as a control environment, not celebrated as an outcome by itself.

This interpretation is consistent with comparative digital-collection evidence. Thriemer et al. (2012) observed fewer errors and omissions after electronic capture in a specific surveillance study, whereas Taber et al. (2020) found faster dataset availability and lower total cost but slower individual digital observations in a road-safety setting. The effects depended on questionnaire design, training, field context, and workflow. The present study has no within-programme comparator and appropriately makes no technology-effect estimate. Its contribution is instead to show what can be verified from the resulting extract and what metadata would be required to assess future programme performance.

## **5.3 Stable identity must distinguish assets, schools, and inspections**

The unique Unit Code, report identifier, and ministerial number are a major strength, but their future semantics should be explicit. A record identifier is not automatically a persistent building identifier; an inspection report is an event; and a ministerial school number describes an educational entity. The high prevalence of multi-school occupancy makes a one-table model fragile. If a shared building changes its occupants, opens a new stage, or is reinspected, the system must preserve the asset while creating new dated relationships and events.

A normalized relational design should therefore contain at least five linked objects: Asset, School, Asset–School Occupancy, Inspection Event, and Inspection Observation. Asset holds the persistent physical-facility identifier and authoritative location. School holds the ministerial organizational identifier. Occupancy links the two with effective dates and operating attributes. Inspection Event holds report ID, date, team, instrument version, and evidence package. Observation stores item-level findings, units, categories, and validation status. Such a structure prevents name-based duplication, supports repeat inspections, and aligns condition evidence with lifecycle asset management (International Organization for Standardization, 2024).

## **5.4 Spatial plausibility should not be reported as positional accuracy**

All complete numeric coordinates fell inside a broad national envelope, and aggregate density showed the expected national footprint. These are useful first-line checks. They cannot determine whether a point lies on the correct school building, entrance, parcel, or campus. Positional accuracy requires a reference of known quality and a documented measurement method. The absent CRS identifier also prevents a complete lineage statement even though the values resemble decimal-degree geographical coordinates.

Operational GIS integration should therefore apply layered controls. Syntax and global limits can be enforced at capture. National-envelope and land/sea checks can run at intake. Administrative containment can compare points

against versioned official boundaries. Asset-level verification can compare points with approved building footprints, imagery, Plus Codes, or site evidence. Repeated coordinates should be evaluated against co-location and occupancy relationships. Each layer answers a different question and should retain its reference dataset, date, tolerance, and disposition.

### **5.5 Missingness must remain visible to decision users**

The evacuation and sprinkler fields show why missingness is a governance issue rather than a cosmetic data-cleaning problem. A blank may mean the item was not included in an instrument version, was conditionally not applicable, was skipped, failed to synchronize, or was omitted during transformation. These mechanisms have different implications, but the extract does not distinguish them. Replacing blanks with no, zero, safe, or not applicable would manufacture information.

Future instruments should encode applicability separately from observation status. A recommended structure is: item applicable (yes/no/unknown), observation completed (yes/no), observed value, reason not observed, evidence reference, and reviewer disposition. Instrument versions should define the conditional logic. Dashboards should display denominator-aware completeness alongside any prevalence or status indicator so that a visually strong percentage cannot hide a sparse base. This approach is consistent with the principle that data should be reusable with adequate metadata and with condition-assessment practice that keeps evidence and interpretation traceable (Begić & Krstić, 2024; Wilkinson et al., 2016).

### **5.6 Institutional governance and independent quality assurance**

The TBC–Ayen programme structure provides a practical basis for clarifying responsibility. TBC is the data owner and authorizing custodian; Ayen is the principal consultant that financed and managed expert field teams and supported nationwide collection, verification, analysis, and GIS representation. For a durable system, those programme roles should be translated into an explicit data-governance matrix: who defines the asset master, approves the instrument, grants access, reviews field exceptions, authorizes corrections, publishes releases, and accepts residual risk.

Independence should be risk-proportionate. Routine schema checks can be automated and reviewed within the field programme. Exceptions affecting identity, location, or high-consequence safety decisions warrant documented second-level review; a sample of closed cases should be independently checked. The reviewer should have access to the original record and evidence, not only the corrected value. This separation protects both the data owner and the delivery consultant by making acceptance criteria and residual limitations explicit rather than personal.

The proposed architecture also separates restricted operational products from aggregate planning products. Facility-level teams may need exact coordinates, photographs, names, and work-order status. Researchers and public dashboards may need only aggregated completeness, coverage, or planning indicators. Access control can coexist with FAIR stewardship: accessibility means that conditions and processes are clear, not that every sensitive record is openly downloadable (Wilkinson et al., 2016).

### **5.7 Decision implications**

The audit supports six immediate programme actions, ordered by dependency rather than presumed cost:

1. Adopt a controlled national asset master and distinguish persistent building IDs from school IDs and inspection-event IDs.
2. Make inspection date, inspector/team ID, instrument version, application version, CRS, and edit/QA timestamps mandatory lineage fields.
3. Publish a data dictionary and conditional-logic specification that distinguishes not applicable, not observed, not synchronized, and true absence.
4. Run the documented conformance, completeness, coordinate, administrative, and plausibility checks at intake and before each analytical release; maintain an exception register with disposition evidence.
5. Preserve immutable source snapshots and issue versioned curated releases with code, metadata, change logs, unresolved exceptions, and purpose-specific quality statements.
6. Link downstream work orders, verification, reinspection, and closure back to the inspection event and publish a new release rather than overwriting history.

These actions are deliberately more specific than a generic recommendation to improve data quality. Each maps an observed limitation to a control, a retained artifact, and a decision gate. They also make the programme evaluable: future studies could measure correction rates, time to exception closure, agreement across repeat inspections, positional error against verified controls, and instrument-version effects.

### **5.8 Transferability**

The framework is transferable to other national public-asset surveys—health facilities, municipal buildings, roads, environmental sites, or utilities—because it separates generic controls from domain-specific thresholds. Stable identity, event lineage, immutable source records, versioned transformations, dimensional quality reporting, restricted and aggregate products, and feedback are broadly applicable. The specific variables, plausible ranges, administrative crosswalks, and acceptance thresholds must be adapted to the asset class and decision. Transferability therefore lies in the governance architecture, not in copying the Saudi screening ranges as universal standards.

### **5.9 Strengths and limitations**

The study has several strengths. It audits a large national extract covering every Saudi region and 149 governorates; preserves the facility-assessment unit; distinguishes source nonmissingness from numeric conformance; evaluates identifiers, coordinates, administrative hierarchy, thematic ranges, and lineage in one reproducible workflow; reports dimension-specific results without an unsupported composite index; and does not delete or repair exceptions without source evidence. It also connects empirical findings to a concrete release and governance architecture rather than ending with a generic call for better data.

The limitations define the boundary of the conclusions. The research extract did not contain the full instrument, score formula and weights, inspection dates, inspector identifiers, instrument or application versions, photographs, audit timestamps, or edit history. Consequently, we cannot assess inter-rater reliability, temporal currency, repeatability, field-team effects, rubric validity, evidence sufficiency, or technology-related time and cost. Coordinate checks use a broad plausibility envelope rather than official boundaries or surveyed ground truth; positional accuracy is not estimated. Administrative modal labels are diagnostic, not authoritative. Plausibility ranges are broad research flags, not

engineering or regulatory criteria. Completeness does not prove correctness, and low field completeness may reflect instrument logic that was not available for review.

The audit also examines one research extract rather than the complete production database, data-entry application, or quality-management system. Absence from the extract does not prove that a field or control was absent from all operational systems; it proves that the lineage was unavailable to the secondary analyst. Finally, the dataset contains facility-assessment records, and the current keys cannot be assumed to represent a fully normalized census of unique physical buildings without the recommended asset–school–inspection model.

### 5.10 Key Results

- Inspection date, inspector/team identifier, instrument version, and edit/QA timestamps were absent, limiting temporal and audit-lineage evaluation.
- Evacuation status and sprinkler availability were only 4.139% and 20.317% complete, respectively.
- Eleven records had nonmodal region labels (0.084%) and require authoritative verification.
- Coordinate checks established broad national plausibility, but not positional accuracy; two repeated-coordinate groups involving four records were flagged.
- 13,042 records (99.954%) met the prespecified core readiness rule.
- All eligible records had complete and unique Unit Codes, inspection-report identifiers, and primary ministerial numbers.
- 13,048 eligible facility-assessment records were retained from 17,008 source records across all 13 Saudi regions and 149 governorates.

### 5.11 Key Recommendations

- Link work orders, verification, reinspection, and closure to inspection events so subsequent corrections create traceable releases rather than overwriting history.
- Preserve immutable source snapshots and publish versioned GIS/relational releases with metadata, change logs, unresolved exceptions, and purpose-specific quality statements.
- Maintain an evidence-based exception register and document correction, acceptance, or reinspection decisions.
- Apply automated conformance, completeness, spatial, administrative, and plausibility checks at intake and before each analytical release.
- Define explicit missingness categories distinguishing not applicable, not observed, not synchronized, and true absence.
- Make inspection date, inspector/team ID, instrument version, application version, CRS, and edit/QA timestamps mandatory lineage fields.
- Establish a controlled national asset master distinguishing building, school, and inspection-event identifiers.

## 6. Conclusion

A nationwide field-survey dataset can support reliable geospatial decision-making only when its identifiers, locations, exceptions, and data lineage are transparent and traceable. The Saudi public-school dataset demonstrates

strong structural readiness: 13,048 eligible assessments have completed and unique core identifiers, while 13,042 meet the predefined identity–coordinate–administrative–score criteria. However, the audit identified important limitations, including numeric and plausibility anomalies, 11 administrative hierarchy flags, two repeated-coordinate groups, substantial gaps in evacuation and sprinkler fields, and missing inspection-event lineage. Therefore, GIS readiness should not be interpreted as evidence of positional accuracy, instrument validity, temporal currency, or regulatory compliance. The proposed six-stage quality-assurance architecture—defining data requirements, validating field capture, managing exceptions, evidence-based adjudication, publishing versioned GIS releases, and incorporating verified updates—provides a practical pathway toward a more reliable, transparent, and auditable national school-asset information system.

## References

- Abulibdeh, A., Al-Ali, M., Al-Quraishi, D., Al-Suwaidi, W., Al-Yafei, B., & Al-Mazawdah, S. (2024). Assessing the spatial distribution and accessibility of public and private schools in Qatar: A GIS-based analysis. *Geomatica*, 76(2), 100015. <https://doi.org/10.1016/j.geomat.2024.100015>
- Batini, C., & Scannapieco, M. (2016). *Data and information quality: Dimensions, principles and techniques*. Springer. <https://doi.org/10.1007/978-3-319-24106-7>
- Begić, H., & Krstić, H. (2024). Comprehensive review and comparative analysis of building condition assessment models. *Results in Engineering*, 22, 102176. <https://doi.org/10.1016/j.rineng.2024.102176>
- Begić Juričić, H., & Krstić, H. (2025). Developing a framework for building condition assessment of schools in Osijek-Baranja County. *Buildings*, 15(9), 1511. <https://doi.org/10.3390/buildings15091511>
- Benchimol, E. I., Smeeth, L., Guttman, A., Harron, K., Moher, D., Petersen, I., Sørensen, H. T., von Elm, E., Langan, S. M., & RECORD Working Committee. (2015). The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLOS Medicine*, 12(10), e1001885. <https://doi.org/10.1371/journal.pmed.1001885>
- Congiu, E., Quaquero, E., Rubiu, G., & Vacca, G. (2024). Building Information Modeling and Geographic Information System: Integrated framework in support of facility management (FM). *Buildings*, 14(3), 610. <https://doi.org/10.3390/buildings14030610>
- Dano, U. L., Balogun, A.-L., Abubakar, I. R., & Aina, Y. A. (2020). Transformative urban governance: Confronting urbanization challenges with geospatial technologies in Lagos, Nigeria. *GeoJournal*, 85, 1039–1056. <https://doi.org/10.1007/s10708-019-10009-1>
- Government Expenditure & Projects Efficiency Authority. (2022). *National manual of assets and facilities management (Vol. 3, Chap. 3): Condition assessment survey (CAS) (Document No. EOM-ZC0-PR-000009 Rev 001)*.
- Hartung, C., Anokwa, Y., Brunette, W., Lerer, A., Tseng, C., & Borriello, G. (2010). Open Data Kit: Tools to build information services for developing regions. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development (Article 18, pp. 1–12)*. Association for Computing Machinery. <https://doi.org/10.1145/2369220.2369236>
- International Organization for Standardization. (2023). *Geographic information—Data quality—Part 1: General requirements (ISO 19157-1:2023)*. <https://www.iso.org/standard/78900.html>
- International Organization for Standardization. (2024). *Asset management—Vocabulary, overview and principles (ISO 55000:2024)*. <https://www.iso.org/standard/83053.html>
- Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S.-T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., & Schilling, L. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *eGEMs*, 4(1), 1244. <https://doi.org/10.13063/2327-9214.1244>

- Taber, N., Mehmood, A., Vedagiri, P., Gupta, S., Pinto, R., & Bachani, A. M. (2020). Paper versus digital data collection methods for road safety observations: Comparative efficiency analysis of cost, timeliness, reliability, and results. *Journal of Medical Internet Research*, 22(5), e17129. <https://doi.org/10.2196/17129>
- Thriemer, K., Ley, B., Ame, S. M., Puri, M. K., Hashim, R., Chang, N. Y., Salim, L. A., Ochiai, R. L., Wierzba, T. F., Clemens, J. D., von Seidlein, L., Deen, J. L., Ali, S. M., & Ali, M. (2012). Replacing paper data collection forms with electronic data entry in the field: Findings from a study of community-acquired bloodstream infections in Pemba, Zanzibar. *BMC Research Notes*, 5, 113. <https://doi.org/10.1186/1756-0500-5-113>
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>