Comparison of Auto-generated DB Design to Textual Analysis

Overlapping Matches

Most of the student's nouns are covered either as direct entity categories, table names, or column fields in proposed schema.

- "artist" → Covered in "Artists" table and Persons category.
- "curator" → Covered in Persons/Roles (e.g., Can Bey as curator).
- "exhibition" → "Exhibitions" table and Events category.
- "gallery" → Organizations category.
- "catalog" → Documents category (e.g., auction catalog).
- "document", "record" → "Documents" table.
- "visitor" → Documents (visitor logs) and "Logs" table (visitor-related logs).
- "folder", "spreadsheet" → Mentioned in data management processes; implied in "Documents" table (file_path).
- "auction" → "Auctions" table.
- "bidder" → "Bidders" table.
- "collector" → Persons category (collectors as external entities).
- "object code", "exhibition code" → Codes category; columns like "object_code", "exhibition_code" in tables.
- "condition report" → Documents category; "condition" column in "Artworks".
- "insurance certificate" → Documents category; "insurance_value", "insurance_status" columns.
- "contract", "agreement", "loan", "consignment" → Documents category (loan agreements, consignment forms).
- "committee" → Processes (artist selection committee); "Selection_Decisions" table.
- "portfolio" → Processes (submissions); linked to "Artists".
- "photograph", "sculpture", "painting", "installation" → Medium types in "Artworks" table (medium column).
- "artwork" → "Artworks" table.
- "label" → Processes (labeling in installation).
- "exhibit" → Events (exhibitions).
- "report" → Documents (e.g., condition reports, summary reports).
- "database", "file", "drive", "password", "system" → Data management processes; implied in schema structure (e.g., restricted access).
- "summary", "sheet" → Documents (e.g., lot summary sheet, selection sheet).
- "reserve price", "commission" → Columns in "Auction_Lots" (reserve_price, commission_tier).
- "timeline" → "Logs" table (timeline logs).
- "event" → Events category (exhibitions, auctions).
- "approval" → Processes (approvals in workflows).
- "note", "memo", "comment" → Documents (notes, memos); "remarks" column in "Selection_Decisions".
- "lot" → "Auction_Lots" table.
- "log" → "Logs" table.
- "archive" → Processes (archiving records).
- "year-end summary" → Documents category.
- "access" → Security notes in schema.
- "exhibition dossier" → Documents (exhibition dossier).
- "artist statement" → Linked to "Artists" (background column).
- "email" → Contact info in tables like "Artists", "Bidders".
- "visitor list" → Documents (visitor logs/mailing lists).

- "reference number" → Codes category.
- "ledger", "register" → Documents (code ledger, master object register).
- "template" → Implied in document creation processes.

Mismatches

These nouns appear in the textual extraction but are not explicitly represented as tables, columns, or key entities in proposed analysis/schema. They might be tangential or attributes rather than core entities:

- "wall", "light" → Mentioned in installation processes (e.g., wall color, lighting), but only as fields in "Exhibition_Artworks" (placement, lighting, wall_color). No dedicated table.
- "budget", "invoice" → Financial aspects mentioned in interviews (e.g., budgets in Can Bey's role), but not explicitly in schema (closest is "Sales" table for payments, but no budget/invoice columns).
- "delivery", "shipment" → Part of logistics processes (artwork movement), but tracked via "status" and "location" in "Artworks"; no specific columns.
- "network" → Possibly refers to file-sharing/network access, but not in schema.
- "dashboard", "interface" → Aspirational for future database, but not current in manual system/schema.
- "backup" → Security process, but not a field.
- "file-sharing tool" → Implied in data management, but not specified.
- "folder structure" → Described in processes, but not a table.
- "catalog proof", "mock copy" → Variants of catalog in Documents, but not exact matches.

These are in proposed analysis/schema but missing from the extracted nouns:

- Specific roles like "director", "coordinator", "assistant", "accountant", "appraiser", "technician", "photographer"
- "sale", "incident", "decision", "bidding", "promotion", "payment", "collection" (from workflows).
- Columns like "dimensions", "medium", "owner_type", "status", "location", "starting_bid", "final_price".

Conclusion
Proposed schema focuses on core, relational data entities for a database, while the extracted list includes broader linguistic nouns (e.g., attributes like "wall" or aspirational terms like "dashboard").

About 10-15% of extracted nouns are attributes or minor elements not warranting separate tables/fields. Conversely, proposed list has more specific terms from interviews not captured in their general extraction.