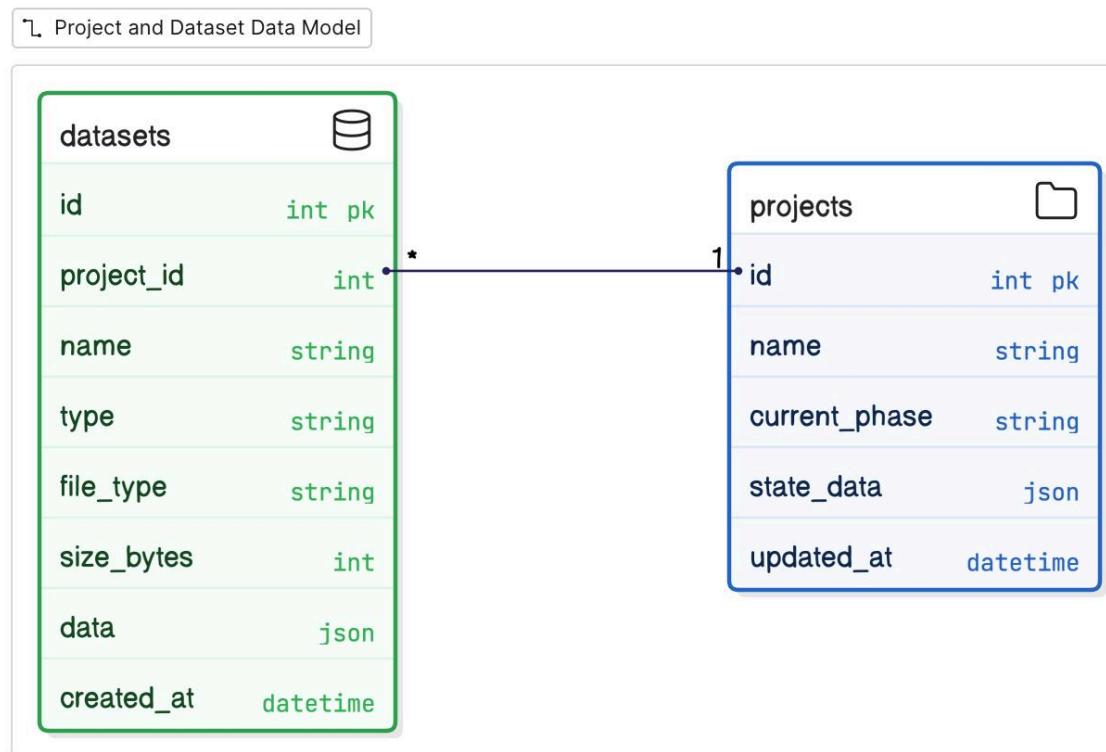


# Easy\_ACC: Visual document representing the database structure and data storage process

## Entity Relationship Diagram



## Workflow Logic

Project Initialization: A project must exist first to generate a project\_id.

Dataset Linking: Every dataset saved via save\_dataset is physically "stamped" with that project\_id.

Scoped Selection: When you view a project, the system only fetches datasets where Dataset.project\_id == active\_project\_id.

## Phase 1: The Gatekeeper (Validation)

Before the database is even touched, the validate\_file\_type function checks the file extension.

Allowed: .csv, .txt, .xlsx

Rejected: Any other extension triggers an UnsupportedFileTypeError.

## Phase 2: Atomic Transaction (The Save)

When save\_dataset is called, the system performs a session transaction:

It checks for the existence of the parent project\_id.

It calculates the size\_bytes of the JSON payload.

It updates/inserts the dataset.

It "touches" the parent project's updated\_at timestamp so the UI shows the project was recently modified.

If any step fails, the session.rollback() ensures no partial data is saved.

## Phase 3: Cascade Cleanup

Because of the Foreign Key constraint with ondelete="CASCADE", deleting a project automatically wipes all its datasets from the disk space.

## Operational Documentation for Developers

Storage Monitoring Use `get_storage_usage()` to retrieve the total byte count of all datasets.

Use Case: Triggering UI alerts when the database nears a size limit (e.g., 2GB).

Data Isolation To retrieve data for a specific project, always use the project\_id filter:

```
[stmt = select(Dataset).where(Dataset.project_id == active_project_id)]
```

**Warning** Deleting a row from the Projects table will permanently and immediately delete all associated rows in the Datasets table.

single database file (projects.db)



The "Project"  
(Process State)

The "Project"  
(Process State)

The "Dataset"  
(Uploaded Data)

Table: projects

Purpose: Primary container for research sessions and workflow state.

Column	Type	Constraints	Description
id	Integer	PK	Unique identifier.
name	String	Unique, Not Null	Human-readable project name.
current_phase	String	-	Workflow status (e.g., EDA, Modeling).
state_data	JSON	-	Serialized UI/Process variables.
updated_at	DateTime	Auto-update	Timestamp of the last user activity.

Table: datasets

Purpose: Stores the actual data content, strictly linked to a parent project.

Column	Type	Constraints	Description
id	Integer	PK	Unique identifier.
project_id	Integer	FK	Reference to <code>projects.id</code> (Cascade on delete).
name	String	Not Null	Filename or internal label.
type	String	-	Category (e.g., "raw", "clean").
file_type	String	-	Restricted to: csv, txt, xlsx.
size_bytes	Integer	-	Data size for storage monitoring.
data	JSON	Not Null	The serialized content of the file.
created_at	DateTime	Default: Now	Timestamp of initial upload.