# Asana Simulation Data Generator – Documentation

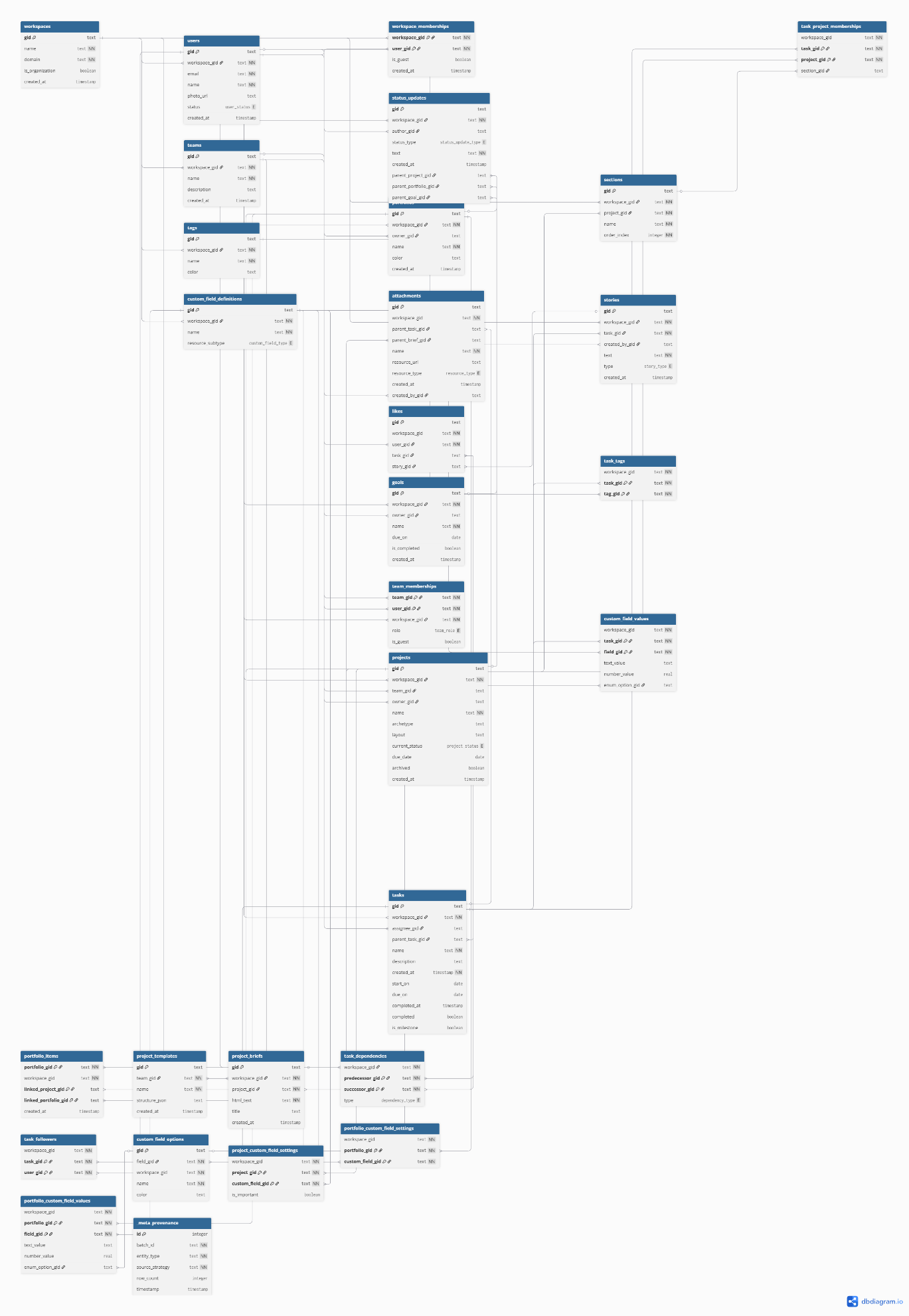
## Section A: Database Schema

### Tables Overview

### The schema consists of 29 tables. See schema.sql for complete DDL.

### Entity-Relationship Diagram

….continued on the following page



### Design Decisions

**1. Custom Fields (vary by project):** Three-table design: custom\_field\_definitions → project\_custom\_field\_settings → custom\_field\_values. Fields are workspace-scoped but project-enabled.

**2. Task Hierarchy (tasks vs. subtasks):** Self-referential FK in single tasks table via parent\_task\_gid. NULL = top-level, non-NULL = subtask.

**3. Workspace Isolation:** Composite unique constraints (gid, workspace\_gid) prevent cross-workspace data leakage.

## Section B: Seed Data Methodology

### Data Sources

| Data Type | Source |
| --- | --- |
| Company names | Y Combinator / Crunchbase naming patterns |
| User names | US Census Bureau demographics |
| Project names | Asana templates, GitHub project boards |
| Task names | GitHub Issues patterns |
| Completion rates | Asana "Anatomy of Work" Index 2023 |

### Table: workspaces

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 first 16 digits (Asana GID format) |
| name | TEXT | Template | Y Combinator-style names: "TechFlow", "DataPulse" |
| domain | TEXT | Derived | Lowercase name + ".com" |
| is\_organization | BOOLEAN | Static | Always 1 (verified org) |
| created\_at | TIMESTAMP | Synthetic | HISTORY\_START - 30 days |

### Table: users

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References parent workspace |
| email | TEXT | Generated | firstname.lastname@{domain} |
| name | TEXT | Census Data | 80% common names, 20% diverse |
| photo\_url | TEXT | Template | ui-avatars.com placeholder |
| status | TEXT | Weighted | 90% active, 5% away, 3% dnd, 2% deactivated |
| created\_at | TIMESTAMP | Synthetic | Exponential growth over 6 months |

### Table: workspace\_memberships

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | References workspaces |
| user\_gid | TEXT (FK) | Derived | References users |
| is\_guest | BOOLEAN | Weighted | 5% guests (contractors) |
| created\_at | TIMESTAMP | Derived | Same as user created\_at |

### Table: teams

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| name | TEXT | Template | Engineering, Product, Design, Marketing, Sales + sub-teams |
| description | TEXT | Template | Team responsibilities text |
| created\_at | TIMESTAMP | Synthetic | Linear distribution across history |

### Table: team\_memberships

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| team\_gid | TEXT (FK) | Derived | References teams |
| user\_gid | TEXT (FK) | Derived | 1-3 teams per user (cross-functional) |
| workspace\_gid | TEXT (FK) | Derived | Composite FK for isolation |
| role | TEXT | Weighted | 85% member, 15% admin |
| is\_guest | BOOLEAN | Weighted | 5% guests |

### Table: portfolios

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| owner\_gid | TEXT (FK) | Derived | Senior users (executives) |
| name | TEXT | Template | "Q1 Initiatives", "Product Roadmap 2024" |
| color | TEXT | Random | Asana color palette |
| created\_at | TIMESTAMP | Synthetic | Exponential growth curve |

### Table: portfolio\_items

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| portfolio\_gid | TEXT (FK) | Derived | References portfolios |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| linked\_project\_gid | TEXT (FK) | Derived | Project in portfolio (mutually exclusive with portfolio) |
| linked\_portfolio\_gid | TEXT (FK) | Derived | Nested portfolio (mutually exclusive with project) |
| created\_at | TIMESTAMP | Derived | After portfolio creation |

### Table: goals

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| owner\_gid | TEXT (FK) | Derived | Senior users |
| name | TEXT | Template | OKR patterns: "Increase ARR by 50%", "Launch v2.0" |
| due\_on | DATE | Synthetic | Quarterly boundaries (Q1, Q2, Q3, Q4 end dates) |
| is\_completed | BOOLEAN | Weighted | 30% completed (based on goal age) |
| created\_at | TIMESTAMP | Synthetic | Quarterly creation pattern |

### Table: projects

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| team\_gid | TEXT (FK) | Derived | 90% have team, 10% workspace-level |
| owner\_gid | TEXT (FK) | Derived | Team admin or senior member |
| name | TEXT | LLM + Template | "[Team] Sprint W03", "[Product] Launch" |
| archetype | TEXT | Weighted | 35% sprint, 25% kanban, 20% launch, 15% ongoing, 5% bugs |
| layout | TEXT | Weighted | 50% list, 35% board, 15% timeline |
| current\_status | TEXT | Weighted | 65% on\_track, 25% at\_risk, 10% off\_track |
| due\_date | DATE | Synthetic | 75% have due, 2-12 weeks out, avoids weekends |
| archived | BOOLEAN | Weighted | 12% archived |
| created\_at | TIMESTAMP | Synthetic | Exponential growth |

### Table: project\_templates

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| team\_gid | TEXT (FK) | Derived | References teams |
| name | TEXT | Template | "Sprint Template", "Launch Checklist" |
| structure\_json | TEXT | Template | JSON with default sections |
| created\_at | TIMESTAMP | Synthetic | Early in history window |

### Table: project\_briefs

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| project\_gid | TEXT (FK) | Derived | One-to-one with project (60% have briefs) |
| html\_text | TEXT | LLM + Template | Project overview, goals, stakeholders |
| title | TEXT | Static | "Overview" |
| created\_at | TIMESTAMP | Derived | Same as project created\_at |

### Table: status\_updates

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| author\_gid | TEXT (FK) | Derived | Project/portfolio owner |
| status\_type | TEXT | Weighted | 65% on\_track, 25% at\_risk, 10% off\_track |
| text | TEXT | LLM + Template | Progress summary, blockers, next steps |
| created\_at | TIMESTAMP | Synthetic | Weekly/bi-weekly intervals |
| parent\_project\_gid | TEXT (FK) | Derived | Mutually exclusive parent references |
| parent\_portfolio\_gid | TEXT (FK) | Derived |  |
| parent\_goal\_gid | TEXT (FK) | Derived |  |

### Table: sections

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| project\_gid | TEXT (FK) | Derived | References project |
| name | TEXT | Template by Archetype | Sprint: [Backlog, To Do, In Progress, In Review, Done] |
| order\_index | INTEGER | Sequential | 0-indexed template order |

### Table: tasks

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | From project or parent task |
| assignee\_gid | TEXT (FK) | Derived | 85% assigned, 15% unassigned (Asana benchmark) |
| parent\_task\_gid | TEXT (FK) | Derived | NULL=top-level, non-NULL=subtask (20% subtasks) |
| name | TEXT | LLM + Template | "[Component] - [Action]", "Fix [issue]" |
| description | TEXT | LLM + Template | 20% empty, 50% short, 30% detailed |
| created\_at | TIMESTAMP | Synthetic | Mon-Wed weighted higher |
| start\_on | DATE | Derived | 35% have start, 1-14 days before due |
| due\_on | DATE | Synthetic | 25% week, 40% month, 20% 1-3mo, 10% none, 5% overdue |
| completed | BOOLEAN | Synthetic | Sprint 70-85%, Kanban 55-70%, Ongoing 40-55% |
| completed\_at | TIMESTAMP | Derived | 1-30 days after creation (log-normal) |
| is\_milestone | BOOLEAN | Weighted | 5% milestones |

### Table: task\_project\_memberships

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| task\_gid | TEXT (FK) | Derived | References tasks |
| project\_gid | TEXT (FK) | Derived | References projects |
| section\_gid | TEXT (FK) | Derived | Section from same project |

### Table: stories

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | From task |
| task\_gid | TEXT (FK) | Derived | References tasks |
| created\_by\_gid | TEXT (FK) | Derived | Team member or follower |
| text | TEXT | LLM + Template | Updates, questions, feedback |
| type | TEXT | Weighted | 80% comment, 20% system |
| created\_at | TIMESTAMP | Synthetic | After task creation, during task lifecycle |

### Table: attachments

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | From parent |
| parent\_task\_gid | TEXT (FK) | Derived | Task attachment (mutually exclusive) |
| parent\_brief\_gid | TEXT (FK) | Derived | Brief attachment (mutually exclusive) |
| name | TEXT | Template | "design\_v2.fig", "requirements.pdf", "screenshot.png" |
| resource\_url | TEXT | Template | Placeholder URLs |
| resource\_type | TEXT | Weighted | 40% image, 30% document, 20% spreadsheet, 10% other |
| created\_at | TIMESTAMP | Derived | After parent creation |
| created\_by\_gid | TEXT (FK) | Derived | Team member |

### Table: tags

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| name | TEXT | Template | "bug", "feature", "urgent", "blocked", "needs-review" |
| color | TEXT | Random | Asana color palette |

### Table: task\_tags

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| task\_gid | TEXT (FK) | Derived | References tasks (30% have tags) |
| tag\_gid | TEXT (FK) | Derived | 1-3 tags per tagged task |

### Table: likes

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | References workspace |
| user\_gid | TEXT (FK) | Derived | Workspace member |
| task\_gid | TEXT (FK) | Derived | 25% of tasks have likes (mutually exclusive) |
| story\_gid | TEXT (FK) | Derived | 15% of comments have likes |

### Table: task\_dependencies

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| task\_gid | TEXT (FK) | Derived | Dependent task |
| dependent\_on\_gid | TEXT (FK) | Derived | Blocking task (same project, 8% have deps) |
| dependency\_type | TEXT | Weighted | 80% finish\_to\_start, 12% start\_to\_start, 8% finish\_to\_finish |

### Table: task\_followers

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| task\_gid | TEXT (FK) | Derived | References tasks |
| user\_gid | TEXT (FK) | Derived | Team members, 1-5 followers per task |

### Table: custom\_field\_definitions

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| workspace\_gid | TEXT (FK) | Derived | Workspace-level |
| name | TEXT | Template | Priority, Story Points, Sprint, Status |
| resource\_subtype | TEXT | Template | "enum", "number", "text" |

### Table: custom\_field\_options

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| gid | TEXT | Generated | UUIDv4 numeric format |
| field\_gid | TEXT (FK) | Derived | References enum field |
| name | TEXT | Template | "P0 - Critical", "Sprint 1", "Approved" |
| color | TEXT | Template | Semantic colors (red=high, green=done) |

### Table: project\_custom\_field\_settings

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| project\_gid | TEXT (FK) | Derived | References projects |
| field\_gid | TEXT (FK) | Derived | 2-4 fields per project |
| is\_important | BOOLEAN | Weighted | 30% important (shown in list) |

### Table: portfolio\_custom\_field\_settings

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| portfolio\_gid | TEXT (FK) | Derived | References portfolios |
| field\_gid | TEXT (FK) | Derived | 1-2 fields per portfolio |
| is\_important | BOOLEAN | Weighted | 50% important |

### Table: custom\_field\_values

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| task\_gid | TEXT (FK) | Derived | Task in project with field enabled |
| field\_gid | TEXT (FK) | Derived | References definitions |
| text\_value | TEXT | LLM | For text fields only |
| number\_value | REAL | Synthetic | Story Points: Fibonacci (1,2,3,5,8,13); Hours: 1-40 |
| enum\_option\_gid | TEXT (FK) | Weighted | Priority: 10% P0, 25% P1, 40% P2, 25% P3 |

### Table: portfolio\_custom\_field\_values

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| workspace\_gid | TEXT (FK) | Derived | Composite FK |
| portfolio\_gid | TEXT (FK) | Derived | References portfolios |
| field\_gid | TEXT (FK) | Derived | References definitions |
| text\_value | TEXT | LLM | For text fields |
| number\_value | REAL | Synthetic | Budget figures, percentages |
| enum\_option\_gid | TEXT (FK) | Weighted | Status distributions |

### Table: \_meta\_provenance

| Column | Data Type | Source Strategy | Methodology & Justification |
| --- | --- | --- | --- |
| table\_name | TEXT | Static | Table being documented |
| record\_count | INTEGER | Computed | Count after generation |
| generation\_strategy | TEXT | Static | Strategy description |
| batch\_id | TEXT | Generated | 8-char UUID for run tracking |
| generated\_at | TIMESTAMP | Static | NOW at generation time |

## LLM Content Generation

### Prompts (in prompts/ folder)

**task\_description.txt:** Context-aware task descriptions **comment.txt:** Status updates, questions, feedback **status\_update.txt:** Project progress summaries **project\_brief.txt:** Project overview documents

### Variety: Temperature 0.7, context injection, template fallback when LLM unavailable

## Temporal Consistency

The schema enforces temporal logic through CHECK constraints and TRIGGERs:

### CHECK Constraints

| Table | Constraint | Rule |
| --- | --- | --- |
| tasks | CHECK (completed\_at IS NULL OR completed\_at >= created\_at) | Tasks cannot be completed before creation |
| tasks | CHECK ((completed = 1 AND completed\_at IS NOT NULL) OR (completed = 0 AND completed\_at IS NULL)) | completed flag and completed\_at must be consistent |
| tasks | CHECK (start\_on IS NULL OR due\_on IS NULL OR start\_on <= due\_on) | Start date cannot be after due date |
| tasks | CHECK (is\_milestone = 0 OR (start\_on IS NULL OR start\_on = due\_on)) | Milestones have same start/due date |
| goals | CHECK (due\_on IS NULL OR due\_on >= DATE(created\_at)) | Goal due date cannot be before creation |
| projects | CHECK (due\_date IS NULL OR due\_date >= DATE(created\_at)) | Project due date cannot be before creation |

### TRIGGER Constraints

| Trigger | Table | Rule |
| --- | --- | --- |
| validate\_task\_parent\_time\_insert | tasks | Child task cannot be created before parent task |
| validate\_task\_parent\_time\_update | tasks | Child task cannot be older than parent task |
| validate\_task\_hierarchy\_recursion | tasks | Prevents circular parent-child references |
| validate\_dependency\_logic\_insert | task\_dependencies | Dependency dates must be logically consistent (finish\_to\_start: predecessor due <= successor start) |
| validate\_dependency\_logic\_update\_tasks | tasks | Moving task dates cannot violate existing dependencies |

### Data Generation Enforcement

| Rule | Implementation Location |
| --- | --- |
| completed\_at >= created\_at | utils/base.py → calculate\_completion\_timestamp() clamps completion to after creation |
| completed\_at <= NOW | utils/base.py → Clamps future dates to NOW - random hours |
| Subtask created after parent | generators/tasks.py → Uses parent's created\_at as minimum timestamp |
| Stories created after task | generators/stories.py → Story timestamp always > task.created\_at |
| Status updates after entity creation | generators/status\_updates.py → Timestamps after project/portfolio/goal creation |

## Relational Consistency

The schema enforces referential integrity through FOREIGN KEYs, UNIQUE constraints, and CHECK constraints:

### Composite Foreign Keys (Workspace Isolation)

| Table | Composite FK | Ensures |
| --- | --- | --- |
| team\_memberships | (team\_gid, workspace\_gid) REFERENCES teams | Team members only in their workspace |
| team\_memberships | (user\_gid, workspace\_gid) REFERENCES users | Users only in their workspace |
| portfolios | (owner\_gid, workspace\_gid) REFERENCES users | Portfolio owner in same workspace |
| projects | (team\_gid, workspace\_gid) REFERENCES teams | Project's team in same workspace |
| projects | (owner\_gid, workspace\_gid) REFERENCES users | Project owner in same workspace |
| tasks | (assignee\_gid, workspace\_gid) REFERENCES users | Task assignee in same workspace |
| tasks | (parent\_task\_gid, workspace\_gid) REFERENCES tasks | Subtask in same workspace as parent |
| task\_project\_memberships | (section\_gid, project\_gid) REFERENCES sections | Section belongs to task's project |
| likes | (user\_gid, workspace\_gid) REFERENCES users | Likes only by workspace members |
| task\_followers | (user\_gid, workspace\_gid) REFERENCES users | Followers only from workspace |
| custom\_field\_values | (enum\_option\_gid, field\_gid) REFERENCES custom\_field\_options | Enum value matches field definition |

### Mutually Exclusive References (CHECK Constraints)

| Table | Constraint | Rule |
| --- | --- | --- |
| portfolio\_items | CHECK ((linked\_project\_gid IS NOT NULL AND linked\_portfolio\_gid IS NULL) OR ...) | Item links to project XOR portfolio, not both |
| portfolio\_items | CHECK (portfolio\_gid != linked\_portfolio\_gid) | Portfolio cannot contain itself |
| status\_updates | CHECK ((parent\_project\_gid IS NOT NULL AND parent\_portfolio\_gid IS NULL AND parent\_goal\_gid IS NULL) OR ...) | Status update has exactly one parent type |
| attachments | CHECK ((parent\_task\_gid IS NOT NULL AND parent\_brief\_gid IS NULL) OR ...) | Attachment belongs to task XOR brief |
| likes | CHECK ((task\_gid IS NOT NULL AND story\_gid IS NULL) OR ...) | Like on task XOR story |
| custom\_field\_values | CHECK (only one of text\_value, number\_value, enum\_option\_gid is set) | Field value matches field type |
| task\_dependencies | CHECK (predecessor\_gid != successor\_gid) | Task cannot depend on itself |
| tasks | CHECK (parent\_task\_gid != gid) | Task cannot be its own parent |

### Unique Constraints

| Table | Constraint | Purpose |
| --- | --- | --- |
| users | UNIQUE(email) | No duplicate emails |
| teams | UNIQUE(workspace\_gid, name) | No duplicate team names per workspace |
| tags | UNIQUE(workspace\_gid, name) | No duplicate tag names per workspace |
| likes | UNIQUE(user\_gid, task\_gid) | User can only like a task once |
| likes | UNIQUE(user\_gid, story\_gid) | User can only like a story once |
| project\_briefs | UNIQUE(project\_gid) | One brief per project |

### Data Generation Enforcement

| Rule | Implementation Location |
| --- | --- |
| Sections from task's project | generators/tasks.py → Section lookup filtered by project\_gid |
| Field values for enabled fields | generators/custom\_fields.py → Checks project\_custom\_field\_settings |
| Likes from workspace members | generators/likes.py → users\_by\_workspace lookup |
| Dependencies within same workspace | generators/dependencies.py → Same workspace filter |
| Team members only from user's teams | generators/teams.py → workspace-scoped assignment |
| Subtasks inherit workspace from parent | generators/tasks.py → Uses parent's workspace\_gid |

## Distribution References

| Metric | Source | Value |
| --- | --- | --- |
| Unassigned tasks | Asana "Anatomy of Work" Index 2023 | 15% |
| Sprint completion rate | Agile velocity benchmarks | 70-85% |
| Completion time distribution | JIRA cycle time research | Log-normal, median 4.5 days |
| Weekend due date avoidance | Calendar planning patterns | 85% |
| Day-of-week task creation | Workplace productivity studies | Mon-Wed higher (weights 1.2-1.3) |
| Team size | McKinsey "State of Organizations" | 40-50 users per team average |
| Teams per user | Cross-functional research | 1-3 teams |