

Decisions · Platform

ADR-0008: The platform consumes the dataset by pin, never by copy

🕒 1 min read · Updated 25 September 2026

Why the platform installs the scripture database from a pinned sggs-data commit and never commits a copy.

<https://docs.gurbanisoul.com/adr/0008-dataset-by-pin/>

Status: accepted (2026-09-24)

Context

The database used to be a Git LFS file committed here. Any commit could change it, a failed cp could truncate it, and nothing proved the served bytes were the ones the scripture gates had checked.

Decision

`dataset.lock.json` names a commit of `Algorythmos-AI/sggs-data` and the database's sha256 and `size.scripts/data/fetch_dataset.py` is the only way the database arrives — in CI, in the API image build and locally (`make dataset`): streamed from sggs-data's LFS, sha256- and size-verified, installed atomically, cached by hash; the image makes it read-only. The integrity check proves the lock and `contract/_meta.json` name the same object, that sggs-data publishes it at the pinned commit, and that the installed database passes `quick_check` with 60,658 lines over Angs 1–1430. The database is never committed here (`/db/` is git-ignored).

Consequences

What production serves is exactly what sggs-data proved, byte for byte, and a revert is a lock revert (every pinned object stays available). A new dataset is a small, reviewable PR. When sggs-data becomes private, the fetch needs a read token (CI's `GH_TOKEN`; a build secret for Render until images are built in CI).

Sri Guru Granth Sahib Ji — Knowledge Base wiki. Generated from commit c3d9360. Updated 25 September 2026. This copy may be out of date; the page is at <https://docs.gurbanisoul.com/adr/0008-dataset-by-pin/>