

Decisions · Data (sggs-data) · pinned from sggs-data

ADR-0003: Every dataset carries its own proofs — ledger, fingerprints, quality contract

🕒 1 min read · Updated 27 October 2026

<https://docs.gurbanisoul.com/data/adr/0003-integrity-proofs/> · a pinned copy of Algorythmos-AI/sggs-data at c36e317

Pinned copy. The canonical file is [sggs-data/docs/adr/0003-integrity-proofs.md](#)¹ at commit c36e317 — edit it there. This wiki publishes it by pin ([ADR-0012](#)); a newer version arrives as a reviewed bump of `docs-site/sources.lock.json`.

Status: accepted (2026-09-24)

Context

The corpus is sacred scripture, extracted verbatim from the source edition. A rebuild, a tool upgrade or an editor must never change a character unnoticed, and “the same data” must be provable across SQLite versions and machines.

Decision

Four independent proofs gate every change (all required CI checks):

- **Reconcile attestation** — the corpus equals the source PDF character for character; the PDF’s sha256 is recorded and a different edition is refused without explicit review.
- **Editorial ledger** (`audit/editorial-ledger.jsonl`, `pipeline/ledger_check.py`) — the only sanctioned text transforms, each itemised and reviewed; any scripture change without a ledger entry fails.
- **Fingerprints** (`pipeline/sggs_integrity.py`, `audit/dataset-fingerprint.json`) — content, not bytes: scripture and T0 hashes, every table (build stamps blanked), every FTS index; stable across SQLite versions.
- **Data-quality contract** (`pipeline/data_quality.py`, `audit/data-quality-baseline.json`) — shape, the `comp_id` gap set (ids are never reused), the heading set, script separation, references, and the rows under scholarly review tracked by their exact Gurmukhi.

Consequences

A rebuild is either provably identical or its differences are named, reviewed and recorded. Baselines are re-recorded only in reviewed PRs. Consumers (the platform, the app) pin a commit and a sha256 and never need to re-prove the text.

Links

1. <https://github.com/Algorythmos-AI/siggs-data/blob/c36e3172bdf38a61070849bb604b38da89ca5d8b/docs/adr/0003-integrity-proofs.md>

Sri Guru Granth Sahib Ji — KnowledgeBase wiki. Generated from commit c3d9360. Updated 27 October 2026. This copy may be out of date; the page is at <https://docs.gurbanisoul.com/data/adr/0003-integrity-proofs/>