

Data · Runbooks · pinned from sggs-data

Runbook: Rebuild the database from the PDF

🕒 1 min read · Updated 27 October 2026

<https://docs.gurbanisoul.com/data/process/runbooks/rebuild-db/> · a pinned copy of Algorythmos-AI/sggs-data at c36e317

Pinned copy. The canonical file is [sggs-data/docs/process/runbooks/rebuild-db.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.

Toolchain trap: the pipeline needs PyMuPDF + numpy + scipy. On this box that is /usr/bin/python3 (3.9), not Homebrew python3 (3.14). make doctor checks it.

```
make doctor                                # confirm interpreter + PDF + LFS DB
cp db/sggs.sqlite /tmp/pre.sqlite         # backup for the proof
bash scripts/data/fetch_translations.sh   # private translation inputs (checksum-
verified)
make rebuild                               # PATH=/usr/bin:$PATH bash
pipeline/rebuild_all.sh                  #
                                         # gates: reconcile char-exact, golden all-
pass
python3 pipeline/verify_regroup.py /tmp/pre.sqlite db/sggs.sqlite # only comp_id/line_no
changed
python3 pipeline/timing/step0_baseline.py --db <pretiming-copy> --force --skip-backup # re-
baseline (see below)
make guard                               # pre-existing tables byte-identical
make reconcile                           # refresh validation/reconcile-
attestation.json
make contract                             # regenerate + assert no unexpected drift
make ios-db                              # rebuild both iOS profiles + license gate
(cd frontend && npm run build:deploy)      # rebuild the served static bundle
```

Re-baselining after a rebuild. step0_baseline.py refuses to run on a DB that already has the timing-layer tables. Snapshot a copy with those 7 tables dropped (timing_sources, raag_timing_claims, shabd_raag_map, shabd_musical_markers, shabd_structural_form, shabd_poetic_genre, timing_migrations), baseline that copy with --force --skip-backup, then make guard against the installed DB.

Note: line_neighbors uses a nondeterministic random projection, so db_sha256 is not bit-reproducible across rebuilds. The scripture guarantees (reconcile, verify, guard) are exact; the analytics layer is regenerated and re-baselined each time.

Links

- <https://github.com/Algorythmos-AI/sggs-data/blob/c36e3172bdf38a61070849bb604b38da89ca5d8b/docs/process/runbooks/rebuild-db.md>