

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

ADR-0004: The dataset is recoverable from its private backups alone, and that is drilled

🕒 1 min read · Updated 27 October 2026

<https://docs.gurbanisoul.com/data/adr/0004-backups-and-restore/> · a pinned copy of Algorythmos-AI/sggs-data at c36e317

Pinned copy. The canonical file is [sggs-data/docs/adr/0004-backups-and-restore.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 source PDF, the English translation sources and the ShabadOS database (the Nitnem input) are not in this repository. The PDF once lived on one laptop; the ShabadOS file was an unpinned third-party download. Losing either would make the dataset impossible to rebuild.

Decision

Every rebuild input is backed up in the private Algorythmos-AI/sggs-source repository (releases `source-pdf-v1`, `translations-en-v1`, `shabados-database-4.8.7`; never replaced), pinned by a committed hash, restored only through verifying scripts, and refused by the rebuild when missing or altered. `scripts/data/dr_drill.sh` restores the dataset from those backups into a fresh clone, rebuilds with every gate and requires identical data; the `dr-drill` workflow runs it on a clean runner every quarter with the pinned toolchain. Results are logged in `audit/dr-drills.md`.

Consequences

Recovery is one command and is known to work, not assumed. Copies 2 and 3 of the 3-2-1 plan (an object-lock bucket, an offline drive) remain owner decisions. The drill needs a read-only token for `sggs-source` (`SGGS_SOURCE_TOKEN`).

Links

1. <https://github.com/Algorythmos-AI/sggs-data/blob/c36e3172bdf38a61070849bb604b38da89ca5d8b/docs/adr/0004-backups-and-restore.md>