

Get started · Exercises

Exercise 03: bump a pin in a scratch branch

🕒 2 min read · Updated 25 September 2026

Change dataset.lock.json to a wrong hash in a scratch branch, watch make dataset-check and the fetcher refuse, restore it — what a dataset bump is and what the integrity gate proves.

<https://docs.gurbanisoul.com/exercises/03-bump-a-pin/>

Goal. See exactly what a dataset bump is and what refuses a bad one. **Needs.** the local setup. 30 minutes. Nothing you do here can change the served data.

1. Read the pin

```
cat dataset.lock.json
make dataset-check
```

Expected: the lock's commit, sha256 and size; then two green lines — the pin agrees with contract/_meta.json, and sggs-data publishes that object at the pinned commit.

2. Break it on purpose

```
git switch -c scratch/exercise-03 integration
python3 - <<'EOF2'
import json, pathlib
p = pathlib.Path('dataset.lock.json'); d = json.loads(p.read_text())
d['database']['sha256'] = '0' * 64          # a hash sggs-data never published
p.write_text(json.dumps(d, indent=2) + '\n')
EOF2
make dataset-check ; echo "exit=$?"
```

Expected: --check-repo fails first — the lock and contract/_meta.json no longer name the same object — with a non-zero exit. Now try to install it:

```
python3 scripts/data/fetch_dataset.py ; echo "exit=$?"
```

Expected: the download is refused or the verification fails; db/sggs.sqlite on disk is untouched (the install is atomic — check its sha256 with shasum -a 256 db/sggs.sqlite before and after).

3. Restore

```
git switch integration && git branch -D scratch/exercise-03
make dataset-check
```

4. What a real bump looks like

Read the dataset pin: a real bump changes the commit **and** the hash **and** the size to values sggs-data published, regenerates the contract with `make contract`, and goes through a pull request where the integrity check repeats what you just saw — plus opens the installed database and counts its lines and Angs.

Self-check

The lock names a sha256 that sggs-data never published. Where is that caught?

- ☐ make dataset-check (the lock vs contract/_meta.json, then the pin vs sggs-data), and again by the integrity workflow on the pull request
- ☐ Only at deploy time
- ☐ Nowhere; the fetch would download whatever the commit holds

How do you roll a dataset back?

- ☐ Revert the lock bump in a pull request
- ☐ Restore db/sggs.sqlite from a backup by hand
- ☐ Delete the newer object in sggs-data