



MAYAN ALFA

17. Dataset Management

Dataset management was defined as a key infrastructure layer of the MAYAN_ALFA Research project.

Canonical CSV datasets represent:

- the reproducibility layer,
- benchmark evidence,
- archive structure,
- and public release data.

Every dataset must be:

- versioned,
- audited,
- public-safe,
- and consistent with release metadata.

Recommended Dataset Naming

`MAYAN_SUMMARY_LIMIT_XXXXXXXX_STEP_XXXXXXXX_TIMESTAMP.csv`

Core Principles

- immutable snapshots,
- canonical outputs,
- release consistency,
- no hidden manipulation.

CSV datasets must be:

- archived,
- hashed,
- linked to DOI releases,
- and linked to release notes.

Public CSV Layer License

CC BY-NC 4.0

The following must never be published:

- internal interpretive data,
- Observation Registry internals,
- heuristics.



MAYAN ALFA

CSV datasets serve as:

- an evidence layer,
- a reproducibility layer,
- and the benchmark foundation of the project.