

# Artifact Sync System Convergence Contract

## v1.1

### 1. Purpose

This contract defines a deterministic, safe convergence process that synchronizes selected Google Docs into version-controlled Markdown artifacts under `artifacts/`.

The system is a reconciler: it repeatedly converges derived repository state to match the current set of eligible source documents.

### 2. Managed Scope

#### 2.1 Managed Subtree (Scoped)

- The subtree `artifacts/` contains managed output **and** unmanaged project-level content.
- The sync process may create, overwrite, move, or delete content **only within managed document units** (see §2.2, §4.3, §7).

#### 2.2 Managed Document Units

A managed document unit is the directory:

- `artifacts/<project_norm>/<docId>/`

The sync process is the publisher-of-record for managed document units and may create, overwrite, move, or delete content **only** within those units, subject to safety gates.

#### 2.3 Unmanaged Project-Level Area (New)

Within a project directory:

- `artifacts/<project_norm>/`

any path **not** under a managed document unit (`<docId>/`) is unmanaged and **MUST NOT** be modified, deleted, or relocated by the sync process.

This includes:

- loose files directly under `artifacts/<project_norm>/`
- subdirectories such as `_mermaid/`, `notes/`, `refs/`, etc.

## 2.4 Unmanaged Area (Repo-Wide) (Retained)

- No files outside `artifacts/` are modified.

## 2.5 Reserved Subpaths (Retained)

- `artifacts/.sync/` is reserved for run outputs (reporting/indexing) and is also managed.

# 3. Source Model

## 3.1 Source Objects

- A “source document” is a Google Doc identified by a stable `docId`.

## 3.2 Inclusion Authority

A source document is eligible for export iff:

- It has the Drive label `artifact`.

## 3.3 Required Source Metadata

If a document is labeled `artifact`, it must also provide:

- `project` (string) (label field)

If `project` is missing/blank, the document is eligible-but-invalid (see §6).

# 4. Derived State Model

## 4.1 Canonical Identity

- Canonical identity is `docId`.
- `docId` is globally unique and stable.

## 4.2 Project Grouping

`project` is used only for grouping/placement under `artifacts/` and may change over time.

`project` is normalized to a safe path segment (`project_norm`) by:

- lowercase
- trim
- whitespace → -
- allow only [a-z0-9-]
- collapse repeated -
- trim leading/trailing -

If normalization yields empty, the document is eligible-but-invalid (see §6).

### 4.3 Output Layout (Doc Unit)

For each successfully exported document:

- `artifacts/<project_norm>/<docId>/doc.md`
- `artifacts/<project_norm>/<docId>/media/*`
- `artifacts/<project_norm>/<docId>/meta.json`

`meta.json` includes at minimum:

- `docId`
- `title` (current)
- `project_norm`
- `exportedAt` (timestamp)

### 4.4 Project-Level Unmanaged Content (New)

The project directory `artifacts/<project_norm>/` MAY contain unmanaged, human-authored content outside any `<docId>/` unit.

This contract does not prescribe structure for unmanaged content.

### 4.5 Discoverability Index (Retained)

Each successful run writes:

- `artifacts/.sync/index.json`

Containing an entry per successfully exported document:

- `docId`
- `title`
- `project_norm`
- `path`
- `exportedAt`

## 5. Convergence Model

### 5.1 Desired Set

On each run, the system constructs a desired set from the current source model:

- `EligibleDocs` = docs with label `artifact`.
- `ExportableDocs` = subset of `EligibleDocs` with valid `project` and valid `project_norm`.
- `DesiredSet` = subset of `ExportableDocs` that were successfully exported in the current run.

### 5.2 Convergence Target (Scoped)

The convergence target applies only to managed document units:

- Under `artifacts/` (excluding `artifacts/.sync/`), the only **managed document unit directories** that should exist are those corresponding to the current desired mapping:
  - (`<project_norm>`, `<docId>`) pairs in `DesiredSet`.

Unmanaged project-level content under `artifacts/<project_norm>/` is excluded from convergence.

### 5.3 Relocation Semantics (Project Change)

If a doc remains labeled `artifact` and its `project` value changes:

- The next successful export places it under the new `artifacts/<project_norm>/<docId>/`.
- The prior location is removed during deletion (see §7), provided deletion gates allow it.

### 5.4 Unlabel Semantics

If a previously exported doc no longer has label `artifact`:

- Its managed document unit directory is removed during the next deletion-permitted run.

## 6. Processing and Error Classification

### 6.1 Per-Document Outcomes

Each eligible doc results in exactly one of:

- `EXPORTED`

- `SKIPPED_METADATA` (e.g., missing/invalid `project`)
- `FAILED_OPERATIONAL` (e.g., export/convert failure for this doc)

## 6.2 Metadata Skips

If `project` is missing/blank or normalizes to empty:

- Outcome: `SKIPPED_METADATA`
- Severity: `WARN`
- Continue processing other documents.

## 6.3 Operational Failures

If export or conversion fails for a specific doc due to:

- network/API transport errors
- conversion tool failure
- unexpected runtime exception during export/convert

Then:

- Outcome: `FAILED_OPERATIONAL`
- Severity: `ERROR`
- The run must abort (see §8) and deletions must not occur.

## 6.4 Run Outputs

Each run writes:

- `artifacts/.sync/errors.json`
- `artifacts/.sync/REPORT.md`

These include:

- counts (eligible, exported, skipped\_metadata, failed\_operational)
- per-doc entries (`docId`, title, outcome, reason)

# 7. Deletion and Safety Gates

## 7.1 Managed Deletion Unit (Unchanged)

Deletion applies only at the directory unit:

- `artifacts/<project_norm>/<docId>/`

No partial deletion within a unit is required by this contract.

## 7.2 Deletion Preconditions (Unchanged)

Deletion is permitted only if all are true:

1. Drive listing succeeded (`EligibleDocs` computed).
2. At least one document was successfully exported in the run (`EXPORTED >= 1`).
3. No global infrastructure failure occurred.
4. No per-document operational failures occurred (`FAILED_OPERATIONAL == 0`).

## 7.3 Deletion Rule (Scoped)

When deletion is permitted, the system converges managed document units:

- `CurrentUnits` = all existing directories matching `artifacts/<project_norm>/<docId>/` (excluding `.sync/`).
- `DesiredUnits` = directories for `DesiredSet` as produced in the run.
- Delete: `CurrentUnits - DesiredUnits`.

Unmanaged project-level content under `artifacts/<project_norm>/` is never a deletion candidate.

## 7.4 No Project Directory Deletion (New)

The sync process MUST NOT delete a project directory:

- `artifacts/<project_norm>/`

even if it contains zero managed document units, and regardless of whether the project appears in the current run's `DesiredSet`.

## 8. Run-Level Failure Model (Retained)

### 8.1 Abort on “All Eligible but Nothing Exported”

If:

- `EligibleDocs > 0` and
- `EXPORTED == 0`

Then:

- Exit status code = 1
- No deletions occur

### 8.2 Abort on Operational Failure

If:

- any document outcome is `FAILED_OPERATIONAL`

Then:

- Exit status code = 1

No deletions occur

### **8.3 Abort on Global Infrastructure Failure**

If:

- authentication fails, or
- Drive listing fails

Then:

- Exit status code = 1
- No deletions occur

## **9. Operational Guarantees**

### **9.1 Safety Guarantee**

The system will not delete managed artifacts unless a run produced at least one successful export and encountered zero operational failures.

### **9.2 Determinism Guarantee**

Given identical source inputs and stable conversion, the system produces identical derived layout for each (`project_norm`, `docId`) unit.

### **9.3 History Guarantee**

Artifact removal and relocation are performed via normal Git commits; prior versions remain available via Git history.

### **9.4 Unmanaged Preservation Guarantee (New)**

Unmanaged project-level content under `artifacts/<project_norm>/` (outside any `<docId>/` unit) is preserved across runs and is not modified or deleted by the sync process.

## **10. Non-Goals (Amended)**

- Preserving Google Drive folder structure.
- Treating document title as identity.
- Rename detection beyond `docId` identity.
- Partial merges of manual edits under `artifacts/<project_norm>/<docId>/`.
- Deletion on “empty desired set” runs.
- Managing or indexing project-level unmanaged content beyond preserving it.

## 11. Implementation Notes (Non-Normative) (Retained)

The current reference implementation:

- runs in an automated CI workflow,
- retrieves documents via Google Drive API,
- exports to an intermediate document format,
- converts to Markdown and extracts embedded media,
- commits results back to the repository.

Tooling and platform details are replaceable; contract semantics remain unchanged.