White Paper on Community Archives of Significant Events

A Guide to Building Community Memory Infrastructure Based on Chinese Archival Governance Experience and Open-Source Technologies

Version: 1.2
Nature: Operational Guide / Policy Recommendation
Audience: Community self-governance organisations, local archives, civil society organisations, community memory project initiators

Preface

The transgressions (legal, moral, political) and disasters (environmental, climatic, public health, and otherwise) that occur within a community constitute the most sensitive—and most systematically forgotten—parts of its collective memory. Power structures tend to place these records in the category of "not for public release" or "to be destroyed," while oral tradition dissipates as witnesses pass away. This white paper draws on several transferable structural principles from China's archival governance system—including tiered management and territorial responsibility, the "four-property testing" (四性检测, sì xìng jiǎn cè) quality control framework, the disaster-recovery logic of off-site and heterogeneous backups, and the downward extension of archival infrastructure to the community level—and combines them with a mature open-source technology stack to propose a construction plan for community-level archives of transgressions and disasters.

This plan does not presuppose any particular political stance. Its core objective is: to enable communities to record their own traumas and conflicts, and to ensure that these records are technically verifiable, institutionally sustainable, and ethically controllable.

Chapter 1: Institutional Framework

Extracting Transferable Principles from Chinese Archival Governance

1.1 Tiered Management and Territorial Responsibility

China's archival system emphasises "unified leadership and tiered management" (统一领导, 分级管理, tǒng yī lǐng dǎo, fēn jí guǎn lǐ), with grassroots archival work incorporated into rural revitalisation and grassroots governance assessments. It encourages a "village archives held by township" (村档乡管, cūn dàng xiāng guǎn) model to address inadequate storage conditions at the village level. For community archives of transgressions and disasters, a transferable variant is:

- Community collection, higher-level custody: Community volunteers or self-governance organisations handle frontline collection and preliminary arrangement, then transfer the final digital archival packages (including metadata and verification information) to higher-level institutions with better technical capacity (such as local public libraries, university archives, or regional community memory alliances) for long-term preservation.

- Separation of ownership and custody: The community retains content ownership and access control rights over its archives, while the custodian bears technical preservation and format migration responsibility. This separation respects community agency while addressing the practical dilemma of small communities lacking climate-controlled storage and dedicated IT staff.

1.2 "Four-Property Testing" (**四性检测**, sì xìng jiǎn cè) as a Quality Baseline

The "four-property testing" in Chinese electronic records management—authenticity (**真实性**, zhēn shí xìng), integrity (**完整性**, wán zhěng xìng), usability (**可用性**, kě yòng xìng), and security (**安全性**, ān quán xìng)—provides an operational quality control checklist. Adapted to the community transgression and disaster archive context:

Dimension	Operational Definition in Community Archive Context	Testing Method
Authenticity	Consistency between recorded content and what occurred at the time; traceable provenance	Hash value recording, collection timestamps, collector identity binding (decentralised identifiers)
Integrity	No missing associated records for a single event/disaster (e.g., witness testimony + on-site footage + official statements + subsequent handling records)	Associated metadata completeness checks, event-to-record mapping tables
Usability	Archives remain readable and retrievable after 5–10 years	Format migration plan (migration to open formats), periodic readability sampling
Security	Sensitive content is not leaked; not tampered with without authorisation	Tiered access control, encrypted storage, operation log auditing

1.3 The "Off-Site, Heterogeneous" Backup Principle

China's archival security standards require important digital resources to achieve off-site (**异地**, yì dì), heterogeneous (**异质**, yì zhì) backups, with online backups completed within 7 days of data creation and offline backups within 30 days. A simplified implementation plan for community archives:

- Online copy: Deployed on community-owned small servers or trusted cloud services (data sovereignty risks must be assessed — see Section 3.1).
- Offline copy: Periodically written to write-once optical discs (such as M-DISC) or encrypted portable hard drives, stored at different physical locations (e.g., one at the community centre, one at the custodian institution).
- Verification mechanism: Each copy is accompanied by a SHA-256 hash manifest; hashes are compared upon recovery.

Chapter 2: Technology Stack

Selection and Integration of Open-Source Tools

2.1 Core Principles: Low Barrier, Collaborative, Exitable

Community archive projects should not depend on any single vendor or proprietary format. Selection criteria are: open-source licence (GPL/MIT/Apache, etc.), graphical interface priority, open data export formats, verifiable community activity. The following tools have all been validated through case studies or have mature documentation.

2.2 Collection and Description Layer

Omeka (or Omeka S) is the most suitable starting point for community scenarios. It is an open-source web publishing platform designed specifically for sharing digital collections and online exhibitions, supporting multimedia content (images, audio, video), the Dublin Core metadata standard, and an active community plugin ecosystem. Communities can use it to quickly build a publicly accessible archive portal while managing collection entries in the backend.

Supplementary collection tools:

- Tropy (open-source, desktop): For organising photographic archives, supporting EXIF extraction and custom metadata templates.
- Oral History Metadata Synchronizer (OHMS) (open-source, web-based): For indexing and synchronising oral history interviews, suitable for fine-grained annotation of disaster survivors' testimony.
- Mobile: Use Open Camera (open-source Android app) to automatically record GPS and timestamps when shooting, avoiding reliance on manufacturers' proprietary metadata.

2.3 Data Cleaning and Standardisation Layer

Community-collected data is often format-messy, inconsistently named, and variably dated. *OpenRefine* is the core tool at this stage. It is an open-source, browser-based data cleaning and transformation platform supporting CSV, TSV, Excel, JSON, and other formats, offering faceted browsing, clustering/merging (for unifying表述 differences such as "XX Road" vs. "XX Street"), and batch format conversion.

Key operations: Before importing into Omeka, use OpenRefine to complete the following standardisation:

- Unify date fields to ISO 8601 format (YYYY-MM-DD).
- Map place names against a community-consensus controlled vocabulary (see Section 3.2).
- Unify event classification tags (transgression type/disaster type) into a controlled vocabulary (see Section 3.3).
- Unify actor and role terms (see Section 3.4) into the same controlled vocabulary.
- Preserve original columns, add cleaned columns, ensuring traceability.

2.4 Long-Term Preservation and Authentication Layer

AtoM (Access to Memory) is an open-source archival description and management system widely used in the international archival community, following international description standards such as ISAD(G), supporting multi-level description (from fonds to item) and rights management. The case of the Mennonite Archival Information Database demonstrates that small archival institutions with limited resources can jointly deploy and maintain AtoM through consortium arrangements, sharing technical costs and sustaining collaboration through regular online meetings.

Authentication and tamper-proofing:

- Generate a SHA-256 hash value for each archived entry and record it.
- For critical testimony or official documents, use *OpenTimestamps* to anchor hash values to the Bitcoin blockchain, providing timestamp proof independent of any single institution. Cost is minimal (near zero) and does not rely on trust in a specific institution.
- If the community has PGP/GPG capability, collectors can sign submitted content with personal keys to achieve provenance authentication.

2.5 Analysis and Presentation Layer

Analysis: For small-to-medium structured data (hundreds to thousands of records), OpenRefine's faceting and clustering functions can already accomplish basic pattern recognition (e.g., temporal clustering of certain transgression types, frequency of disaster reports in certain areas). For deeper analysis, export data as CSV and import into *Metabase* (open-source BI tool) for visual exploration.

Presentation:

- Omeka's built-in exhibition builder can create narrative exhibitions for public audiences.
- For projects requiring greater interactivity, use Omeka as a data backend and *Wax* (open-source, IIIF-based) or *CollectiveAccess* (open-source, suited for complex relational data) for frontend customisation.

2.6 Technology Stack Summary Table

Stage	Tool	Licence	Deployment	Role in the Archive	Priority	Community Suitability Notes
Collection	Open Camera	GPL	Local (Android)	Captures GPS and timestamps without vendor-proprietary metadata	[P]	Zero cost, no network required
Collection	Tropy	MIT	Local (desktop)	Organises photographs; extracts EXIF; applies custom	[P]	Graphical interface; no coding required

Stage	Tool	Licence	Deployment	Role in the Archive	Priority	Community Suitability Notes
				metadata templates		
Collection	OHMS (Oral History Metadata Synchronizer)	Open source (MIT)	Web (self-hosted)	Indexes and synchronises oral history interviews; fine-grained annotation	[E]	Requires basic web hosting; strong for testimony work once oral history programme matures
Collection / Web archiving	ArchiveBox	MIT	Local or server	Periodic crawling and local archiving of publicly available but removable online content	[P]	Command-line oriented; schedule via cron
Collection / Web archiving	Conifer	Open source	Local or server	Interactive web capture (browser-based) of dynamic pages and official notices	[A]	Use where ArchiveBox fails on dynamic pages; interchangeable at the capture stage
Description & Publishing	Omeka / Omeka S	GPL	Self-hosted or shared hosting	Public-facing archive portal; Dublin Core metadata; exhibition builder; role-based access plugins	[P]	User-friendly GUI; rich plugin ecosystem
Data Cleaning & Standardisation	OpenRefine	BSD	Local or intranet server	Batch cleaning, clustering/merging, date and place standardisation, controlled-vocabulary mapping	[P]	Visual, no coding; preserves original columns
Archive Management	AtoM (Access to Memory)	GPL	Server (consortium sharing	Multi-level archival description (ISAD(G)); rights	[E]	Professional standards; best deployed jointly once collection

Stage	Tool	Licence	Deployment	Role in the Archive	Priority	Community Suitability Notes
			recommended)	management; long-term management		volume outgrows Omeka
Integrity Authentication	SHA-256 (via coreutils / built-in tooling)	GPL / system	Local	Hash manifests for every archival package and every backup copy	[P]	Baseline; mandatory at every transfer
Integrity Authentication	OpenTimestamps	Open source (LGPL)	CLI / API	Anchors SHA-256 hashes to the Bitcoin blockchain for independent timestamp proof	[E]	Near-zero cost; add once baseline hashing is routine
Integrity Authentication	GnuPG (PGP/GPG)	GPL	Local	Collector key-signing of submitted content for provenance authentication	[A]	Requires key-management training; alternative to OpenTimestamps for provenance
Encrypted storage	VeraCrypt	Apache 2.0 / TrueCrypt licence	Local	Encrypted containers for sealed records	[P]	Plausibly deniable volumes where lawful
Attribution management	Omeka S resource templates + custom fields (or AtoM actor records)	GPL	Server	Stores actor type, role, attribution status, evidence links, and right-of-reply records	[P]	Native to both platforms; no additional tool required
Analysis	Metabase	AGPL	Server	Drag-and-drop dashboards for pattern exploration of exported CSV	[E]	No coding; connects to common databases; add when analysis needs exceed OpenRefine
Presentation	Wax	Apache 2.0	Web	IIIF-based narrative exhibitions using	[E]	Requires frontend skills; add for public-facing

Stage	Tool	Licence	Deployment	Role in the Archive	Priority	Community Suitability Notes
(advanced)				Omeka as data backend		curated exhibitions
Presentation (advanced)	CollectiveAccess	GPL	Server	Complex relational data presentation and management	[A]	Steeper learning curve; alternative to AtoM+Omeka for larger consortia with complex relations
Controlled Vocabulary	SKOS (W3C standard, RDF/XML or Turtle export)	W3C standard	File-based	Versioned multilingual vocabulary; persistent URIs; language-tagged labels	[P]	Exportable as dated files stored with the archive
Diagrams / process documentation	Mermaid	MIT	Local or web	Flowcharts for risk assessment, classification, attribution, safety escalation, committee structure	[P]	Text-based; version-controllable alongside policy documents

Priority definitions:

- [P] Primary: Adopt at launch. The archive cannot meet its baseline obligations (collection, description, integrity, access control) without it.
- [A] Alternative: Interchangeable substitute for a Primary tool at the same stage. Choose one, not both, unless there is a specific reason to run parallel.
- [E] Extension: Adopt later, once volume, staffing, or programme maturity justifies it. Not required for a functioning pilot.

Deployment note: Tools marked ”server” (AtoM, Metabase, OHMS, Omeka S, CollectiveAccess) can be consolidated on a single small virtual private server for a community with modest technical support; tools marked ”local” (Open Camera, Tropy, OpenRefine, VeraCrypt, GnuPG, SHA-256) run on volunteer devices and require no server. ArchiveBox and Conifer may run either locally or on the server depending on the crawling schedule.

Chapter 3: Bootstrapping Frameworks

This chapter provides ready-to-use starter frameworks for the five most critical risk and standardisation areas. Each framework begins with a conceptual overview and is followed by a comprehensive schema expressed as field-level tables.

3.1 Data Sovereignty Risk Assessment Framework

Community archives frequently rely on external platforms (cloud hosting, social media for collection, third-party transcription services). Each introduces sovereignty risk. The following framework allows a community to score and compare options before committing.

3.1.1 Risk Dimensions

#	Dimension	What It Asks	Weight
D1	Jurisdictional exposure	Where is data physically stored? Which legal regime can compel disclosure?	25%
D2	Vendor lock-in	Can all content and metadata be exported in open formats without loss?	20%
D3	Terms-of-service volatility	Can the provider unilaterally change terms, suspend accounts, or delete content?	15%
D4	Access continuity	What happens to the data if the provider shuts down or the community loses funding?	15%
D5	Insider/outsider access control	Can the community independently control who sees sealed or restricted records?	15%
D6	Metadata leakage	Does the platform collect or expose usage metadata (who viewed what, when)?	10%

3.1.2 Scoring Rubric

Per dimension, 1–5

Score	Meaning
5	Community retains full control; no external dependency
4	Minor exposure; mitigations exist and are documented
3	Moderate exposure; mitigations require ongoing effort
2	Significant exposure; community cannot independently verify

Score	Meaning
1	Unacceptable; community has no meaningful control

3.1.3 Decision Thresholds

- Weighted score **≥** 4.0: Acceptable for public-content hosting.
- Weighted score 3.0–3.9: Acceptable for restricted content only if a local offline master copy exists and is verified quarterly.
- Weighted score ¡ 3.0: Not acceptable for any transgression or disaster records; use only for fully public, already-published material.

3.1.4 Platform Assessment Record Schema

Every platform assessed must be logged using the following schema so that decisions are auditable and comparable across time.

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`assessmentId`	identifier	Yes	`asmt-YYYY-NNNN`	Unique assessment identifier	`asmt-2024-0007`
`platformName`	string	Yes	—	Name of the platform or service	`ExampleCloud S3`
`platformVendor`	string	Yes	—	Legal entity operating the platform	`Example Cloud Ltd.`
`assessmentDate`	date	Yes	ISO 8601	Date assessment completed	`2024-07-15`
`assessor`	string	Yes	—	Person or role conducting assessment	`Technical coordinator`
`reviewCycle`	enum	Yes	`annual`, `biannual`, `on-terms-change`	When the assessment is next due	`annual`
`d1Score`	integer	Yes	1–5	Jurisdictional exposure score	`3`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`d1Notes`	string	Yes	—	Evidence for D1 score (storage locations, treaties)	`EU + US regions; MLAT with home country`
`d2Score`	integer	Yes	1–5	Vendor lock-in score	`4`
`d2Notes`	string	Yes	—	Evidence for D2 score (export test results)	`Full CSV/JSON export verified 2024-07-10`
`d3Score`	integer	Yes	1–5	Terms-of-service volatility score	`4`
`d3Notes`	string	Yes	—	Evidence for D3 score (ToS archive reference)	`ToS archived with OpenTimestamps`
`d4Score`	integer	Yes	1–5	Access continuity score	`3`
`d4Notes`	string	Yes	—	Evidence for D4 score (exit plan)	`Exit plan documented; migration within 90d`
`d5Score`	integer	Yes	1–5	Access control score	`3`
`d5Notes`	string	Yes	—	Evidence for D5 score (per-record permissions)	`Per-record ACLs supported`
`d6Score`	integer	Yes	1–5	Metadata leakage score	`2`
`d6Notes`	string	Yes	—	Evidence for D6 score (analytics policy)	`Usage analytics on by default; opt-out`
`weightedScore`	decimal	Yes	1.00–5.00	Computed: Σ(score × weight)	`3.35`
`decision`	enum	Yes	`approved-public`, `approved-`	Outcome per Section 3.1.3	`approved-restricted`

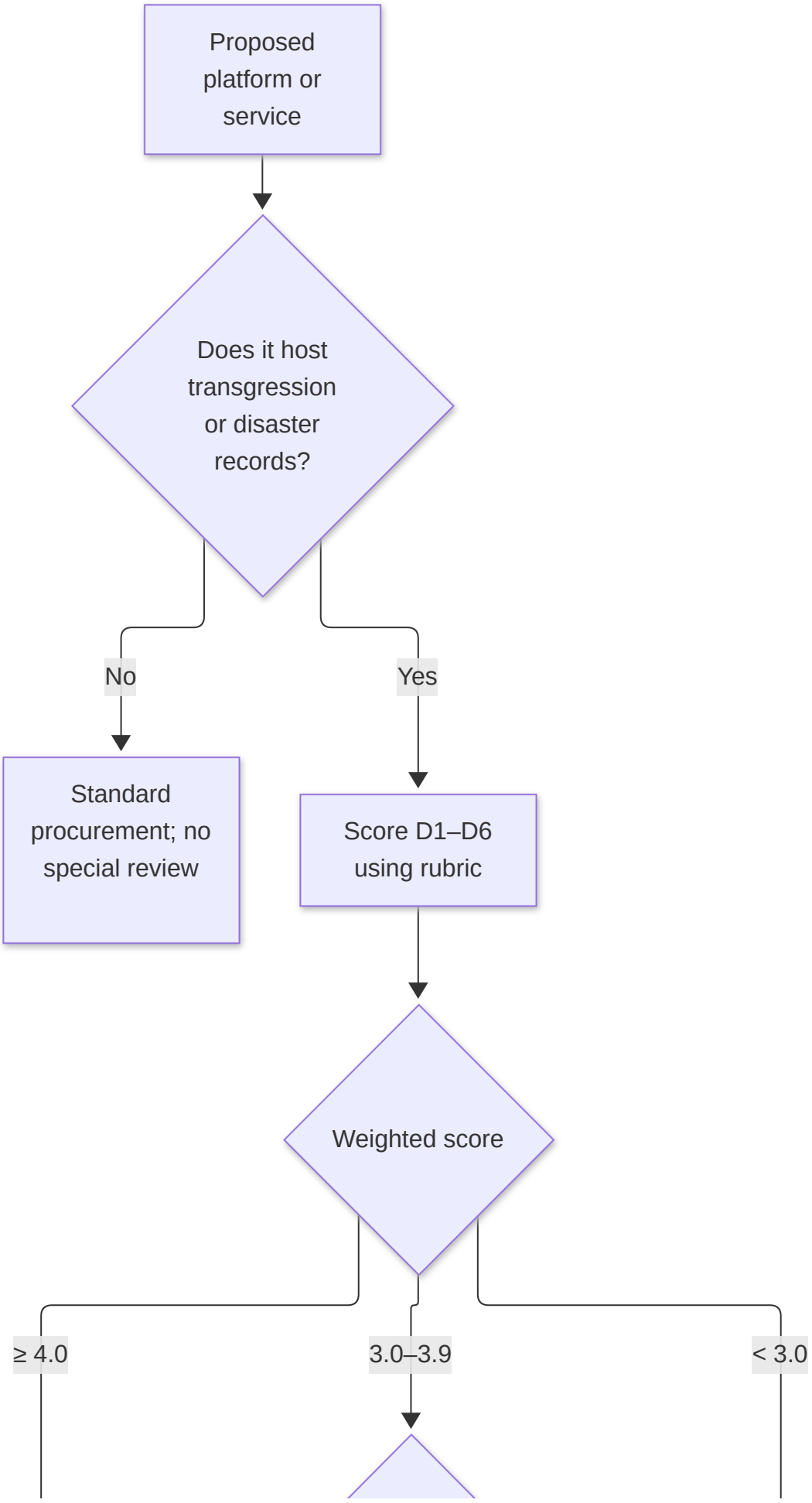
Field	Type	Required	Controlled Vocabulary / Format	Description	Example
			`restricted`, `rejected`		
`conditions`	string	No	—	Conditions attached to approval	`Offline master verified quarterly`
`nextReviewDate`	date	Yes	ISO 8601	Date of next assessment	`2025-07-15`

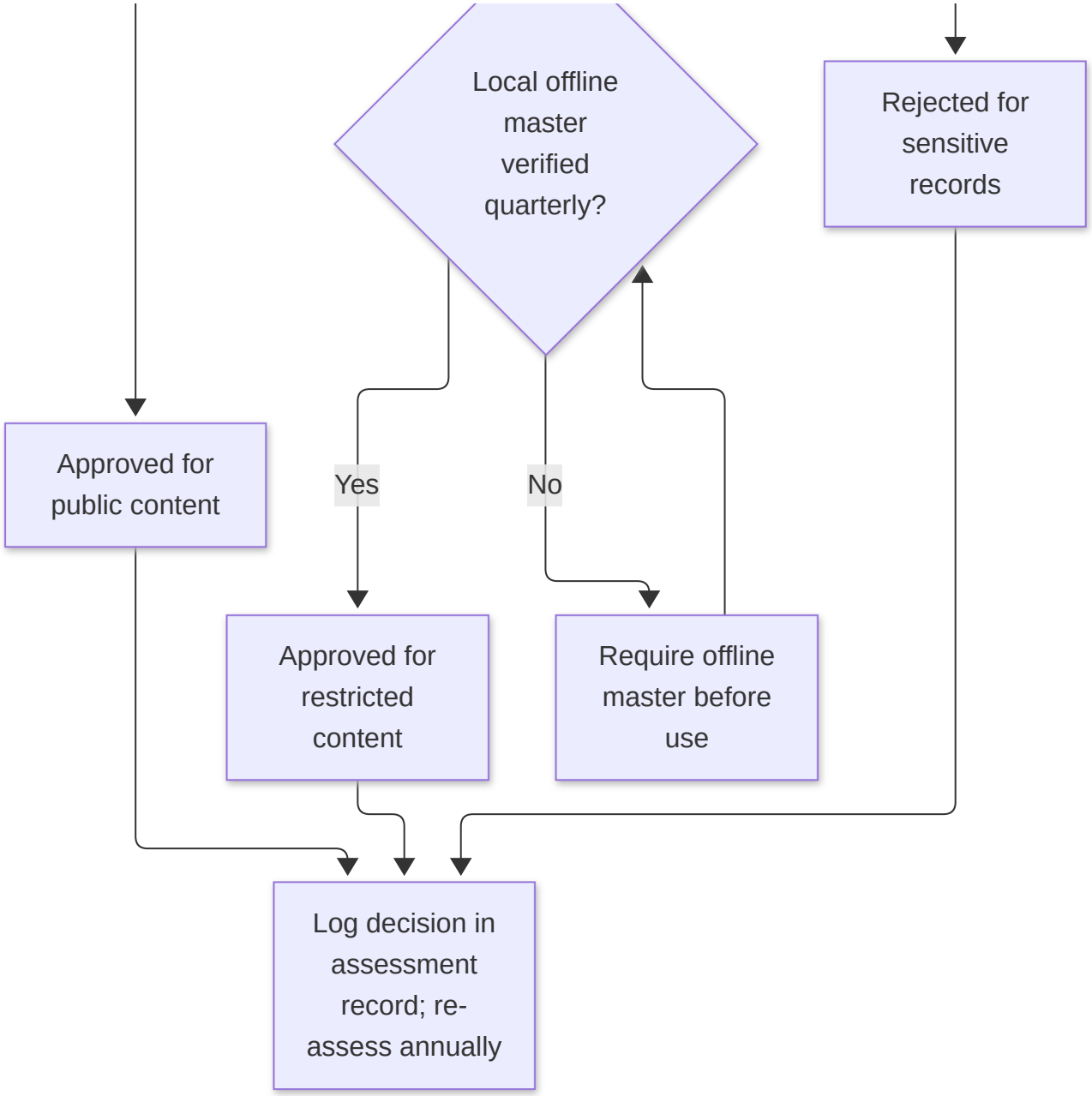
3.1.5 Bootstrap Checklist

Copy and complete before adopting any platform

```
[ ] D1: Identify storage jurisdiction(s). List any mutual legal assistance treaties
       that could compel disclosure.
[ ] D2: Test full export. Confirm metadata survives in CSV/JSON with no proprietary
fields.
[ ] D3: Archive the current terms of service with a timestamp (use OpenTimestamps).
[ ] D4: Document an exit plan: where will data go, who executes migration, by when?
[ ] D5: Confirm the platform supports per-record access levels, not just per-account.
[ ] D6: Review privacy policy for analytics/tracking; prefer providers that allow
disabling.
[ ] Record weighted score and date using the schema in 3.1.4. Re-assess annually or upon
any
    terms change.
```

3.1.6 Risk Assessment Flow


Proposed platform or service
Does it host transgression or disaster records?
No
Yes
Standard procurement; no special review
Score D1–D6 using rubric
Weighted score
≥ 4.0
3.0–3.9
< 3.0



3.2 Controlled Vocabulary in a Multilingual Setting

Communities are rarely monolingual. A controlled vocabulary must accommodate multiple languages without forcing false equivalence. The framework below uses a *concept-based, language-tagged* structure: each concept has one stable identifier, and each language variant is a label attached to that identifier.

3.2.1 Structural Rules

1. One concept, one identifier. Use a persistent URI (e.g., `cv:transgression.theft`) as the authoritative key. Labels in any language are attached to it.
2. Language tags are mandatory. Every label carries an ISO 639-1/639-3 tag (e.g., `en`, `zh`, `yue` for Cantonese, `ug` for Uyghur).
3. No preferred language is privileged in storage. Display preference is a user setting, not a data property.

4. Scope notes are per-language. A concept's definition may differ subtly across languages; record each language's scope note separately rather than translating one master definition.
5. Non-equivalence is flagged. Where a term in one language has no exact equivalent, mark the mapping as `closeMatch` (not `exactMatch`) and record the residual difference in a note.
6. Every term has a status and a version. Terms are never deleted; they are deprecated with a date and a pointer to the replacement term (if any).

3.2.2 Comprehensive Concept Schema (SKOS-compatible)

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`id`	URI	Yes	`cv:` namespace	Persistent, language-neutral identifier; never changes	`cv:disaster.flood`
`prefLabel@[lang]`	string	Yes	ISO 639 tag on field name	One preferred label per language; at least one language required	`prefLabel@en = flood`; `prefLabel@zh = 洪灾` (hóng zāi)
`altLabel@[lang]`	string[]	No	ISO 639 tag on field name	Alternative labels, synonyms, spelling variants, historical names	`altLabel@en = inundation`
`hiddenLabel@[lang]`	string[]	No	ISO 639 tag on field name	Misspellings or search-only variants not for display	`hiddenLabel@en = flod`
`definition@[lang]`	string	Yes	ISO 639 tag on field name	Full definition in each language; may differ across languages	`definition@en = "Water covering normally dry land…"`
`scopeNote@[lang]`	string	No	ISO 639 tag on field name	Usage guidance: what the term includes and excludes	`scopeNote@en = "Excludes tsunamis and storm surges."`
`example@[lang]`	string[]	No	ISO 639 tag on field name	Concrete examples of appropriate use	`example@en = "2024-07-12 river flood, Lower Ward"`
`broader`	URI[]	No	`cv:` namespace	Parent concept(s) in the hierarchy	`cv:disaster.weather`
`narrower`	URI[]	No	`cv:` namespace	Child concept(s); usually inferred, may be explicit	`cv:disaster.weather.flood.flash`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`related`	URI[]	No	`cv:` namespace	Associative (non-hierarchical) links	`cv:response.official.emergency-declaration`
`exactMatch`	URI[]	No	External vocabulary URI	Equivalent concept in another vocabulary (e.g., Wikidata, UNESCO thesaurus)	`https://www.wikidata.org/wiki/Q8068`
`closeMatch`	URI[]	No	External vocabulary URI	Near-equivalent concept; residual difference must be noted	`closeMatch = external:deluge`
`broadMatch`	URI[]	No	External vocabulary URI	External concept that is broader	—
`narrowMatch`	URI[]	No	External vocabulary URI	External concept that is narrower	—
`mappingNote@[lang]`	string	No	ISO 639 tag on field name	Records residual non-equivalence between languages or vocabularies	`"zh 洪灾 implies large-scale; yue 水浸 may be local."`
`status`	enum	Yes	`proposed`, `approved`, `deprecated`	Lifecycle status	`approved`
`proposedBy`	string	Yes	—	Person or role who proposed the term	`Volunteer collector A`
`proposedDate`	date	Yes	ISO 8601	Date proposed	`2024-06-01`
`approvedBy`	string	No	—	Committee decision reference	`Committee minutes 2024-06-15`
`approvedDate`	date	No	ISO 8601	Date approved	`2024-06-15`
`deprecatedDate`	date	No	ISO 8601	Date deprecated (if applicable)	—
`replacedBy`	URI	No	`cv:` namespace	Successor term, if deprecated	—

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`deprecationNote@[lang]`	string	No	ISO 639 tag on field name	Reason for deprecation and migration guidance	—
`vocabularyVersion`	string	Yes	Semantic version `MAJOR.MINOR.PATCH`	Version of the vocabulary in which this term last changed	`1.2.0`
`createdDate`	date	Yes	ISO 8601	Date the concept was first created	`2024-06-01`
`lastModifiedDate`	date	Yes	ISO 8601	Date of last change to any field	`2024-06-15`
`changeNote@[lang]`	string[]	No	ISO 639 tag on field name	History of substantive changes	`"Scope note narrowed 2024-06-15."`

3.2.3 Worked Example: Flood Concept

Field	Value
`id`	`cv:disaster.weather.flood`
`prefLabel@en`	`flood`
`prefLabel@zh`	`洪灾` (hóng zāi)
`prefLabel@yue`	`水浸` (seoi2 zam3)
`altLabel@en`	`inundation`
`definition@en`	"Water covering normally dry land; includes river, coastal, and flash flooding."
`definition@zh`	"水流淹没通常干燥的陆地；包括河流、沿海和山洪。"
`scopeNote@en`	"Excludes tsunamis and storm surges (see separate concepts)."
`broader`	`cv:disaster.weather`
`related`	`cv:response.official.emergency-declaration`
`exactMatch`	`https://www.wikidata.org/wiki/Q8068`
`mappingNote@en`	"zh 洪灾 implies large-scale; yue 水浸 may be local."

Field	Value
`status`	`approved`
`vocabularyVersion`	`1.2.0`

3.2.4 Governance Rules

- Term proposals come from collectors and community members; each proposal must include at least one language label and a scope note in that language, using the schema in 3.2.2.
- Review by the archive committee (see Chapter 4); approval requires that no existing concept already covers the meaning.
- Deprecation, not deletion: retired terms are marked `deprecated`, given a `deprecatedDate`, and linked to a `replacedBy` term where applicable; they retain their identifier so historical records remain valid.
- Versioning: the vocabulary is exported as a dated SKOS file (RDF/XML or Turtle) and stored with the archive; every record references the `vocabularyVersion` used at the time of cataloguing.

3.2.5 Bootstrap Starter Set

Extend as needed

```
cv:transgression.legal.*        (e.g., theft, fraud, assault)
cv:transgression.moral.*        (e.g., betrayal, exploitation)
cv:transgression.political.*    (e.g., suppression, corruption)
cv:disaster.environmental.*     (e.g., pollution, deforestation)
cv:disaster.weather.*           (e.g., flood, drought, typhoon)
cv:disaster.public-health.*     (e.g., outbreak, contamination)
cv:response.official.*          (e.g., statement, inquiry, compensation)
cv:response.community.*         (e.g., mutual aid, protest, memorial)

cv:actor.individual.*
cv:actor.state-agency.*
cv:actor.state-official.*
cv:actor.organisation.*
cv:actor.corporation.*
cv:actor.collective.*
cv:actor.unknown
cv:actor.natural

cv:role.primary-cause
cv:role.contributing
cv:role.failed-to-prevent
cv:role.responded
cv:role.benefited
cv:role.unknown

cv:scale.fam
cv:scale.com
cv:scale.reg
```

```
cv:scale.nat
cv:scale.int
```

3.3 Unified Event Classification Tags (Multi-Scale)

The same event may be recorded at family, community, regional, national, or international scale. A single flat tag list causes collisions (e.g., "flood" at family scale vs. "flood" at national scale). The framework below uses a *scale-prefixed, faceted tag* structure so that any record can be filtered by scale without losing the event type.

3.3.1 Tag Syntax

```
[scale].[category].[subcategory]
```

Component	Allowed Values	Example
`scale`	`fam` (family), `com` (community), `reg` (regional), `nat` (national), `int` (international)	`com`
`category`	`transgression` or `disaster`	`disaster`
`subcategory`	from controlled vocabulary (Section 3.2)	`weather.flood`

Example: `com.disaster.weather.flood` = a flood affecting a community.

3.3.2 Comprehensive Event Record Schema

Every event recorded in the archive must be described using the following schema. This schema sits alongside the attribution schema (Section 3.4.2) and the archival package schema (Section 3.5.3).

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`eventGroupId`	identifier	Yes	`evt-YYYY-MMDD-slug-NN`	Unique per event; shared across all records of the same event	`evt-2024-0712-flood-01`
`recordId`	identifier	Yes	`rec-NNNN`	Unique per record	`rec-0045`
`title@[lang]`	string	Yes	ISO 639 tag on field name	Short human-readable title in each language	`title@en = "July 2024`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
					`Lower Ward flood"`
`description@[lang]`	string	Yes	ISO 639 tag on field name	Narrative description; minimum 100 characters recommended	`"River overtopped embankment…"`
`primaryTag`	string	Yes	`[scale].[category].[subcategory]`	Primary classification tag	`com.disaster.weather.flood`
`affects`	enum[]	No	`fam`, `com`, `reg`, `nat`, `int`	Other scales affected; must not duplicate `primaryTag` scale	`["fam","reg"]`
`secondaryTags`	string[]	No	`[scale].[category].[subcategory]`	Additional classifications (e.g., a flood that was also a political failure)	`["com.transgression.political.negligence"]`
`eventStartDate`	date	Yes	ISO 8601	Date event began (or earliest known)	`2024-07-12`
`eventStartTime`	time	No	ISO 8601	Time event began, if known	`03:40`
`eventEndDate`	date	No	ISO 8601	Date event ended (or latest known)	`2024-07-18`
`datePrecision`	enum	Yes	`exact`, `day`, `month`, `year`, `decade`, `circa`	Precision of the dates recorded	`day`
`locationName@[lang]`	string	Yes	ISO 639 tag on field name	Place name as locally used	`locationName@en = "Lower Ward"`
`locationGeo`	geomet	No	GeoJSON Point/Polygon	Coordinates or boundary; WGS84	`{"type":"Point","coordinates"`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`m`	ry				`:[…]}`
`locationAdmin`	string[]	No	—	Administrative hierarchy (village, district, region, country)	`["Lower Ward","Riverside District","Example Region"]`
`locationGazetteerRef`	URI	No	Gazetteer URI	Link to authoritative gazetteer entry	`https://www.geonames.org/…`
`severity`	enum	No	`minor`, `moderate`, `major`, `catastrophic`, `unknown`	Severity assessment	`major`
`severityBasis`	string	No	—	Basis for severity assessment (casualties, damage, area)	`"3 deaths; 40 homes damaged"`
`casualtiesDeaths`	integer	No	≥ 0	Known deaths	`3`
`casualtiesInjured`	integer	No	≥ 0	Known injuries	`12`
`casualtiesDisplaced`	integer	No	≥ 0	Known displacements	`140`
`casualtiesMissing`	integer	No	≥ 0	Known missing	`0`
`casualtySource`	string	No	—	Source of casualty figures	`"Official statement 2024-07-20"`
`economicLos`	decimal	No	—	Estimated economic loss	`250000.00`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`s`					
`economicLossCurrency`	string	No	ISO 4217	Currency of loss estimate	`USD`
`economicLossBasis`	string	No	—	Method and source of estimate	`"Insurance assessment"`
`status`	enum	Yes	`ongoing`, `resolved`, `dormant`, `unknown`	Current status of the event	`resolved`
`accessLevel`	enum	Yes	`public`, `restricted`, `sealed`	Access level (see Section 3.5.2)	`restricted`
`sealedUntil`	date	No	ISO 8601	Date seal expires; required if `accessLevel = sealed`	`2044-07-18`
`consentStatus`	enum	Yes	`informed`, `dynamic`, `withdrawn`, `not-applicable`	Consent status for human-subject content	`dynamic`
`sensitivityFlags`	enum[]	No	`minors`, `sexual-violence`, `self-harm`, `official-secrets`, `active-investigation`, `inter-community-conflict`	Flags triggering additional review	`["minors"]`
`contextNote@[lang]`	string	No	ISO 639 tag on field name	Contextual notes to prevent misreading of isolated records	`"Casualty figures disputed by community group."`
`relatedRecordIds`	identifier[]	No	`rec-NNNN`	Other records related to this event	`["rec-0046","rec-0061"]`
`sourceRefs`	URI[]	No	—	External sources (official reports,	`["https://…/report.pdf"]`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
				news)	
`language`	string	Yes	ISO 639-1/639-3	Primary language of the record	`en`
`collectorId`	string	Yes	—	Collector identifier (pseudonymous permitted)	`volunteer-0042`
`collectorRole`	enum	Yes	`volunteer`, `staff`, `partner`, `automated`	Role of collector	`volunteer`
`collectionDate`	date	Yes	ISO 8601	Date of collection	`2024-07-13`
`collectionMethod`	enum	Yes	`field-notes`, `interview`, `photo`, `video`, `document-scan`, `web-capture`, `sensor`	Method of collection	`photo`
`hashSha256`	string	Yes	64 hex characters	SHA-256 hash of the archival package	`a3f5…`
`hashTimestamp`	string	No	OpenTimestamps proof	Blockchain anchor of the hash	`…`
`signatureRef`	string	No	—	Reference to PGP/GPG signature	`sig-0042`
`vocabularyVersion`	string	Yes	Semantic version	Controlled vocabulary version used	`1.2.0`
`schemaVersion`	string	Yes	Semantic version	Version of this schema	`1.2.0`
`createdDate`	datetime	Yes	ISO 8601	System creation timestamp	`2024-07-13T10:00:00Z`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`lastModifiedDate`	datetime	Yes	ISO 8601	Last modification timestamp	`2024-07-20T14:30:00Z`
`auditTrailRef`	string	Yes	—	Reference to audit log entry	`audit-2024-07-13-0045`

3.3.3 Cross-Scale Linking Rule

A single event may generate records at multiple scales. Rather than duplicating tags, each record carries one primary scale tag plus an optional `affects` field listing other scales:

Field	Example
`primaryTag`	`com.disaster.weather.flood`
`affects`	`fam`, `reg`
`eventGroupId`	`evt-2024-0712-flood-01` (shared across all records of the same event)

This allows queries such as "all records of events that affected families but were primarily recorded at community scale" without ambiguity.

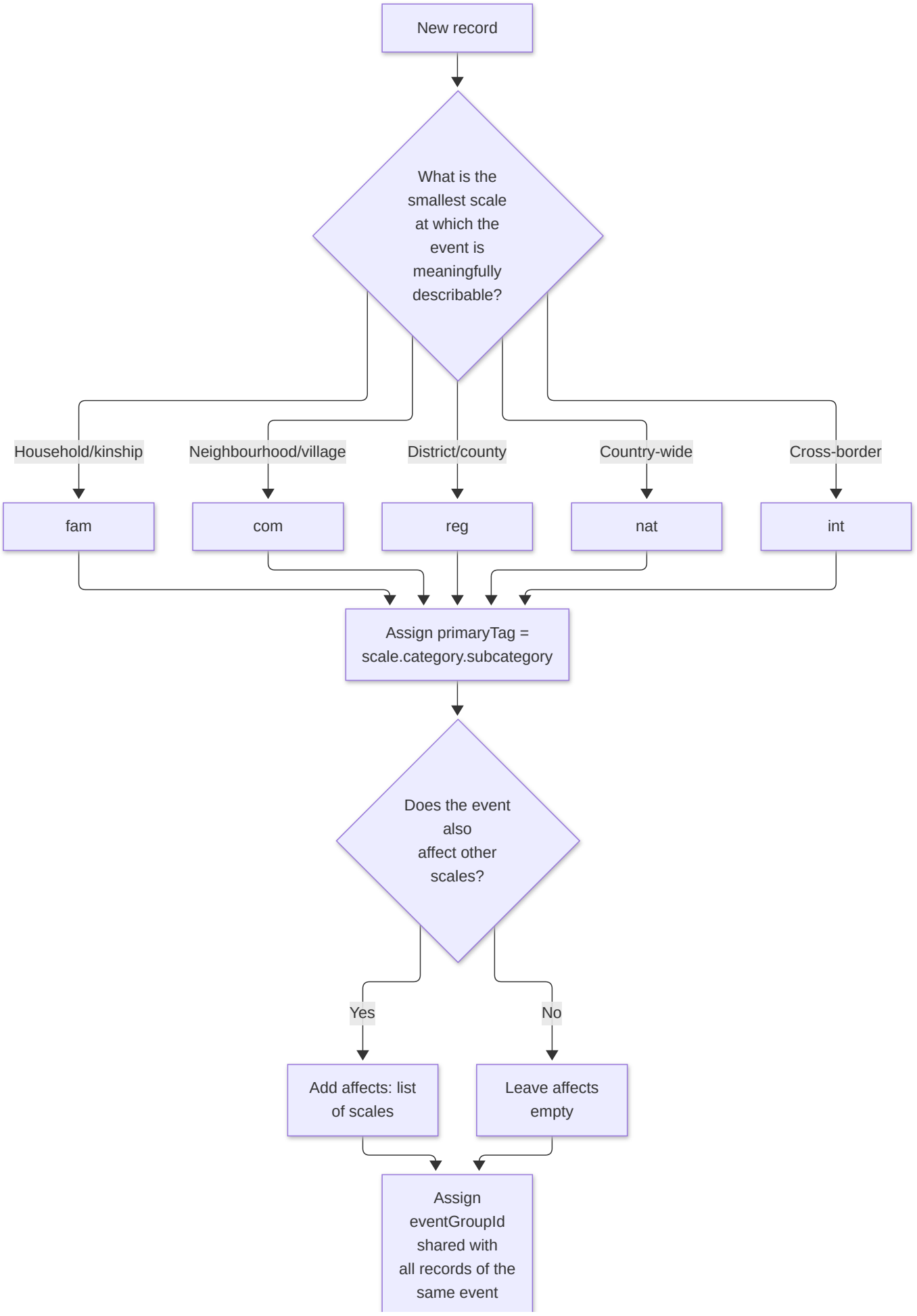
Attribution to actors is handled separately in Section 3.4. A single event may have multiple attribution entries linked by the same `eventGroupId`.

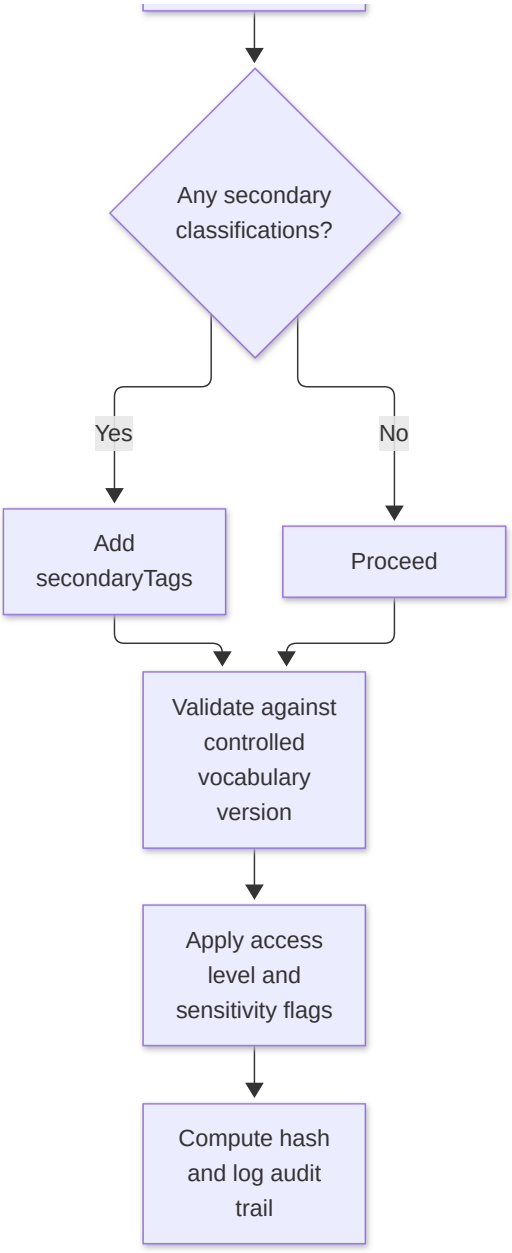
3.3.4 Scale Definitions

Scale	Definition	Typical Recorder
`fam`	Affecting one household or kinship group	Family member, community worker
`com`	Affecting a neighbourhood, village, or shared institution	Community volunteer, local organisation
`reg`	Affecting a district, county, or multi-community area	Local archive, NGO, journalist
`nat`	Affecting a country as a whole or requiring national response	National archive, researcher

Scale	Definition	Typical Recorder
`int`	Cross-border or internationally significant	Consortium, academic institution

3.3.5 Classification Decision Flow


New record
What is the smallest scale at which the event is meaningfully describable?
Household/kinship
Neighbourhood/village
District/county
Country-wide
Cross-border
fam
com
reg
nat
int
Assign primaryTag = scale.category.subcategory
Does the event also affect other scales?
Yes
No
Add affects: list of scales
Leave affects empty
Assign eventGroupId shared with all records of the same event



3.3.6 Validation Rules

- `primaryTag` must exist in the current vocabulary version.
- `affects` values must be from the allowed scale list and must not duplicate `primaryTag`'s scale.
- `secondaryTags` must exist in the current vocabulary version and must not duplicate `primaryTag`.
- `eventGroupId` must be unique per event, not per record; multiple records sharing an event share the same ID.
- `sealedUntil` is mandatory when `accessLevel = sealed`.
- `sensitivityFlags` must trigger the corresponding mitigation in Section 3.6.1.
- Rejected tags are logged with a reason; the record is held in a `pending-classification` queue rather than discarded.

3.4 Attribution of Event to Actor (Individual, State, Organisation)

Event classification (Section 3.3) records *what* happened and at *what scale*. Attribution records *who or what is alleged to have caused, enabled, or failed to prevent* the event. Attribution is distinct from classification because it is contestable, often plural, and legally sensitive. The framework below treats attribution as a structured, evidence-linked, and dispute-aware field rather than free text.

3.4.1 Structural Rules

1. Attribution is an allegation, not a finding. Every attribution record carries an `attributionStatus` field: `alleged`, `admitted`, `found-by-inquiry`, `found-by-court`, `disputed`, or `unknown`. No record may carry an attribution without a status.
2. Multiple actors per event are permitted and expected. A disaster may have a primary corporate actor, a negligent regulator, and an individual official. Each is a separate attribution entry linked to the same `eventGroupId`.
3. Actor type is a controlled field. Allowed values: `individual`, `state-agency`, `state-official`, `organisation` (NGO, religious body, political party), `corporation`, `collective` (e.g., unnamed group, mob), `unknown`, `natural` (for disasters with no attributable human actor).
4. Attribution must link to evidence. Each attribution entry references at least one record in the archive (testimony, document, official finding) that supports it. An attribution with no supporting record is held in `pending-evidence` and is not publicly displayed.
5. Right of reply is built in. Any named individual or organisation may submit a response, which is stored as a linked record and displayed alongside the attribution. The response does not remove the attribution; it annotates it.

3.4.2 Comprehensive Attribution Schema

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`attributionId`	identifier	Yes	`attr-YYYY-MMDD-slug-NN-aN`	Unique per attribution entry	`attr-2024-0712-flood-01-a1`
`eventGroupId`	identifier	Yes	`evt-…`	Links to the event (Section 3.3.2)	`evt-2024-0712-flood-01`
`actorType`	enum	Yes	`individual`, `state-agency`, `state-official`, `organisation`, `corporation`, `collective`, `unknown`, `natural`	Type of actor attributed	`corporation`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`actorName@[lang]`	string	Yes	ISO 639 tag on field name	Name of actor as commonly used; may be redacted in public view	`actorName@en = "Example Water Ltd."`
`actorLegalName`	string	No	—	Registered legal name, if different	`"Example Water Limited"`
`actorIdentifier`	string	No	LEI, company registry number, national ID	Persistent identifier for the actor	`LEI:5493001KJTIIGC8Y1R12`
`actorIdentifierScheme`	enum	No	`LEI`, `company-registry`, `national-id`, `local-registry`, `none`	Scheme of `actorIdentifier`	`LEI`
`actorRole`	enum	Yes	`primary-cause`, `contributing`, `failed-to-prevent`, `responded`, `benefited`, `unknown`	Role in the event (3.4.3)	`primary-cause`
`actorRoleDetail@[lang]`	string	No	ISO 639 tag on field name	Free-text clarification of role	`"Operated the embankment gate"`
`attributionStatus`	enum	Yes	`alleged`, `admitted`, `found-by-inquiry`, `found-by-court`, `disputed`, `unknown`	Evidentiary status of the attribution	`alleged`
`attributionBasis@[lang]`	string	Yes	ISO 639 tag on field name	Short statement of what the attribution is based on	`"Witness testimony and maintenance logs"`
`evidenceRefs`	identifier[]	Yes	`rec-NNNN`	At least one required; each must exist in the archive	`["rec-0045","rec-0061"]`
`evidenceStrength`	enum	No	`weak`, `moderate`, `strong`	Assessed strength of the evidence	`moderate`

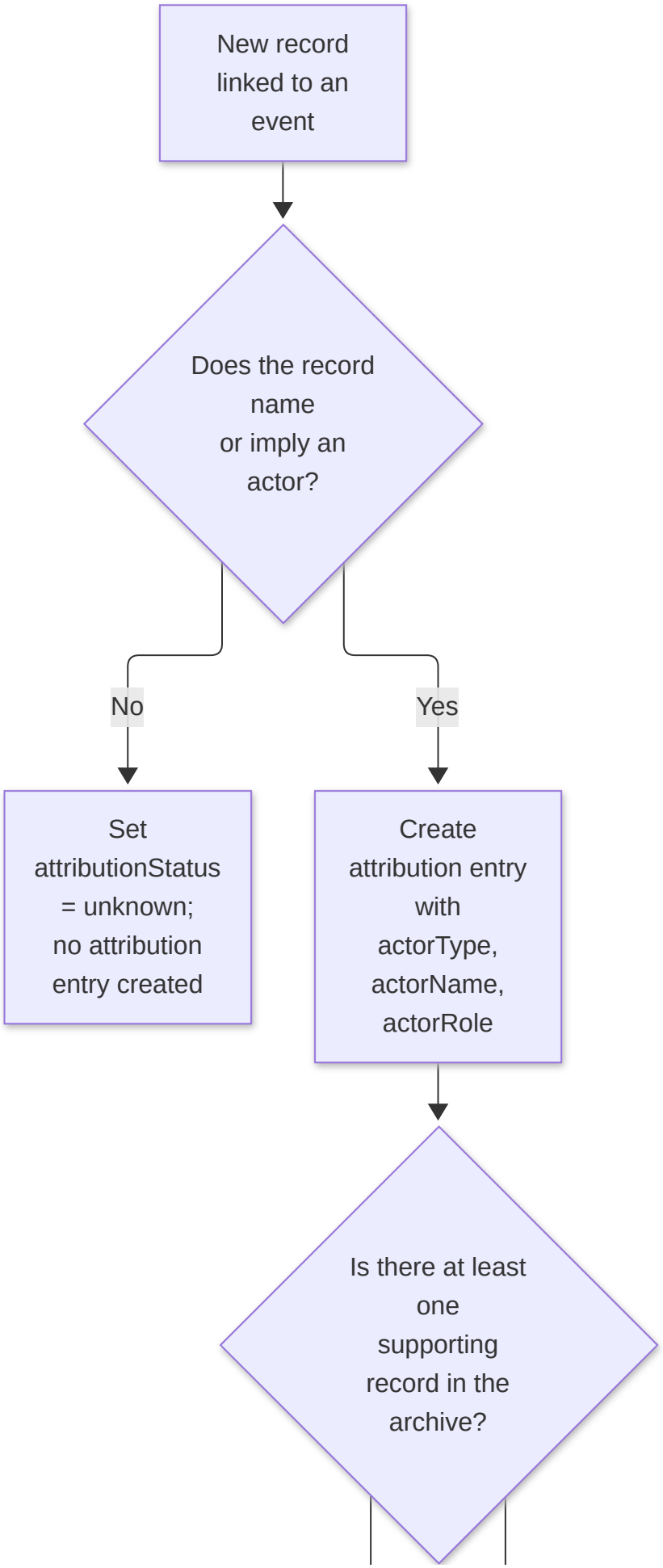
Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`responseRefs`	identifier[]	No	`rec-NNNN`	Actor's reply, if any	`["rec-0078"]`
`responseRequestedDate`	date	No	ISO 8601	Date right of reply was communicated	`2024-07-25`
`responseReceivedDate`	date	No	ISO 8601	Date response was received	`2024-08-01`
`legalReviewStatus`	enum	Yes	`not-required`, `pending`, `cleared`, `flagged`	Legal review status for the attribution	`cleared`
`legalReviewRef`	string	No	—	Reference to legal review record	`legal-2024-07-22-003`
`confidence`	enum	No	`low`, `medium`, `high`	Committee's confidence in the attribution (distinct from evidence strength)	`medium`
`firstRecordedDate`	date	Yes	ISO 8601	Date this attribution was first recorded	`2024-07-20`
`lastModifiedDate`	datetime	Yes	ISO 8601	Last modification timestamp	`2024-08-01T09:15:00Z`
`language`	string	Yes	ISO 639-1/639-3	Language of the attribution statement	`en`
`accessLevel`	enum	Yes	`public`, `restricted`, `sealed`	Access level of the attribution (may differ from the event record)	`public`

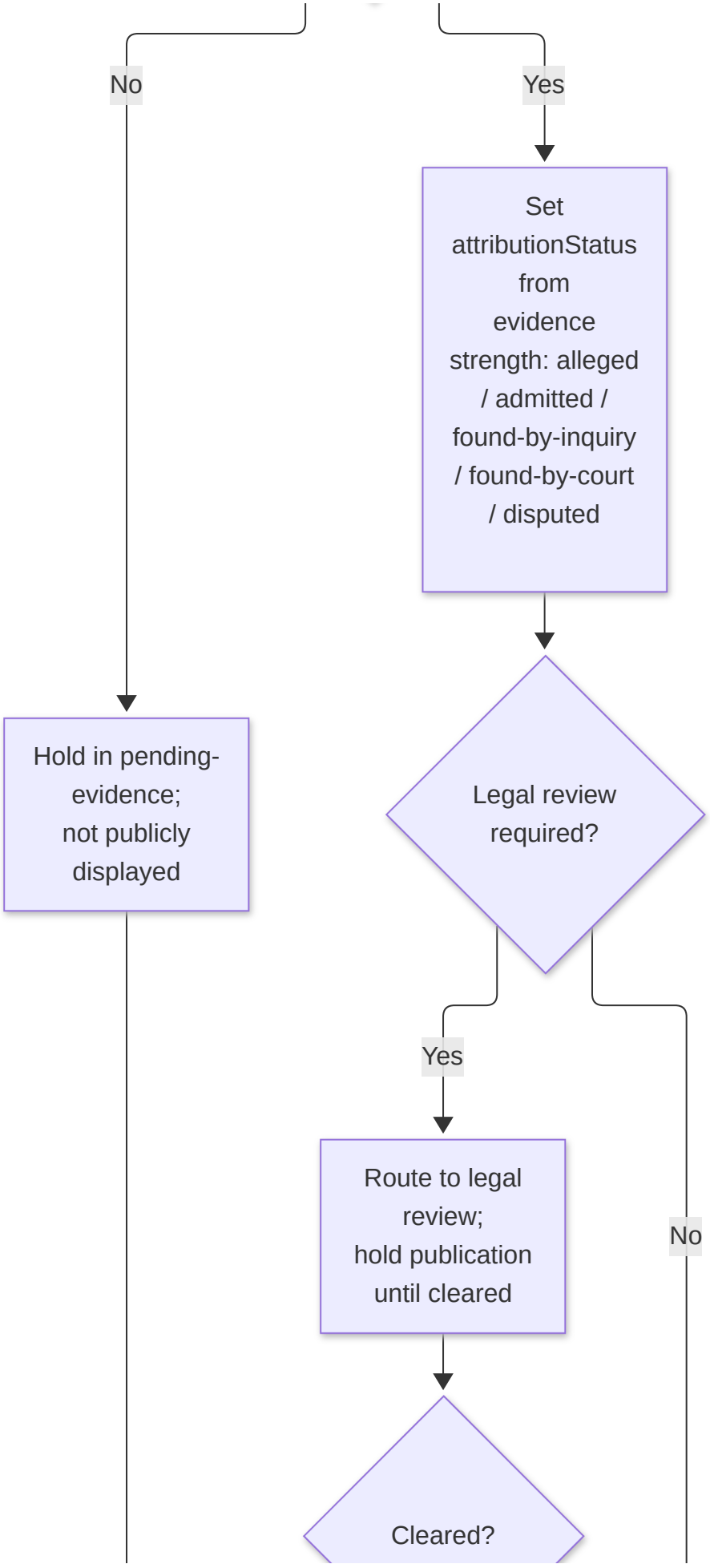
Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`contextNote@[lang]`	string	No	ISO 639 tag on field name	Context to prevent misreading	`"Actor disputes the finding; see response."`
`auditTrailRef`	string	Yes	—	Reference to audit log entry	`audit-2024-07-20-0012`

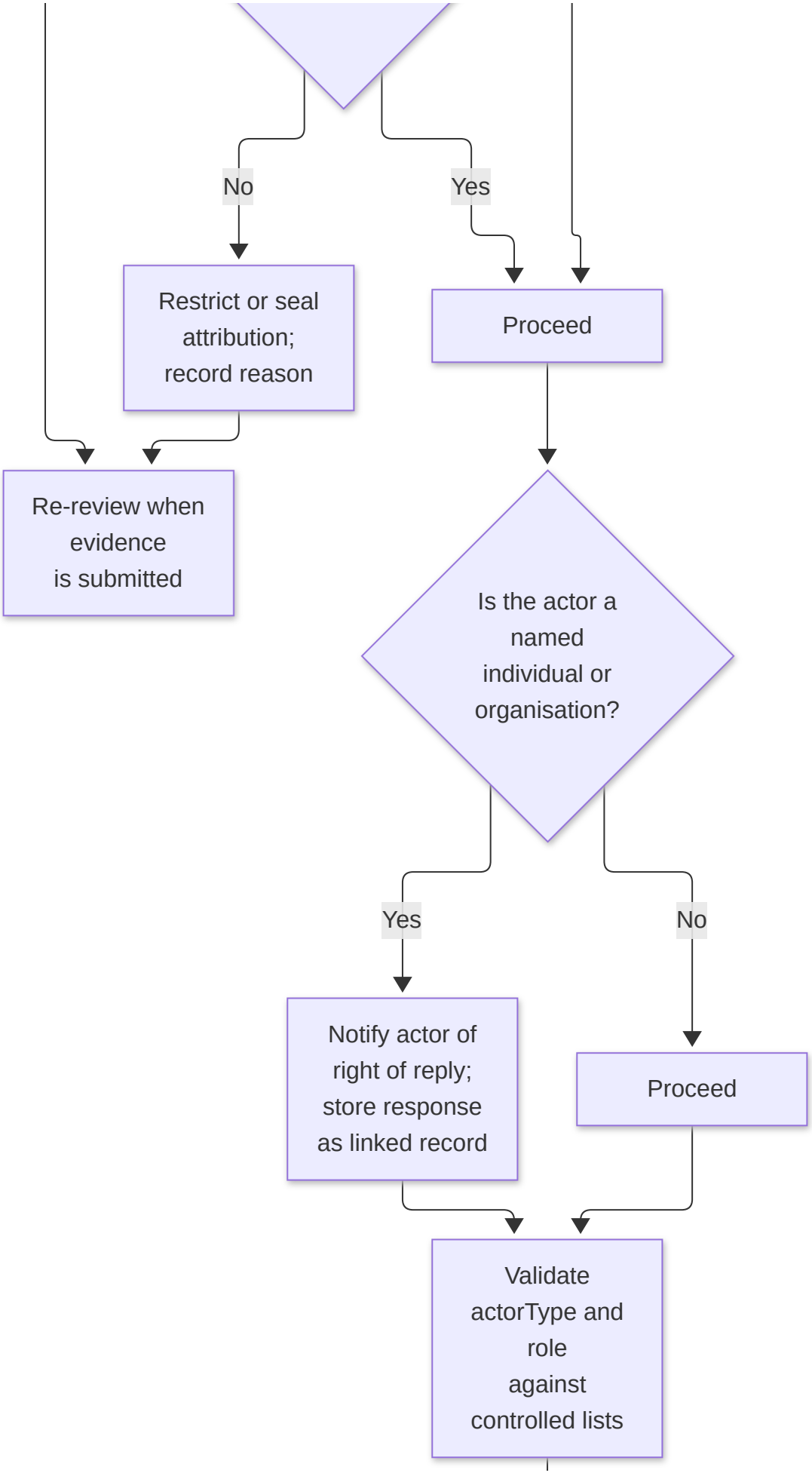
3.4.3 Role Vocabulary

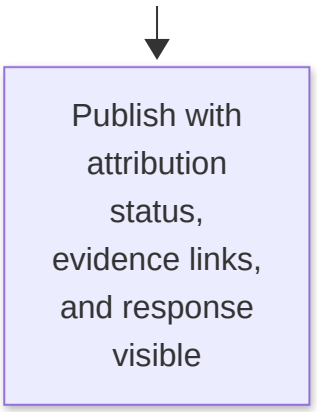
Role	Definition
`primary-cause`	Attributed as the principal agent of the event
`contributing`	Attributed as one of several contributing agents
`failed-to-prevent`	Attributed with a duty of care that was not met (regulators, officials)
`responded`	Attributed with the response (may be positive or negative)
`benefited`	Attributed with benefiting from the event without causing it
`unknown`	Actor not identified

3.4.4 Attribution Decision Flow


New record linked to an event
Does the record name or imply an actor?
No
Yes
Set attributionStatus = unknown; no attribution entry created
Create attribution entry with actorType, actorName, actorRole
Is there at least one supporting record in the archive?


No
Yes
Set attributionStatus from evidence strength: alleged / admitted / found-by-inquiry / found-by-court / disputed
Hold in pending-evidence; not publicly displayed
Legal review required?
Yes
Route to legal review; hold publication until cleared
No
Cleared?


No
Yes
Restrict or seal attribution; record reason
Proceed
Re-review when evidence is submitted
Is the actor a named individual or organisation?
Yes
No
Notify actor of right of reply; store response as linked record
Proceed
Validate actorType and role against controlled lists



3.4.5 Display Rule

Public display of an attribution must always show, in the same view: the `actorName`, the `attributionStatus`, the linked evidence records, and any `responseRefs`. An attribution may never be displayed in isolation from its status and evidence links. This prevents selective citation of allegations as if they were findings (see Section 4.3).

3.4.6 Relationship to Classification

Attribution does not replace or modify the event classification tags. A record may carry:

- `primaryTag`: `com.disaster.weather.flood`
- `affects`: `fam`, `reg`
- `eventGroupId`: `evt-2024-0712-flood-01`
- plus one or more attribution entries (e.g., `corporation` / `primary-cause` / `alleged`; `state-agency` / `failed-to-prevent` / `found-by-inquiry`).

Queries such as "all events where a state agency is attributed with failed-to-prevent and the status is found-by-inquiry" become possible without altering the classification scheme.

3.5 Archival Package and Record Schema

Every item deposited in the archive is wrapped in an archival package with its own metadata. This schema ensures that the "four-property testing" dimensions (Section 1.2) are computable from the metadata alone.

3.5.1 Comprehensive Archival Package Schema

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`packageId`	identifier	Yes	`pkg-YYYY-NNNN`	Unique package identifier	`pkg-2024-0045`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`recordId`	identifier	Yes	`rec-NNNN`	Record this package belongs to	`rec-0045`
`eventGroupId`	identifier	Yes	`evt-…`	Event this package documents	`evt-2024-0712-flood-01`
`contentType`	enum	Yes	`image`, `audio`, `video`, `document`, `dataset`, `web-archive`, `transcript`, `mixed`	Type of content in the package	`image`
`format`	string	Yes	IANA media type or format registry	File format(s)	`image/jpeg`
`formatVersion`	string	No	—	Version of the format	`1.0`
`fileCount`	integer	Yes	≥ 1	Number of files in the package	`3`
`totalSizeBytes`	integer	Yes	≥ 0	Total size of the package in bytes	`12582912`
`originalFilename`	string	No	—	Original filename(s) as collected	`IMG_20240712_034012.jpg`
`hashSha256`	string	Yes	64 hex characters	SHA-256 hash of the package manifest	`a3f5…`
`hashManifestRef`	string	Yes	—	Reference to the manifest file listing per-file hashes	`manifest-0045.txt`
`hashTimestamp`	string	No	OpenTimestamps proof	Blockchain anchor of the hash	`…`
`signatureRef`	string	No	—	Reference to PGP/GPG signature	`sig-0042`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`signerIdentity`	string	No	—	Identity of the signer	`volunteer-0042`
`collectedBy`	string	Yes	—	Collector identifier	`volunteer-0042`
`collectedDate`	date	Yes	ISO 8601	Date collected	`2024-07-12`
`collectedTime`	time	No	ISO 8601	Time collected	`03:40`
`collectionMethod`	enum	Yes	`field-notes`, `interview`, `photo`, `video`, `document-scan`, `web-capture`, `sensor`, `donation`	Method of collection	`photo`
`captureDevice`	string	No	—	Device used for capture	`Open Camera 1.53 on Pixel 6a`
`captureSoftware`	string	No	—	Software used, with version	`Open Camera 1.53`
`exifPresent`	boolean	No	—	Whether EXIF metadata is present	`true`
`exifStripped`	boolean	No	—	Whether EXIF was stripped (for privacy)	`false`
`gpsPresent`	boolean	No	—	Whether GPS coordinates are embedded	`true`
`gpsStripped`	boolean	No	—	Whether GPS was stripped for public release	`true`
`language`	string	No	ISO 639-1/639-3	Language of content (for text/audio)	`en`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`transcriptRef`	identifier	No	`rec-NNNN`	Reference to transcript record, if any	`rec-0046`
`accessLevel`	enum	Yes	`public`, `restricted`, `sealed`	Access level of the package	`restricted`
`accessConditions`	string	No	—	Conditions for restricted access	`"Committee approval required"`
`sealedUntil`	date	No	ISO 8601	Date seal expires	—
`consentStatus`	enum	Yes	`informed`, `dynamic`, `withdrawn`, `not-applicable`	Consent status	`dynamic`
`consentRef`	identifier	No	`rec-NNNN`	Reference to consent record	`rec-0047`
`sensitivityFlags`	enum[]	No	`minors`, `sexual-violence`, `self-harm`, `official-secrets`, `active-investigation`, `inter-community-conflict`, `personal-data`	Flags triggering additional review	`["personal-data"]`
`retentionPeriod`	string	No	ISO 8601 duration	Retention period before review	`P10Y`
`retentionReviewDate`	date	No	ISO 8601	Date of next retention review	`2034-07-12`
`preservationAction`	enum	No	`none`, `migration`, `emulation`, `normalisation`	Preservation action taken	`migration`
`preservationActionDate`	date	No	ISO 8601	Date of last preservation action	`2024-07-15`
`preservationAc`	string	No	—	Description of preservation	`"JPEG normalised`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`tionNote`				action	`to sRGB"`
`fixityCheckedDate`	date	No	ISO 8601	Date of last fixity check	`2024-07-15`
`fixityResult`	enum	No	`pass`, `fail`, `not-checked`	Result of last fixity check	`pass`
`replicationStatus`	enum	Yes	`single`, `replicated-onsite`, `replicated-offsite`, `replicated-both`	Backup replication status	`replicated-both`
`replicationRefs`	string[]	No	—	References to replica locations	`["server-A","M-DISC-01"]`
`schemaVersion`	string	Yes	Semantic version	Version of this schema	`1.2.0`
`createdDate`	datetime	Yes	ISO 8601	System creation timestamp	`2024-07-12T10:00:00Z`
`lastModifiedDate`	datetime	Yes	ISO 8601	Last modification timestamp	`2024-07-15T11:30:00Z`
`auditTrailRef`	string	Yes	—	Reference to audit log entry	`audit-2024-07-12-0045`

3.5.2 Access Level Definitions

Level	Definition	Who May Access
`public`	Fully open	Anyone
`restricted`	Access requires approval	Researchers, journalists, and community members approved by the committee
`sealed`	Closed until `sealedUntil`; then reviewed by committee	No one until expiry; committee after expiry

3.5.3 Audit Trail Schema

Every action on a record or package generates an audit entry.

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`auditId`	identifier	Yes	`audit-YYYY-MMDD-NNNN`	Unique audit entry identifier	`audit-2024-07-13-0045`
`timestamp`	datetime	Yes	ISO 8601	When the action occurred	`2024-07-13T10:15:00Z`
`actor`	string	Yes	—	Person or system performing the action	`volunteer-0042`
`actorRole`	enum	Yes	`volunteer`, `staff`, `committee`, `system`, `external`	Role of the actor	`volunteer`
`action`	enum	Yes	`create`, `update`, `delete-request`, `access-grant`, `access-revoke`, `seal`, `unseal`, `migrate`, `verify-fixity`, `export`, `publish`	Action performed	`create`
`targetType`	enum	Yes	`record`, `package`, `attribution`, `vocabulary`, `assessment`	Type of object acted upon	`package`
`targetId`	identifier	Yes	—	Identifier of the object acted upon	`pkg-2024-0045`
`beforeState`	string	No	—	State before the action (for updates)	`accessLevel = public`
`afterState`	string	No	—	State after the action	`accessLevel = restricted`

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`reason@[lang]`	string	No	ISO 639 tag on field name	Reason for the action	`"Committee decision 2024-07-13"`
`ipAddress`	string	No	IPv4/IPv6	IP address of actor (privacy-sensitive; may be hashed)	`hashed:…`
`previousEntryHash`	string	No	64 hex characters	Hash of previous audit entry (for chaining)	`b2e4…`
`entryHash`	string	Yes	64 hex characters	Hash of this entry	`c3f5…`

3.6 Safety Precautions Framework

Recording transgressions and disasters can expose collectors, witnesses, and the community to retaliation, legal action, physical harm, or re-traumatisation. The following framework is organised by *who is at risk* and *at what stage*.

3.6.1 Comprehensive Risk Matrix

Risk	Affected Party	Stage	Mitigation
Physical retaliation	Collector, witness	Collection	Avoid solo collection in contested spaces; use check-in protocol; never collect during active violence
Legal exposure (defamation, privacy, official-secrets laws)	Collector, committee, custodian	Collection, publication	Legal review before publishing identifiable allegations; default to restricted access; publish context notes
Defamation or reprisal claim by a named actor	Committee, custodian, collector	Publication	Attribution status field mandatory; evidence link mandatory; right of reply built in; legal review before publishing `alleged` attributions against named individuals

Risk	Affected Party	Stage	Mitigation
Digital surveillance	All parties	Collection, storage	Use encrypted messaging for coordination; avoid collecting on devices tied to real identity where risk is high; strip GPS from files intended for public release
Device seizure	Collector	Collection, storage	Keep no single device holding both identity-linked and sensitive material; use encrypted volumes with plausibly deniable passphrases where lawful
Re-traumatisation	Witness, survivor	Interview, publication	Trained interviewers; right to stop and withdraw; dynamic consent (Section 4.2); access to support services
Community polarisation	Whole community	Publication	Committee review; context notes; sealed periods for incendiary material
Insider threat (committee or volunteer misuse)	All parties	All stages	Role separation; audit logs; two-person rule for sealed-record access
Data loss through technical failure	All parties	Storage	Off-site, heterogeneous backups (Section 1.3); quarterly fixity checks
Re-identification through combined metadata	Witness, survivor, collector	Publication	Strip GPS and EXIF from public files; aggregate or redact location data for sensitive events; review combinations of fields that could re-identify
Coercion of collector by authorities	Collector	Collection, storage	Do not store unencrypted copies; use duress codes where supported; document any coercion attempts as a record

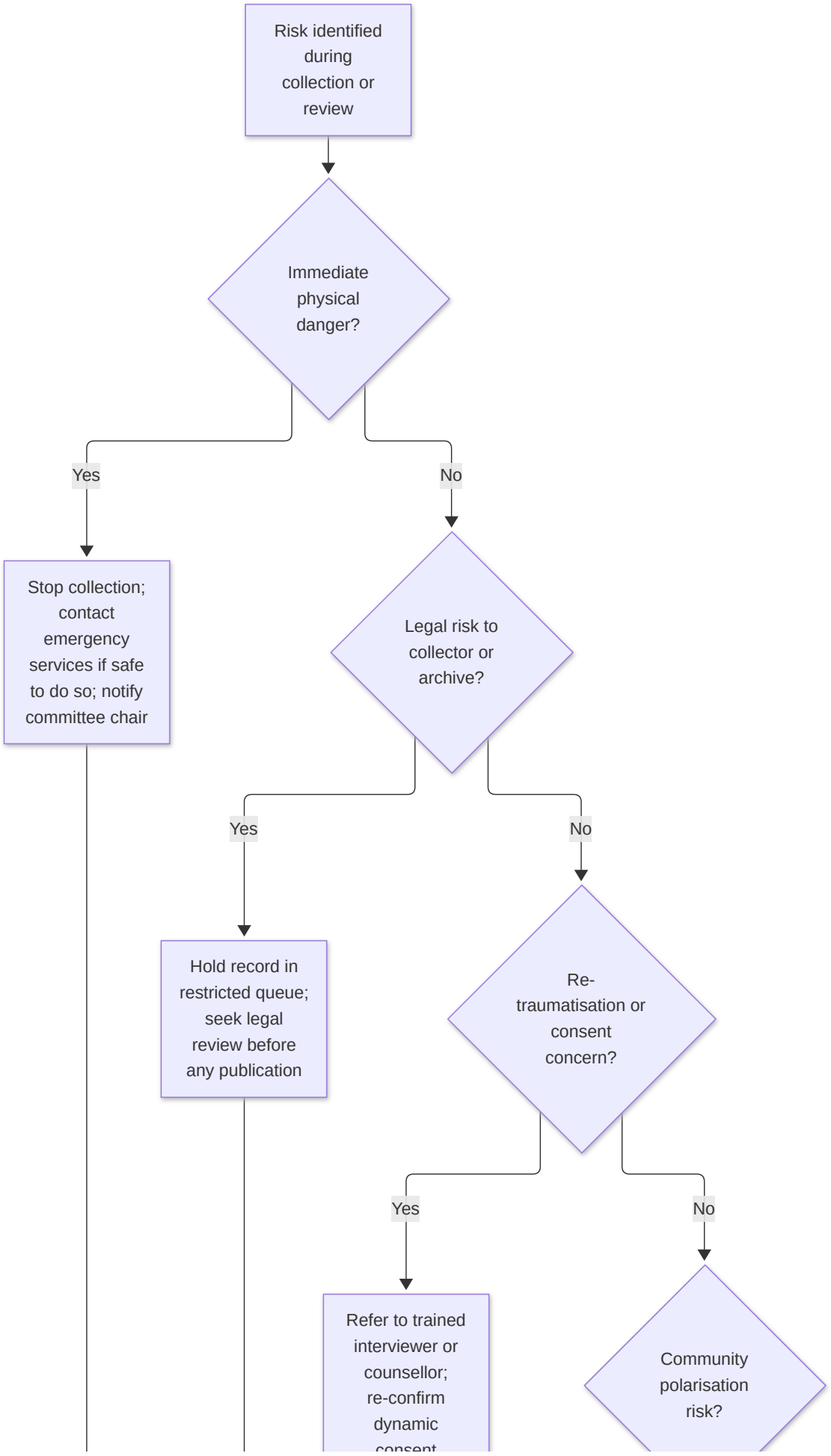
3.6.2 Pre-Collection Safety Checklist

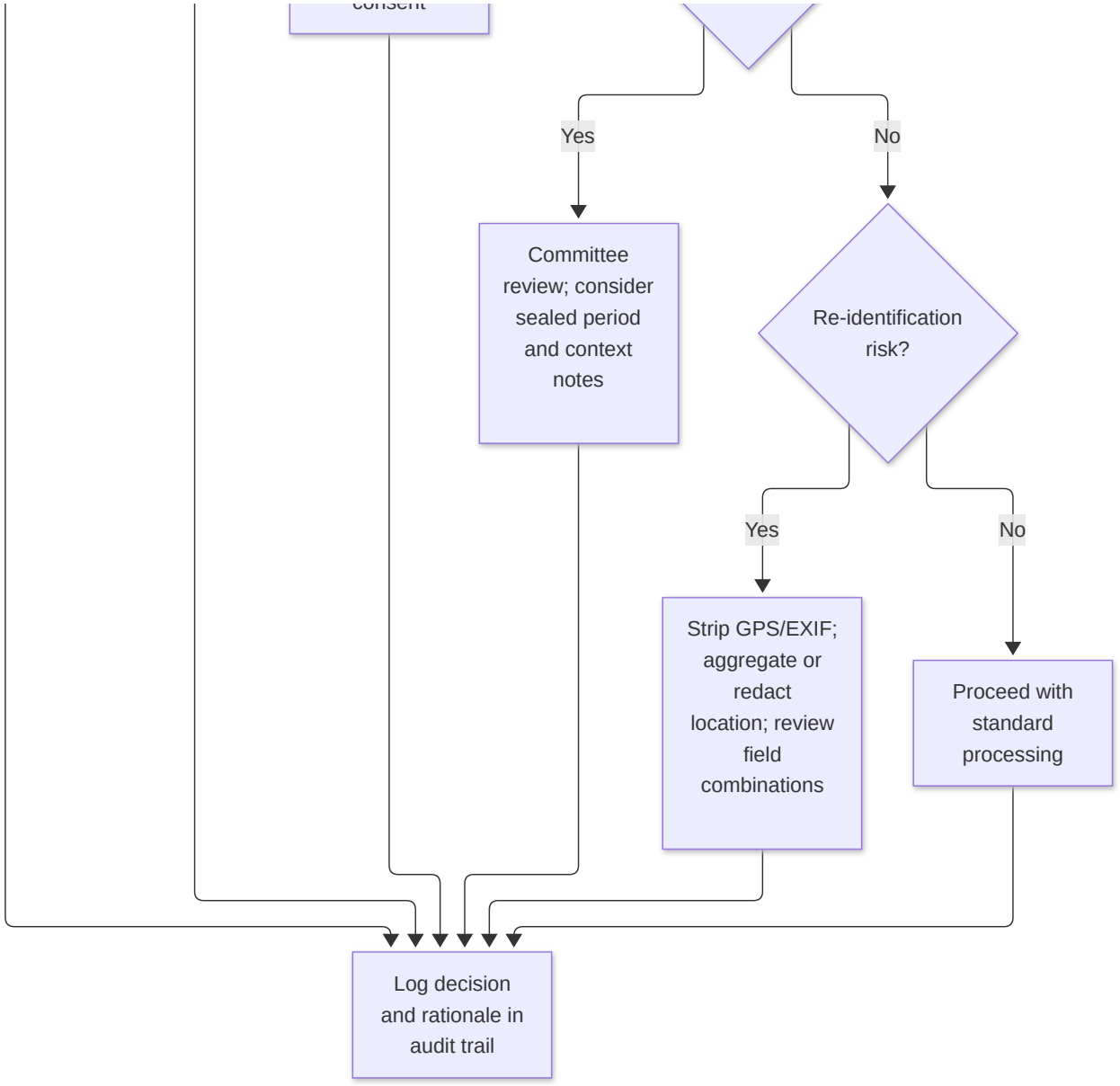
```
[ ] Identify the specific risk(s) to the collector and to witnesses for this event.
[ ] Confirm a check-in/check-out contact and schedule.
[ ] Confirm the device is encrypted and does not carry identity-linked material
    if the risk level is high.
[ ] Confirm the witness understands: purpose, access level, right to withdraw,
    and any legal limits on confidentiality.
[ ] Record consent (audio or written) with language-tagged template.
[ ] Agree a safe word or signal to end the session immediately.
[ ] Note any sensitivity flags that will apply (Section 3.5.1).
```

3.6.3 Post-Collection Safety Checklist

```
[ ] Transfer files to the archive system within 48 hours; remove from collection device.
[ ] Verify hash values after transfer; log transfer in the audit trail.
[ ] Assess whether any content requires escalation to the committee (e.g., immediate
    danger, legal risk, identifiable minor).
[ ] Confirm with the witness which access level applies and record their choice.
[ ] Apply sensitivity flags and strip GPS/EXIF from files intended for public release.
[ ] Schedule a follow-up contact (wellbeing check, consent confirmation).
```

3.6.4 Safety Escalation Flow


Risk identified during collection or review
Immediate physical danger?
Yes
No
Stop collection; contact emergency services if safe to do so; notify committee chair
Legal risk to collector or archive?
Yes
No
Hold record in restricted queue; seek legal review before any publication
Re-traumatisation or consent concern?
Yes
No
Refer to trained interviewer or counsellor; re-confirm dynamic
Community polarisation risk?



Chapter 4: Governance and Ethics

4.1 Community Archive Committee

Drawing on the implementation of "archival administrative responsibility" (档案行政管理主体责任, dàng àn xíng zhèng guǎn lǐ zhǔ tǐ zé rèn) in China's archival system, communities should establish a diverse and representative archive committee, including at minimum: a representative of the community self-governance organisation, a representative of collection volunteers, a representative of the custodian institution, and an external advisor with an archival/information management background. The committee's core responsibility is not "censoring content" but rather:

- Developing and revising *collection scope and access rules* (subject to public community discussion).
- Handling *access requests and objections* (e.g., recorded individuals requesting restricted access).
- Overseeing the implementation of *technical maintenance and backup*.

4.1.1 Committee Member Schema

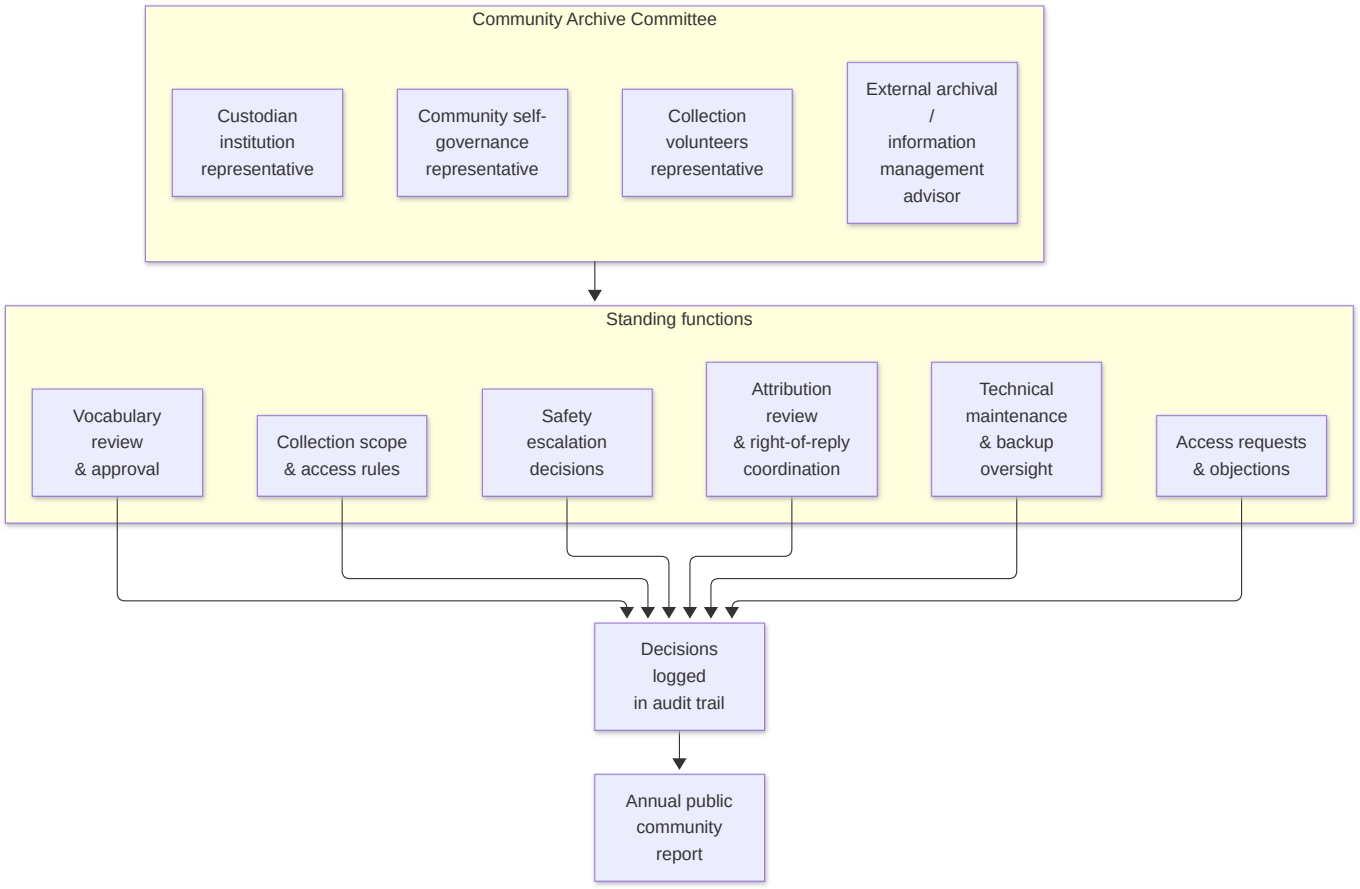
Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`memberId`	identifier	Yes	`cm-NNN`	Unique member identifier	`cm-004`
`displayName`	string	Yes	—	Name or pseudonym used in committee records	`A. Rivera`
`role`	enum	Yes	`chair`, `secretary`, `member`, `external-advisor`	Committee role	`external-advisor`
`constituency`	enum	Yes	`community`, `volunteers`, `custodian`, `external`	Whom the member represents	`external`
`appointedDate`	date	Yes	ISO 8601	Date appointed	`2024-03-01`
`termEndDate`	date	No	ISO 8601	End of current term	`2026-03-01`
`conflictOfInterest`	string	No	—	Declared conflicts of interest	`"Family member of named actor"`
`recusalScope`	string	No	—	Scope of recusal if any	`"All events involving Example Water Ltd."`
`contactRef`	string	No	—	Reference to contact record (not stored in public view)	`contact-004`
`status`	enum	Yes	`active`, `resigned`, `removed`	Membership status	`active`

4.1.2 Decision Rules

- Quorum: at least three members, including the external advisor or the custodian representative.
- Two-person rule for any decision affecting sealed records or safety escalations.

- Recusal: any member with a personal connection to the event or parties must recuse.
- Appeal: any affected party may request reconsideration within 30 days; the committee must respond in writing with reasons.

4.1.3 Committee Structure



4.2 Informed Consent and "Dynamic Consent"

Oral histories and audio-visual recordings must obtain *explicit, revocable informed consent*. A "dynamic consent" model is recommended: contributors may request at any time that their contributed content be removed from public access (though the archive system retains a metadata record of "removed"—i.e., recording the fact that "a certain item once existed but has been restricted upon request"—to maintain archival integrity).

4.2.1 Consent Record Schema

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`consentId`	identifier	Yes	`consent-YYYY-NNNN`	Unique consent record identifier	`consent-2024-0042`
`contributorId`	string	Yes	—	Pseudonymous contributor identifier	`contributor-0042`
`recordIds`	identifier[]	Yes	`rec-NNNN`	Records covered by this consent	`["rec-0045","rec-0046"]`
`consentType`	enum	Yes	`informed`, `dynamic`, `withdrawn`	Type of consent	`dynamic`
`consentDate`	date	Yes	ISO 8601	Date consent given	`2024-07-12`
`consentMethod`	enum	Yes	`written`, `audio`, `video`, `digital-form`	Method of recording consent	`audio`
`consentRef`	string	Yes	—	Reference to the consent artefact	`consent-audio-0042.mp3`
`language`	string	Yes	ISO 639-1/639-3	Language of consent	`en`
`scope@[lang]`	string	Yes	ISO 639 tag on field name	What the contributor agreed to	`"Public display of interview transcript"`
`accessLevelAgreed`	enum	Yes	`public`, `restricted`, `sealed`	Access level agreed at consent	`public`
`rightToWithdrawExplained`	boolean	Yes	—	Whether right to withdraw was explained	`true`
`withdrawalDate`	date	No	ISO 8601	Date consent withdrawn (if applicable)	—

Field	Type	Required	Controlled Vocabulary / Format	Description	Example
`withdrawalScope`	string	No	—	Scope of withdrawal	—
`withdrawalEffect`	enum	No	`restrict`, `seal`, `remove-public`	Effect of withdrawal	—
`reviewDate`	date	No	ISO 8601	Date of next consent review	`2025-07-12`
`auditTrailRef`	string	Yes	—	Reference to audit log entry	`audit-2024-07-12-0042`

4.3 Preventing Archives from Becoming Weapons

Transgression and disaster archives are inherently political. Mechanisms must be established to prevent archives from being used for:

- Retaliation against whistleblowers: Sealing periods and access approval are the first line of defence.
- Selective citation by external actors: When publishing publicly, simultaneously provide *contextual notes* (historical background of events, data limitations, known points of controversy) and enforce the attribution display rule (Section 3.4.5) to avoid misreading of isolated entries or allegations as findings.
- Internal community polarisation: The committee should include individuals with conflict mediation experience in access policy development.

Chapter 5: Launch Roadmap

Months 1–2: Form a core group (3–5 people); identify a custody partner; complete basic installation and testing of Omeka/OpenRefine; draft collection scope and access rules, and the attribution vocabulary (actor types and roles), using the frameworks in Chapter 3.

Months 3–4: Conduct first volunteer training (half-day: collection ethics, basic photography, metadata entry, safety precautions from Section 3.6); select a *low-sensitivity historical event* as a pilot (e.g., a natural disaster from years past) and complete the full pipeline from collection to publication, including at least one attribution entry with evidence links and one consent record.

Months 5–6: Pilot review and rule revision; deploy AtoM (or confirm a consortium custody arrangement); establish routine backup procedures; publish a public community report explaining the archive's existence, scope, and access methods.

Months 7–12: Expand collection scope to more sensitive events; establish formal custody agreements with local libraries/archives; apply for small grants for hardware (servers/storage) and training.

Conclusion

The core contradiction of community archives of transgressions and disasters is this: *the things most in need of recording are often the hardest to record.* What is transferable from China's archival governance experience is not its political logic but a set of *technical disciplines* tested through large-scale practice: tiered responsibility, quality testing, redundant backup, and institutionalised maintenance. The open-source technology stack offers the possibility of escaping proprietary dependence and maintaining community autonomy. The combination of the two points toward a *small but resilient* community memory infrastructure —one that does not pursue grand narratives but ensures one thing: when the community next experiences rupture or disaster, at least some verified records will survive, available for later community members to judge for themselves.

Reference Standards and Literature for further reading

- DA/T 70-2018 "General Requirements for Testing Document-Based Electronic Records" (四性检测, sì xìng jiǎn cè indicators and schemes)
- DA/T 99-2024 "Specification for the Implementation of Backup of Archival Digital Resources" (backup cycles and carrier requirements)
- GB/T 18894-2016 "Specification for Electronic Document Archiving and Electronic Records Management" (metadata and system security requirements)
- "The 14th Five-Year Plan for the Development of National Archival Undertakings" (grassroots archival work and informatisation direction)
- Mennonite Archival Information Database AtoM case study (community consortium collaboration model)
- W3C SKOS Reference (Simple Knowledge Organization System) for controlled vocabulary structure
- ISO 8601 (Date and time representation)
- ISO 639-1 / ISO 639-3 (Language codes)
- ISO 4217 (Currency codes)
- ISAD(G) (General International Standard Archival Description)
- GeoJSON (RFC 7946) for geometry representation
- LEI (Legal Entity Identifier) standards, ISO 17442