



Primmortal: Primary Sources, Made Immortal

Concept White Paper, Draft v0.2.1, September 26, 2026

Naphtali Moore, Founder

primmortal.org

Draft for discussion with prospective partners. Not for citation. Items marked
[FOUNDER TO CONFIRM] or [TBD] are open.

Executive summary

Researchers, students, journalists, and courts depend on citations that point to primary sources. Those citations fail more often than most people assume. A 2014 study of legal scholarship found that more than 70% of URLs in the journals examined, and 50% of URLs in U.S. Supreme Court opinions, no longer produced the information originally cited [1]. A 2024 Pew Research Center analysis found that 38% of webpages that existed in 2013 were no longer accessible a decade later [3]. Physical originals are lost too: the 1973 fire at the National Personnel Records Center destroyed an estimated 16 to 18 million military personnel files, and no duplicate copies had been made [4]. Meanwhile, convincing synthetic media makes it harder to tell whether a digital image of a document is what it claims to be [6].

Primmortal proposes a narrow, verifiable layer for primary sources. Consortium stewards register high-resolution captures of physical originals and packages of born-digital sources. For each item, a cryptographic hash and provenance metadata are recorded on a permissioned ledger, and the object is held in content-addressed storage replicated across stewards. A registered record is never edited. Corrections and new captures become new records that link to the old one. Each record receives a **citation token**, a short, verifiable citation identifier that lets anyone confirm that the file they are looking at is bit-for-bit the file a named steward registered on a given date.

Primmortal fixes the record and labels uncertainty honestly. It does not interpret sources, rank them, or filter them for moral comfort. Its ethics framework places technical design below ethical obligations: capture must be non-destructive, source communities are respected, and methods, selection, and governance are public.

The draft is deliberately frank about limits. A permissioned ledger moves trust to whoever governs admission; it does not remove it. "Immutable" describes a policy and a replication practice, not a law of physics. Tokens invite speculation. And the only working demonstration so far is a local v0 prototype: it registers files, issues citation tokens, verifies files and the ledger, and records corrections as new linked records on a hash-chained, signed ledger, but it runs on synthetic sample data with a single steward on one computer, and it is not hosted and not a production archive. The roadmap therefore begins small: a Phase 0 prototype, then a Phase 1 pilot with one partner library and one rights-clear collection, before any multi-steward consortium or public service.

Section 10 looks ahead to a graph of how ideas spread, built on registered citations. It explains why Primmortal will not offer identification of individuals as a feature, and it sets guardrails that keep any influence analysis separate from the neutral archive.

1. The problem

1.1 Citations that stop pointing at their sources

Scholarship assumes that a reader can follow a citation back to the evidence. On the web that assumption has weakened. Zittrain, Albert, and Lessig documented "reference rot" (links that fail or no longer show the cited content) in more than 70% of URLs in the legal journals they studied and in 50% of URLs in Supreme Court opinions [1]. Klein and colleagues, analyzing millions of scholarly articles in science, technology, and medicine, found that about one in five articles published between 2009 and

2012 suffered reference rot [2]. Pew's 2024 study found that 38% of webpages from 2013 had disappeared by 2023 [3]. Persistent-link services such as Perma.cc, proposed in [1], address part of the problem for web pages. No comparable shared service exists for primary sources in general.

1.2 Loss of originals

Physical records remain vulnerable to single events. On 12 July 1973, a fire at the National Personnel Records Center in St. Louis destroyed approximately 16 to 18 million Official Military Personnel Files. The National Archives notes that "no duplicate copies of these records were ever maintained, nor were microfilm copies produced" [4]. Digitization does not replace originals, but a well-documented master capture, held in several places, can preserve evidence of what an original contained.

1.3 Silent alteration

Records can be altered without an obvious trace. Historians have documented the systematic retouching of photographs in the Soviet Union to remove people who had fallen from favor [5]. Digital files are easier still to change, and unless a fixity value (a hash) was recorded and published at the time of capture, an altered copy can be indistinguishable from the original. Repositories routinely compute fixity internally. What is rarely available is a public, independently checkable attestation that ties a particular file to a particular steward and date.

1.4 Synthetic media

The U.S. National Institute of Standards and Technology describes a range of technical approaches to "digital content transparency," including provenance metadata and watermarking. It notes that these methods help but do not by themselves guarantee authenticity [6]. The Coalition for Content Provenance and Authenticity (C2PA) publishes an open standard, "Content Credentials," for recording the origin and edits of digital media [7]. Primmortal is complementary: C2PA travels with a media file from the moment of creation, while Primmortal records an archival steward's registration of a capture of a historical source, often long after that source was made.

1.5 Contested records and institutional mistrust

Public trust in institutions that mediate information is low. Gallup reported in October 2025 that 28% of U.S. adults trust the mass media "a great deal" or "a fair amount," the lowest figure in its trend [8]. In a low-trust environment, a record whose integrity depends on trusting a single institution is easy to dismiss. Primmortal's response is to spread custody across several independent stewards and to let anyone check a file against the public record with open-source tools, without having to trust Primmortal itself.

1.6 What is already solved

Much of this is already well handled by existing practice, and Primmortal should not pretend otherwise. LOCKSS networks keep many copies and repair damage by comparing them [11][12]. CLOCKSS provides a governed dark archive for scholarly publishing [13]. HathiTrust preserves digitized library collections at scale [14]. The National Digital Newspaper Program has made historic newspapers freely

available through Chronicling America [15]. The gap Primmortal targets is narrower: a **cross-institutional, publicly verifiable, citable registration** of primary-source captures, designed so that a student can check a citation in seconds.

2. What Primmortal is and is not

Primmortal is	Primmortal is not
A registry of hashes and provenance metadata for captures of primary sources, signed by named stewards	A wiki, a discussion platform, or an interpretive resource
A replicated, content-addressed store of the registered objects, held by consortium stewards	A replacement for the physical originals or for the stewards' own preservation programs
A citation-token system that lets anyone verify that a file matches a registered record	A cryptocurrency or an investment product
Governed by a nonprofit consortium with public policies	Controlled by any single company, university, or government
Honest about uncertainty (dates, attribution, capture defects)	A judge of which sources are true, good, or acceptable
Append-only: records are never edited; corrections are new, linked records	A system that silently deletes or rewrites anything

3. The archive model

3.1 The record

The unit of registration is a **record**: one capture package, its manifest of files and hashes, its provenance metadata, and the digital signature of the registering steward. Once registered, a record's content and metadata are never altered and never versioned in place. If a better capture is made, an error is found in the metadata, or a file format must be migrated, a **new record** is registered with an explicit relationship to the earlier one (for example, "recapture of," "corrects metadata of," "migrated from"). The earlier record stays visible and verifiable.

3.2 Scan levels

Primmortal uses two separate numbered scales. **Scan levels**, described here, record how a physical original was imaged. **Citation levels** (Section 4.7) describe how a source is cited. The two answer different questions and should not be confused.

Level 1: Basic optical scan. An ordinary optical image capture of the original, such as a photograph or a flatbed-style scan. Published cultural-heritage imaging guidance, such as the FADGI Technical Guidelines for Digitizing Cultural Heritage Materials (3rd edition, 2023) [20], is a suitable reference for capture practice.

Level 2: Non-destructive multispectral scanning. Multispectral scanning of every non-destructive kind: multispectral and hyperspectral imaging, ultraviolet and infrared imaging, raking light, and similar techniques. Such captures can show features of the original that a basic optical scan misses. Nothing about a Level 2 capture may damage the original.

Level 3: Holographic scan (theoretical). A holographic capture of the original. This level is theoretical: no such scanning method exists today.

At every level, capture must be non-destructive; no item may be disbound or otherwise altered to be registered. A capture package includes the unprocessed or minimally processed master files, color and scale targets where applicable, capture-device metadata, the operator and date, and a condition note, and the registering steward states the scan level in the record's metadata. A later capture of the same original, at the same or a different level, is registered as a new record linked to the earlier one ("recapture of," Section 3.1). In the v0 prototype, the scan level is self-declared by the steward and not checked, and the prototype's capture-level labels predate these definitions and will be updated.

Born-digital sources. Scan levels describe the imaging of physical originals. Where born-digital sources (for example, a newspaper's digital edition, an email archive, or a government dataset) fit in the scan levels is an open question [FOUNDER TO CONFIRM]. Born-digital sources are still registered as received: the record holds the original bitstreams exactly as received, packaged with a manifest of files and checksums. The BagIt packaging format (RFC 8493) is a suitable starting point [23], and preservation events should be described using the PREMIS data dictionary [29]. Where a source was issued in versions, each version is its own record, linked in sequence.

Primes and derivatives. The registered original capture, whether a scan of a physical original or born-digital files as received, is the **prime**, and primes are canonical. OCR text, transcriptions, web-friendly images, and translations are **derivatives**. They are not primes and not a scan level, and **they are never canonical**. Each derivative is linked to the prime it came from: it carries its own hash and a pointer to its prime, and can be regenerated or superseded without affecting the prime. Citations should always resolve to a prime, and to a derivative only as a convenience.

3.3 Labeling uncertainty

Every record carries structured fields for what is known and what is not: date (exact, approximate, or unknown, with the basis stated), attribution and its confidence, custodial history as known, known gaps or damage in the original, and known defects in the capture. Primmortal records these statements as metadata supplied by the steward. It does not adjudicate them, and it does not add interpretation.

3.4 Selection

What is registered, by whom, and why must be public. Each steward publishes a selection statement for each collection it registers. The consortium publishes general criteria: primary sources only; rights and permissions documented; community protocols respected (Section 6). The first collecting scope is [TBD] and should be set with the Phase 1 partner.

3.5 Restricted material and legal takedowns: no silent deletion

An archive that promises never to alter records must still obey the law and respect communities. Primmortal separates the **registration** (the ledger entry, which is never removed) from **access** to the object. If an object must be withdrawn (a court order, a valid rights claim, a privacy obligation to a living person, or a community protocol), stewards stop serving it publicly and append a **status event** to the ledger stating that access is withheld, with a reason category and date. The hash remains, so the fact of registration and the integrity of any lawfully held copy can still be verified. Nothing disappears silently. Where the law requires actual destruction of every copy, the status event records that, too. Legal counsel must review this policy before Phase 2 [TBD].

4. Technical architecture

4.1 Overview

The pipeline has six parts:

1. **Capture and packaging.** A steward produces a capture package (a scan of a physical original at a stated scan level, Section 3.2) or a born-digital package, and an open-source packaging tool writes the manifest and metadata.
2. **Hashing.** The tool computes cryptographic hashes of every file and of the manifest.
3. **Steward signature.** The steward signs the registration with its organizational key.
4. **Ledger registration.** The signed registration is submitted to the permissioned ledger.
5. **Content-addressed storage.** The package is added to content-addressed storage and replicated (pinned) by several stewards.
6. **Citation and verification.** A citation token is issued. Anyone can resolve it and verify the file independently.

4.2 Hashes

Each file and manifest is hashed with SHA-256, plus a second, independent algorithm (for example SHA-512 or SHA3-256) so that a future weakness in one algorithm does not undermine the record. The choice of algorithms is recorded in each registration. When the consortium adopts a new algorithm, it re-hashes the existing objects and records the new values as additional status events. The original values are never overwritten.

4.3 Content-addressed storage

In content-addressed storage, an object's address is derived from its contents, so retrieving an object by its address also checks its integrity. The InterPlanetary File System (IPFS) is the best-known design [16]. Content addressing guarantees that what you retrieve is what was addressed. **It does not guarantee that anyone still holds a copy.** Persistence depends entirely on stewards choosing to keep (pin) the data. Primmortal therefore sets a replication policy as a proposal: at least three independent

stewards in at least two regions hold every canonical object, with scheduled audits in the spirit of LOCKSS [11]. Whether the alpha uses IPFS itself or another content-addressed system is [FOUNDER TO CONFIRM].

4.4 The permissioned ledger

The ledger records, for each record: a record identifier; the file and manifest hashes; the content address; a hash of the metadata; the steward's identity and signature; the registration time; and subsequent status events (withheld, key revoked, re-hashed, succeeded-by). **Objects are never stored on the ledger.**

The alpha targets **Hyperledger Fabric**, an open-source permissioned blockchain platform [17] hosted by the Linux Foundation's LF Decentralized Trust. Fabric 3.0 (September 2024) added a Byzantine-fault-tolerant ordering service [18]. A permissioned design fits Primmortal because stewards are known institutions, consensus does not require an energy-intensive mining process or a native currency, and every write is attributable to an accountable organization. Smart contracts ("chaincode") would enforce append-only rules: a registration can be followed by status events but never edited.

A full blockchain may not be necessary. An append-only transparency log of the kind used for Certificate Transparency (RFC 6962) [30], with signed checkpoints held by several stewards, might provide most of the same guarantees with far less operational complexity. Phase 0 should prototype the simpler option first, and Phase 2 should compare the two before committing to Fabric.

4.5 Signed steward registrations

Each steward holds organizational signing keys, issued through Fabric's membership service in the alpha. Keys are rotated on a published schedule. If a key is compromised, a revocation event is appended. Records signed with that key stay in place but are flagged for review. They are never deleted.

4.6 Citation tokens

In this paper, a **citation token is a verifiable citation identifier, not a cryptocurrency**. A token combines the record identifier with a short fingerprint of the canonical hash, in a form that can be printed in a footnote. An illustrative format (not final): `pm:1:<record-id>:<first 16 hex characters of SHA-256>`. The exact syntax is [TBD]. A resolver turns the token into the record's metadata and retrieval addresses. A reader who has a copy of the file can recompute the hash and confirm that it matches.

4.7 Citation levels

Citation levels describe how a source is cited. They are separate from the scan levels in Section 3.2. There are only two.

Level 1 citation: the universal citation token. The citation token described in Section 4.6. Anyone can use it, with no account, to point to a registered record and to verify that a file matches it. The v0 prototype issues Level 1 tokens today, in a provisional format that differs from the illustrative format in Section 4.6.

Level 2 citation: a registered citing work. The citing work itself (for example, an article, a paper, or a news story) is registered in the ledger and linked to the source record it cites, so the connection between the two is part of the record rather than only a footnote. Professionals such as academics and journalists make Level 2 citations. Level 2 citations are designed but not yet implemented in the

prototype, which links records to one another only through correction notices. How citing works are submitted for registration, and by whom, given that only admitted stewards write to the ledger (Section 5.3), is [TBD].

There is no Level 3 citation.

4.8 Independent verification

Verification must not depend on trusting Primmortal's own servers. The verifier will be open source and runnable offline. The consortium will publish regular signed checkpoints of the ledger state (for example, a Merkle root) to several independent locations, such as each steward's website. That way, anyone can detect if the ledger were rewritten. Anchoring checkpoints in an external public system is an option to evaluate, not a requirement.

4.9 What the ledger does not prove

A hash proves only that a particular bitstream was registered by a particular steward at a particular time. It does not prove that the capture faithfully represents the original, that the original is authentic, or that its metadata are correct. Lemieux makes this point about blockchain records generally: the technology can support the integrity of records after creation, but it cannot establish that the information was reliable in the first place [19]. Primmortal's trustworthiness therefore rests on steward practice, documented capture standards, and public review as much as on cryptography.

5. Governance

5.1 A nonprofit consortium

Primmortal would be governed by a nonprofit consortium of stewards, envisioned as major university libraries and newspapers of record, with a structure modeled loosely on the National Digital Newspaper Program. In that program, the National Endowment for the Humanities and the Library of Congress partner with state institutions that select and digitize newspapers, and the Library of Congress maintains Chronicling America [15]. Primmortal would like the Library of Congress, or a comparable national body, to serve as a standards or hosting partner. **No such relationship exists today.** The legal form and state of incorporation of the consortium are [TBD].

5.2 Roles

- **Stewards:** institutions that capture, register, sign, and replicate records, and run ledger nodes.
- **Members:** individuals and organizations who support the archive through a modest membership fee and receive services and a voice in public review, but no control over the record.
- **Verifiers:** anyone, with no account required.
- **Service providers:** companies that build citation, verification, and teaching tools on top of the open archive under a published license (Section 7).
- **Advisory councils:** ethics (including Indigenous and community representation), technical and security, and legal.

5.3 Admission: where trust actually lives

A permissioned ledger is only as trustworthy as the process that decides who may write to it. Primmortal proposes the following:

- Publish admission criteria.
- Hold a public comment period before admitting a steward.
- Record admission and removal decisions on the ledger itself.
- Require that no single organization operate more than one third of ordering or consensus nodes.
- Commission periodic external audits of steward practice.

These measures reduce the problem. They do not eliminate it. The consortium should say so plainly.

5.4 Public review

Selection statements, capture standards, the ethics framework, governance minutes, and this white paper are public documents, open to comment. Substantive changes to policy take effect only after a published notice period.

5.5 Steward succession and format migration

Institutions and programs end. The MetaArchive Cooperative, the first and longest-running private LOCKSS network, sunset on 31 March 2025. Roughly half its members took their data back, and the rest moved to a transitional network operated by the LOCKSS Program [24]. Primmortal should plan for this from the start:

- Each steward signs a succession agreement naming how its replicas and keys pass to another steward.
- Replication minimums are enforced so that no single departure puts content at risk.
- CLOCKSS-style trigger rules, under which content is released by supermajority board vote after defined events [13], govern what happens if the consortium itself winds down.

Format migration follows the record rule: migrated files are registered as new records linked to the originals, and the original bitstreams are kept.

5.6 Durability language

Primmortal will avoid promising permanence it cannot guarantee. Public materials should say that registered records are **never edited**, that they are **replicated across independent stewards**, and that the consortium **commits to succession and migration plans**. They should not say that records are "permanent" or "indestructible." The name expresses an aspiration; the technical claims must stay precise.

6. Ethics and the Archival Oath

Primmortal's ethics framework places technical design below ethical obligations. Its commitments include the following:

- **Non-destructive capture.** No original may be damaged or altered in order to be registered.
- **Respect for source cultures and living communities.** Some materials, especially Indigenous cultural materials, are governed by protocols that restrict who may see them. An immutable public archive can conflict directly with those protocols. Primmortal will follow the lead of community-directed practice such as the Mukurtu platform and the Plateau Peoples' Web Portal at Washington State University [28], and the Traditional Knowledge Labels developed by Local Contexts [31].

Communities decide whether their materials are registered at all. They may choose hash-only registration, with the object held by the community, or restricted access.

- **No filtering for moral comfort.** Offensive or painful records are part of the historical record and are not excluded for that reason. This is distinct from obligations to living people, such as privacy, which are handled through the withheld-access process in Section 3.5.
- **Transparency.** Methods, selection criteria, and governance are published and open to public review.

Participants affirm the **Archival Oath**:

We preserve what is primary. We guard it with integrity. We entrust it to the future.

7. Sustainability and funding

The core is a public good. Verification, public access to unrestricted records, and the reference software should always be free and open source. Funding sources in likely order of arrival:

1. **In-kind steward contributions.** Staff time, storage, and ledger nodes, as in existing preservation cooperatives.
2. **Grants through partners.** Most humanities and preservation grants require an eligible institutional applicant, so Primmortal will need a university partner or fiscal sponsor. Funding is also uncertain. For example, NEH's Digital Humanities Advancement Grants program is listed as cancelled for fiscal year 2026 [26].
3. **Modest membership fees** from individuals and organizations. The proposed amount is [FOUNDER TO CONFIRM].
4. **Value-added services by private firms.** Examples include citation-manager plug-ins, verification APIs for publishers, and classroom tools. Firms would pay license or service fees but could never gain governance control or exclusive access to the core archive.

Experimental later phase: tradable tokens and background "useful work." Earlier versions of the concept imagined citation tokens being exchanged, and members' computers doing useful background work: hash verification, re-pinning and storage proofs, and generating OCR and other derivatives. Such work would have to be checked redundantly before stewards accept it. This paper treats all of this as **an experiment for a later phase, not a core requirement**, for three reasons. First, anything tradable may raise securities, money-transmission, and tax questions that need formal legal review. Second, speculation could distort incentives and damage the archive's credibility. Third, the core archive works without it. Nothing in Phases 0 through 2 depends on tokens having monetary value.

8. Related work and precedents

Precedent	What it does	What Primmortal borrows or differs on
ARCHANGEL (UK National Archives with University of Surrey and partners) [9][10]	Records hashes of archival documents and video on a blockchain shared by archives, to detect tampering	Closest precedent for blockchain fixity; Primmortal adds public citation tokens and replicated storage of the objects
LOCKSS [11][12]	Peer-to-peer preservation: many copies, compared and repaired	Replication and audit model for storage
CLOCKSS [13]	Community-governed dark archive with trigger-based release	Succession and trigger governance
MetaArchive Cooperative [24]	Private LOCKSS network, 2004 to 2025	Cautionary lesson on sustainability and exit planning
HathiTrust [14]	Shared repository of digitized library collections, founded 2008	Consortium scale and governance
NDNP / Chronicling America [15][27]	NEH and Library of Congress program; state partners digitize newspapers	Governance and selection model; possible pilot content
IPFS [16]	Content-addressed, peer-to-peer file system	Addressing scheme for stored objects
Hyperledger Fabric [17][18]	Permissioned blockchain platform	Alpha ledger target
Certificate Transparency (RFC 6962) [30]	Public append-only logs with signed checkpoints	Simpler alternative ledger design
C2PA Content Credentials [7]	Provenance manifests attached to media files	Complementary; could be ingested as metadata
Perma.cc [1]	Library-backed preserved links for citations	Citation-stability precedent for web sources
OAIS (ISO 14721:2025) [22]; NDSA Levels [21]; FADGI [20]; BagIt [23]; PREMIS [29]	Reference model, preservation levels, imaging, packaging, and metadata standards	Standards Primmortal should adopt rather than reinvent

9. Risks and honest limitations

- **Trust in governance.** The permissioned ledger relocates trust to admission governance (Section 5.3). A captured or careless consortium could register bad records.
- **Registration is not authentication.** A hash proves integrity after registration, not faithfulness or authenticity (Section 4.9) [19].
- **Durability depends on people and money.** Without sustained replication, succession, and migration, content-addressed data can disappear. MetaArchive shows that even long-running cooperatives end [24]. Single large archives are vulnerable too: in October 2024, attacks took the Internet Archive's services offline for days [25].

- **Legal conflicts.** Court orders, copyright, and privacy law can conflict with "never delete." The withheld-access model (Section 3.5) must be reviewed by counsel.
- **Ethical conflicts.** Immutability can harm communities whose materials were taken without consent. Community control over registration is mandatory, not optional.
- **Key and algorithm risk.** Steward keys can be compromised, and hash algorithms can weaken. Mitigations are rotation, revocation events, and dual hashing.
- **Token speculation.** It could undermine credibility, which is why tokens are deferred and subject to legal review (Section 7).
- **Complexity.** Fabric is operationally demanding for small archives, and a simpler log may be enough (Section 4.4).
- **Adoption.** Archives already run fixity checks internally, so Primmortal must show a clear benefit (public verifiability and citation) at low cost to stewards.
- **Stage and capacity.** The only working demonstration is a local v0 prototype on synthetic sample data, with a single steward; it is not hosted and not a production archive. The project currently has a single founder. Partners should weigh this.

10. Dual Use, Misuse, and the Influence Graph

10.1 A future direction: the influence graph

Once citations point to registered originals, the links become data. Author-asserted citations, plus links that AI tools might later suggest in existing documents (an unattributed quotation, a close paraphrase, a reused passage), would form a graph of how ideas spread. Versions of this graph already exist:

- Northeastern University's Viral Texts project used text-reuse detection on *Chronicling America* newspapers to map nineteenth-century reprinting [32].
- Researchers have proposed network methods for detecting "citation cartels" [33].
- The Problematic Paper Screener flags papers containing machine-generated text or "tortured phrases" for human review [34].
- The Election Integrity Partnership, including UW's Center for an Informed Public, traced how false 2020 election claims spread across platforms and merged into larger narratives [35].

Primmortal could add what these efforts usually lack: trusted registration timestamps. If two registered documents share a passage, the ledger shows which was registered first, so some priority and backdating questions become answerable from evidence.

10.2 Limits

- **Registration time is not creation time.** A registration proves that a document existed by a certain date. It does not prove when the document was written, or that an earlier-registered text was the source of a later one.
- **Similarity is not copying.** Independent discovery, shared sources, formulaic language, and common quotations all produce matches without any influence between the texts.
- **Asserted and inferred links are different kinds of claim.** An author's citation is a fact about a document. An AI-inferred link is a hypothesis. Inferred links must be stored in a separate layer, labeled with the method, model version, date, and a confidence score, and open to challenge by

authors and stewards, with disputes recorded publicly. They must never be merged into the canonical record.

- **Adversaries adapt.** Automated paraphrasing is already used to evade reuse detection, as the tortured-phrases research shows [34]. Any detection method will be probed and gamed.
- **The graph covers only what is archived.** If Primmortal holds no earlier source, that is not evidence that none existed. Coverage will also skew toward institutions that can afford to register material.

10.3 Dual-use risk

The methods that trace ideas can also identify people. Stylometry, the statistical analysis of writing style, has a long record. In 1963, Mosteller and Wallace used word-frequency statistics to attribute the disputed Federalist Papers, published under the shared pseudonym "Publius," to James Madison [36]. In 2013, computational analyses done for The Sunday Times supported the identification of "Robert Galbraith" as J. K. Rowling [37]. Juola also describes an asylum case that turned on whether the applicant wrote anonymous articles critical of his government [37].

Applied to a large archive of timestamped texts, these methods could be turned against whistleblowers, dissidents, and pseudonymous writers who depend on anonymity for their safety. License-plate-reader networks show the pattern. The ACLU documented in 2013 that they kept location data on innocent drivers for long periods [38]. In May 2025, reporting showed a Texas sheriff's office searching more than 83,000 Flock cameras nationwide, including in states where abortion is legal, for a woman who had self-managed an abortion; the office said her family was concerned for her safety [39][40]. Once data are pooled and searchable, capabilities built for one purpose get used for others.

Primmortal does not offer identification of individuals as a feature. It will not build authorship attribution or de-anonymization into the archive or its reference services.

10.4 Guardrails

- **A neutral archive layer.** The canonical record states only facts: this exact file, registered at this time, by this steward. It contains no inferred links, authorship scores, or influence rankings.
- **A separate, clearly labeled analysis layer.** Influence analysis is built by researchers or service providers, presented as interpretation, with published methods.
- **Consortium policy on person-identifying analysis.** Licensed services and consortium projects may not offer identification of living pseudonymous or anonymous people. Historical attribution scholarship, of the Federalist kind, is distinguished from analysis of living persons. Where and how that line is drawn, and what ethics review applies, is [TBD].
- **A public, tamper-evident log of bulk access.** Bulk queries and downloads above a threshold ([TBD]) are recorded in an append-only public log, using the registry's own mechanism, with the requesting organization, date, scope, and stated purpose. Individual readers' lookups are not logged.
- **Library privacy values.** Article VII of the American Library Association's Library Bill of Rights states that "all people, regardless of origin, age, background, or views, possess a right to privacy and confidentiality in their library use" [41]. Primmortal will follow this principle: verification requires no account, and the resolver keeps no reader-level records beyond what operations require (retention period [TBD]).

11. Roadmap

Phase 0: Concept and prototype (now).

- Circulate this draft for critique.
- Write a two-page threat model.
- Build an open-source demo that hashes and packages files, signs a registration, appends it to a simple append-only log, issues a citation token, and verifies a file offline.
 - Status: a local v0 prototype now registers files, signs each registration, appends it to a hash-chained, append-only ledger file, issues a Level 1 citation token, records corrections as new linked records, and verifies files and the ledger offline against a public export. It runs on synthetic sample data with a single steward on one computer, and it is not hosted and not a production archive. It registers single files rather than packages with manifests, its code is not yet published, and its capture-level labels predate the scan levels in Section 3.2 and will be updated.
- Hold partner conversations.
- Identify a university partner or fiscal sponsor.
- Timeline: [TBD].

Phase 1: Pilot with one partner library and one collection.

- Register a small, rights-clear collection with one steward. Already-digitized public-domain newspaper pages from an NDNF state project are a natural candidate, subject entirely to the partner's agreement [FOUNDER TO CONFIRM].
- Size: [TBD].
- Test the scan levels, the metadata schema, the withheld-access process, and the verifier with real students.
- Success criteria: [TBD], for example, the time it takes a student to verify a citation, and steward staff hours per item.

Phase 2: Multi-steward consortium alpha on Fabric.

- Three to five stewards in at least two regions.
- Comparative evaluation of Fabric against a transparency log.
- Draft consortium charter, admission policy, and succession agreements.
- Independent security review, legal review of takedown and token questions, and an ethics advisory council with community representation.

Phase 3: Public citation service.

- Public resolver and verification API.
- Integration with citation managers.
- Licensing framework for private service providers.
- Membership program launch.

12. Call to partners

Primmortal is looking for:

- **Archivists and librarians** willing to critique the model.
- **One partner library** with a small, rights-clear collection for a Phase 1 pilot.
- **Security and distributed-systems researchers** to review the threat model.
- **Community and Indigenous-data advisors** to shape the ethics framework.
- **A university partner or fiscal sponsor** to make grant applications possible.

The first request is always small: a conversation and honest feedback. Contact: Naphtali Moore, [FOUNDER TO CONFIRM: contact email], <https://primmortal.org>.

References

1. Zittrain, J., Albert, K., & Lessig, L. (2014). Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations. *Harvard Law Review Forum*, 127, 176. <https://harvardlawreview.org/forum/vol-127/perma-scoping-and-addressing-the-problem-of-link-and-reference-rot-in-legal-citations/>
2. Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLOS ONE*, 9(12), e115253. <https://doi.org/10.1371/journal.pone.0115253>
3. Chapekis, A., Bestvater, S., Remy, E., & Rivero, G. (2024, May 17). *When Online Content Disappears*. Pew Research Center. <https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears/>
4. U.S. National Archives and Records Administration. The 1973 Fire, National Personnel Records Center. <https://www.archives.gov/personnel-records-center/fire-1973>
5. King, D. (1997). *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia*. New York: Metropolitan Books.
6. National Institute of Standards and Technology (2024). *Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency* (NIST AI 100-4). <https://doi.org/10.6028/NIST.AI.100-4>
7. Coalition for Content Provenance and Authenticity (C2PA). Content Credentials and the C2PA Specification. <https://c2pa.org/>
8. Gallup (2025, October 2). Trust in Media at New Low of 28% in U.S. <https://news.gallup.com/poll/695762/trust-media-new-low.aspx>
9. Collomosse, J., Bui, T., Brown, A., Sheridan, J., Green, A., Bell, M., Fawcett, J., Higgins, J., & Thereaux, O. (2018). ARCHANGEL: Trusted Archives of Digital Public Documents. *Proceedings of the ACM Symposium on Document Engineering (DocEng 2018)*. <https://arxiv.org/abs/1804.08342>
10. Bui, T., Cooper, D., Collomosse, J., Bell, M., Green, A., Sheridan, J., Higgins, J., Das, A., Keller, J., Thereaux, O., & Brown, A. (2019). ARCHANGEL: Tamper-Proofing Video Archives Using Temporal Content Hashes on the Blockchain. *CVPR Workshops 2019*. <https://arxiv.org/abs/1904.12059>
11. Maniatis, P., Roussopoulos, M., Giuli, T. J., Rosenthal, D. S. H., & Baker, M. (2005). The LOCKSS Peer-to-Peer Digital Preservation System. *ACM Transactions on Computer Systems*, 23(1). <https://doi.org/10.1145/1047915.1047917>
12. LOCKSS Program, Stanford Libraries. What is LOCKSS? <https://www.lockss.org/about/what-lockss>
13. CLOCKSS. Triggered Content; CLOCKSS Access Policy. <https://clockss.org/triggered-content/>; https://documents.clockss.org/index.php/CLOCKSS:_Access_Policy
14. HathiTrust. Our Mission & History. <https://www.hathitrust.org/about/mission-history/>

15. Library of Congress. National Digital Newspaper Program: Award Recipients; Chronicling America: About this Collection. <https://www.loc.gov/ndnp/awards/>; <https://www.loc.gov/collections/chronicling-america/about-this-collection/>
16. Benet, J. (2014). IPFS: Content Addressed, Versioned, P2P File System. arXiv:1407.3561. <https://arxiv.org/abs/1407.3561>
17. Androulaki, E., et al. (2018). Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. *Proceedings of EuroSys 2018*. <https://arxiv.org/abs/1801.10228>
18. LF Decentralized Trust (2024, September 16). Version 3.0 of Hyperledger Fabric, an LF Decentralized Trust Project, Now Available. <https://www.lfdecentralizedtrust.org/announcements/version-3.0-of-hyperledger-fabric-an-lf-decentralized-trust-project-now-available>
19. Lemieux, V. L. (2016). Trusting Records: Is Blockchain Technology the Answer? *Records Management Journal*, 26(2), 110–139. <https://doi.org/10.1108/RMJ-12-2015-0042>
20. Federal Agencies Digital Guidelines Initiative (2023). *Technical Guidelines for Digitizing Cultural Heritage Materials*, 3rd ed. <https://www.digitizationguidelines.gov/guidelines/digitize-technical.html>
21. National Digital Stewardship Alliance (2019). Levels of Digital Preservation, Version 2.0. <https://ndsa.org/publications/levels-of-digital-preservation/>
22. ISO 14721:2025 / CCSDS 650.0-M-3. Space data and information transfer systems: Open archival information system (OAIS) Reference model.
23. Kunze, J., Littman, J., Madden, E., Scancella, J., & Adams, C. (2018). *The BagIt File Packaging Format (V1.0)*. RFC 8493. <https://www.rfc-editor.org/info/rfc8493>
24. MetaArchive Cooperative (2025, March 28). MetaArchive Cooperative Sunset Announcement. Educopia Institute. <https://educopia.org/wp-content/uploads/2025/03/MetaArchive-Cooperative-Sunset-Announcement-MetaArchive.pdf>
25. Internet Archive (2024, October 21). Internet Archive Services Update: 2024-10-21. <https://blog.archive.org/2024/10/21/internet-archive-services-update-2024-10-21/>
26. National Endowment for the Humanities. Digital Humanities Advancement Grants (program page). <https://www.neh.gov/grants/odh/digital-humanities-advancement-grants>
27. LSU Libraries. Digitizing Louisiana Newspapers Project (DLNP) research guide. <https://guides.lib.lsu.edu/c.php?g=1431778>
28. Washington State University, Center for Digital Scholarship and Curation. Plateau Peoples' Web Portal: About. <https://plateauportal.libraries.wsu.edu/about>
29. Library of Congress. PREMIS Data Dictionary for Preservation Metadata. <https://www.loc.gov/standards/premis/>
30. Laurie, B., Langley, A., & Kasper, E. (2013). *Certificate Transparency*. RFC 6962. <https://www.rfc-editor.org/info/rfc6962>
31. Local Contexts. Traditional Knowledge (TK) Labels. <https://localcontexts.org/labels/traditional-knowledge-labels/>
32. Cordell, R., & Smith, D. A. (2024). *Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines*. <https://viraltxts.org>; see also Smith, D. A., Cordell, R., & Dillon, E. M. (2013). Infectious Texts: Modeling Text Reuse in Nineteenth-Century Newspapers. *Proceedings of the IEEE Workshop on Big Humanities*. <https://doi.org/10.1109/BigData.2013.6691675>
33. Fister, I., Jr., Fister, I., & Perc, M. (2016). Toward the Discovery of Citation Cartels in Citation Networks. *Frontiers in Physics*, 4, 49. <https://doi.org/10.3389/fphy.2016.00049>
34. Cabanac, G., Labbé, C., & Magazinov, A. (2021). Tortured Phrases: A Dubious Writing Style Emerging in Science. arXiv:2107.06751 <https://arxiv.org/abs/2107.06751>; and Cabanac, G., Labbé, C., & Magazinov, A. (2022). The 'Problematic Paper Screener' Automatically Selects Suspect Publications for Post-Publication (Re)assessment. arXiv:2210.04895 <https://arxiv.org/abs/2210.04895>
35. Center for an Informed Public, University of Washington (2021, March 2). Election Integrity Partnership Releases Final Report on Mis- and Disinformation from the 2020 U.S. Election ("The Long Fuse").

<https://www.cip.uw.edu/2021/03/02/election-integrity-partnership-final-report-long-fuse/>

36. Mosteller, F., & Wallace, D. L. (1963). Inference in an Authorship Problem. *Journal of the American Statistical Association*, 58(302), 275–309. <https://doi.org/10.1080/01621459.1963.10500849>
37. Juola, P. (2015). The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions. *Digital Scholarship in the Humanities*, 30(suppl. 1), i100–i113. <https://doi.org/10.1093/llc/fqv040>
38. American Civil Liberties Union (2013, July). *You Are Being Tracked: How License Plate Readers Are Being Used to Record Americans' Movements*. <https://www.aclu.org/documents/you-are-being-tracked-how-license-plate-readers-are-being-used-record-americans-movements>
39. Cox, J., & Koebler, J. (2025, May 29). A Texas Cop Searched License Plate Cameras Nationwide for a Woman Who Got an Abortion. *404 Media*. <https://www.404media.co/a-texas-cop-searched-license-plate-cameras-nationwide-for-a-woman-who-got-an-abortion/>
40. Electronic Frontier Foundation (2025, May). She Got an Abortion. So A Texas Cop Used 83,000 Cameras to Track Her Down. <https://www.eff.org/deeplinks/2025/05/she-got-abortion-so-texas-cop-used-83000-cameras-track-her-down>
41. American Library Association. Library Bill of Rights, Article VII. <https://www.ala.org/advocacy/intfreedom/librarybill>

Appendix: open items for the founder

- [FOUNDER TO CONFIRM] Where born-digital sources fit in the scan levels (Section 3.2).
- [FOUNDER TO CONFIRM] Whether the alpha uses IPFS itself or another content-addressed system (Section 4.3).
- [FOUNDER TO CONFIRM] Membership fee amount (Section 7).
- [FOUNDER TO CONFIRM] Phase 1 candidate collection, subject to partner agreement (Section 11).
- [FOUNDER TO CONFIRM] Public contact email (Section 12).
- [TBD] First collecting scope (Section 3.4); legal review of takedown policy (Section 3.5); citation-token syntax (Section 4.6); how citing works are registered for Level 2 citations, and by whom (Section 4.7); consortium legal form and state of incorporation (Section 5.1); Phase 0 timeline, Phase 1 size and success criteria (Section 11); boundary and ethics review for person-identifying analysis, bulk-access logging threshold, and resolver log retention (Section 10.4).

Revision history

- **v0.2.1 (September 26, 2026), draft.** Section 3.2 rewritten: the proposed three-level capture hierarchy (master optical capture, born-digital package, derivatives) is replaced by the founder's scan levels (Level 1 basic optical scan; Level 2 non-destructive multispectral scanning; Level 3 theoretical holographic scan). Derivatives are described as linked to primes rather than as a level, and where born-digital sources fit in the scan levels is marked as an open question. New Section 4.7, Citation levels (Level 1 universal citation token; Level 2 registered citing work, designed but not yet implemented); former Sections 4.7 and 4.8 renumbered 4.8 and 4.9. Section 4.1 no longer refers to "Level 1 or Level 2" packages. Executive summary, Section 9, and the Phase 0 roadmap now describe the local v0 prototype instead of saying there is no working demonstration, and note that the prototype's capture-level labels predate the scan levels. Phase 1 roadmap and appendix wording updated to match. No other changes of substance.
- **v0.2 (September 2026), draft.** Previous version, archived.
- **v0.1 (September 2026), draft.** First draft.