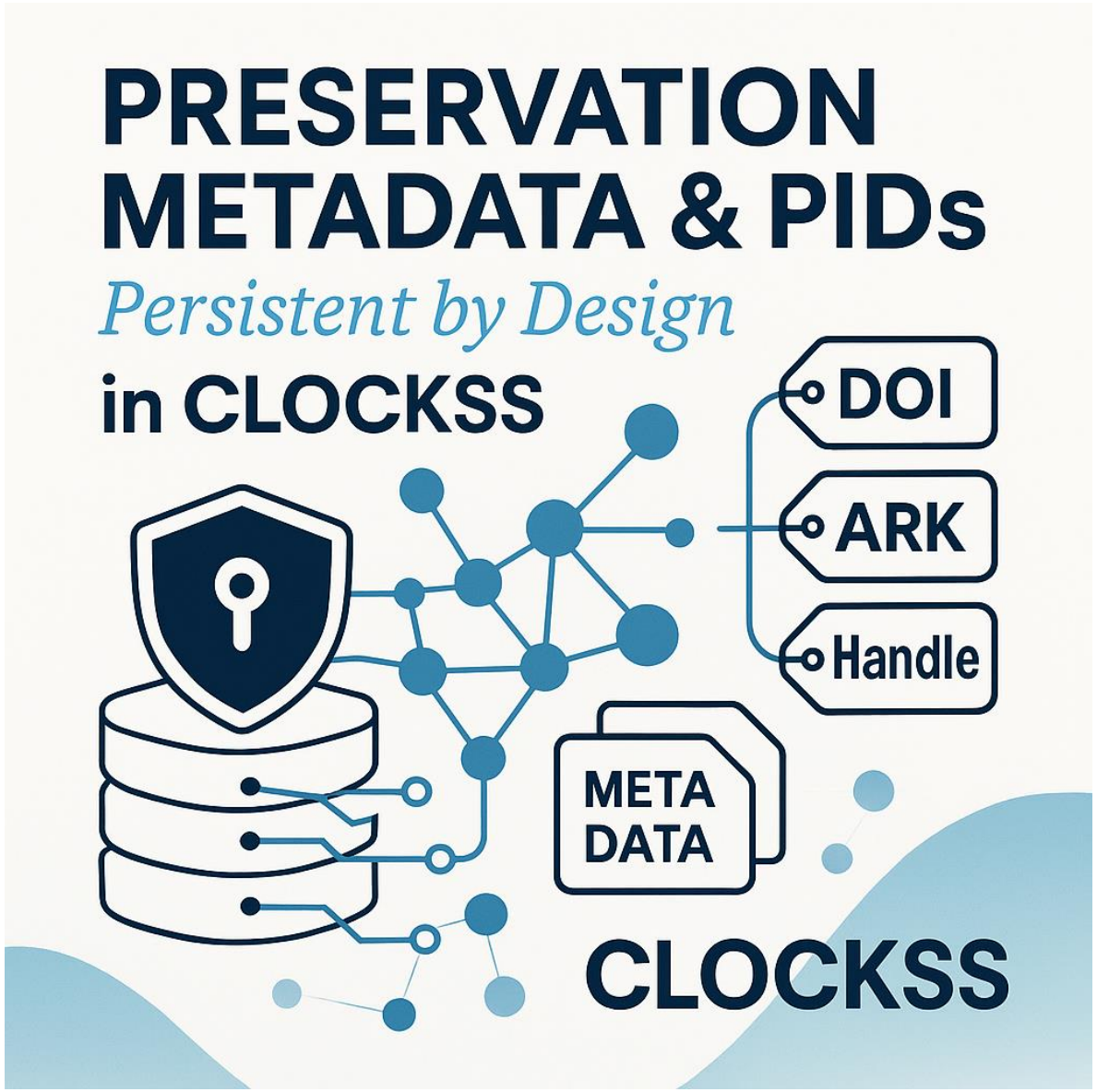


Preservation Metadata, PIDs, and the Long View: Persistent by Design in CLOCKSS

What is digital preservation?

- The active management of digital content to ensure it remains accessible, usable, and authentic for the very long term.
- Involves ongoing processes rather than a one-time action, recognizing that digital materials are vulnerable to technological change, file degradation, and loss.
- Ensures that future users can still open, read, interpret, and trust digital content, regardless of evolving formats, platforms, or hardware.
- Protects the scholarly, cultural, and historical record by maintaining continuity of access over decades or even centuries.

Why Digital Preservation Matters?

- Safeguards knowledge so it remains available even during crises such as war, natural disasters, or cyberattacks that can destroy or disrupt access to information.
- Protects valuable content from accidental loss caused by human error, technical failure, or organizational instability.
- Ensures continuity of scholarship, cultural memory, and institutional knowledge for future generations.
- Strengthens trust in the scholarly record by ensuring that research outputs remain stable, verifiable, and accessible over the long term.

Some Examples

- **PKP Preservation Network (PKP PN):**
  A free preservation service for journals using the Open Journal Systems platform. It automatically deposits content into a geographically distributed, LOCKSS-based network to guarantee long-term access.
- **SAFE Network:**
  A secure, decentralized storage system that distributes encrypted content across numerous independent nodes. This approach removes single points of failure and protects content even if parts of the network go offline.
- **Internet Archive:**
  A large-scale, nonprofit digital preservation initiative that captures and stores snapshots of websites, digital books, media, and datasets. Through the Wayback Machine and extensive replicated storage, it provides long-term access to materials that might otherwise disappear due to organizational changes, domain loss, or server failures.

About CLOCKSS

- Community-Governed Digital Archive.
- Non-profit partnership of libraries and publishers ensuring trustworthy stewardship.
- Dark Archive Model.
- Content remains inaccessible unless triggered by loss, ensuring preservation without duplicating access.
- Preserves All Content Types: Journals, books, monographs, datasets, and grey literature across disciplines.
- Global Redundancy and Security12 secure nodes worldwide guarantee resilience against technological and institutional failure.
- Supports Open Access and Equity.
- Triggered content becomes freely available to all, regardless of institutional affiliation.

The Importance of Digital Preservation in Ukraine During Wartime

Protecting cultural memory	Digital copies safeguard heritage and research that may be lost to physical destruction.
Ensuring academic continuity	Preserved content keeps education and research accessible when campuses or systems are disrupted.
Defending against cyberattacks	Secure, redundant storage protects information targeted by hacking or digital sabotage.
Documenting truth and accountability	Preserved records support fact-checking, legal evidence, and resistance to misinformation.
Maintaining off-site backups	Distributed preservation ensures survival of digital assets even if local infrastructure is damaged.

See our webinar on Ukraine preservation efforts https://youtu.be/6E5SeHVGYLM?si=5al9gLv2bDupBLH9

Preservation Metadata in CLOCKSS

Metadata is the foundation that makes long-term preservation possible

Without trustworthy metadata, we might still have the file years later but not understand it or trust it.

Integrity verification is continuous, not one-time event:

Digital preservation isn't "save it once and forget it."
Files can change or get damaged over time, so the system checks them regularly to make sure nothing has been altered, lost, or corrupted.

Formats and versions:

What format it's in (PDF, XML, JPEG), what version, and how it was created.

This information helps future systems read and interpret the file correctly, even when technology changes.

Audits and Repairs:

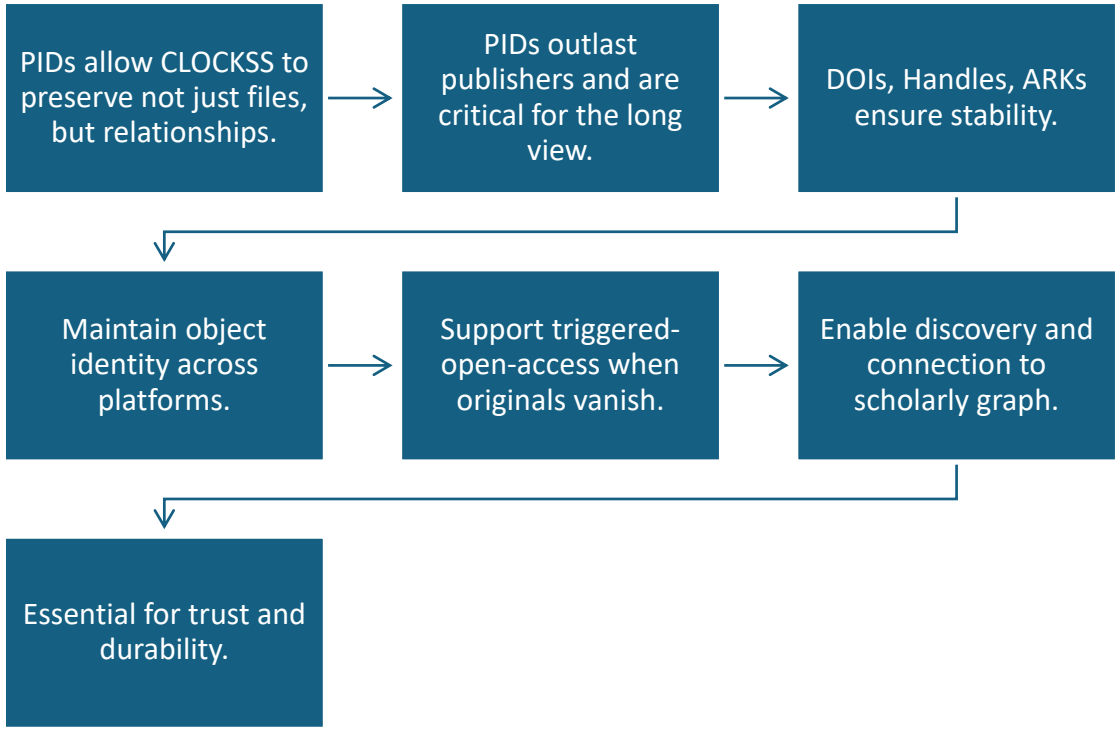
Regular checks and repairs are crucial to keep content safe.

By keeping all this information, we can be confident that:

Authenticity: The file is real and trustworthy.

Completeness: Nothing is missing.

Role of Persistent Identifiers (PIDs)



PRESERVED IN CLOCKSS

The Long View in CLOCKSS

- Preservation is not static; systems must evolve.
- CLOCKSS is structured to thrive beyond current technologies.
- Designed for decades, not product cycles.
- Resilient to tech turnover and publisher change.
- Durable community governance.
- Metadata standards built for longevity.
- Redundant, geographic replication.

Persistent by Design: CLOCKSS Practices

- Metadata-rich ingest workflows.
- Automated audit & repair cycles.
- Dark archive model reduces risk.
- Distributed nodes for resilience.
- Design embeds persistence at every layer.

Why Persistence Matters

- Prevents loss of scholarly knowledge.
- Protects against publisher/platform collapse.
- Guards against technological obsolescence.
- Ensures long-term, equitable access.
- Metadata + PIDs + CLOCKSS = sustainable preservation system.

How Ukrainian Publishers Can Protect Their Publications for the Long Term

Join trusted preservation networks: Register content with services such as CLOCKSS, Portico, or PKP Preservation Network.

Deposit content in multiple repositories: Share publications with national libraries, institutional repositories, and international archives for redundancy.

Use persistent identifiers: Assign DOIs, ORCID IDs, and standardized metadata to ensure long-term discoverability.

Publish in stable, open formats: Use well-supported formats such as EPUB, PDF/A, and XML to improve long-term accessibility.

Maintain clear preservation policies: State how content is archived, backed up, and migrated to new formats over time.

Back up locally and off-site: Store encrypted copies in geographically distributed locations to guard against physical and digital threats.

Thank you for your time and attention

Please feel free to get in touch at any time:

Alicia Wise - awise@clockss.org
Gali Halevi – ghalevi@clockss.org