# Generative-Lives-Bluesky [2025] – Working Brief

*Last updated: 2025-10-28*

## 1) Project Overview

A living document to coordinate the Generative-Lives-Bluesky [2025] research project. Serves as the shared reference for scope, methods, metadata, ethics, and decisions. We’ll update this as the project evolves.

## 2) Objectives (initial)

* Design an archival strategy for Bluesky/AT Protocol social media content.
* Balance provenance/series-level context with item-level description.
* Document capture, preservation, and access workflows that are resilient to platform change.
* Record decisions, open questions, and next actions in one place.

## 3) Scope & Boundaries (draft)

* **Platform focus:** Bluesky (AT Protocol). Consider cross-posts only insofar as they appear in Bluesky records.
* **Accounts/collections:** [List specific accounts or hashtags/campaigns here].
* **Inclusions:** Creator’s posts; [decide: include/exclude] public replies, quotes, reposts by others.
* **Exclusions:** Private/DM content; deleted content unless independently captured prior to deletion and ethically cleared.
* **Time span:** [Start date]–[End/ongoing].

## 4) Archival Approach (model)

* **Arrangement:** Account (series) → Thread/Campaign (file) → Post (item).
* **Principles:** Respect des fonds (provenance/original order); multi-level description; relationship-rich modeling.
* **Preservation:** OAIS lifecycle; PREMIS events for capture, normalization, fixity, migration.
* **Packaging:** Store ATProto repository **CAR** exports alongside **WARC** crawls of public profiles/media; maintain checksums and capture logs.

## 5) Metadata Schema (starter)

### A. Series (Account) fields

* Identifier(s): DID, handle(s), account ID; PDS/host history.
* Creator/Agent: name, role, affiliation.
* Scope & Content: purpose, themes, collecting policy.
* Dates: active date range, capture date ranges.
* Technical: capture methods (CAR/WARC), frequency, storage locations.
* Rights/Ethics: access restrictions, consent notes, takedown policy.
* Related: linked accounts, websites, projects.

### B. File (Thread/Campaign) fields (optional layer)

* Root post URI; summary/scope; hashtags; key events; date span.
* Relationships: in-reply-to chain, quotes, reposts.

### C. Item (Post) fields

* Stable ID/URI/CID; timestamp; text; language.
* Media: blob refs, media types, durations/dimensions.
* Context: in-reply-to, quote-of, repost-of, mentions, tags.
* Capture: method, date/time, checksum, normalization.
* Access: rights/visibility, redactions if any.

## 6) Capture & Preservation Workflow (draft)

1. **Selection & Scheduling:** Define targets and frequency (e.g., weekly CAR export + monthly WARC crawl).
2. **Acquisition:** Pull com.atproto.sync.getRepo CAR for each target account; perform crawl of public profile pages/media for replay.
3. **Normalization:** Verify fixity; normalize media derivatives for access; retain originals.
4. **Description:** Apply series/file/item metadata; link relationships.
5. **Storage:** Write AIPs to preservation storage with versioning; DIP generation for access.
6. **Monitoring:** Periodic fixity checks; log PREMIS events; track API/protocol changes.

## 7) Ethics & Rights (draft)

* Minimize harm: consider consent and expectations for third-party replies.
* Redaction/takedown: document requests and actions; maintain audit trail.
* Data minimization: where feasible, store references/IDs rather than duplicating third-party content; respect robots/noindex where applicable.

## 8) Tools & Endpoints (notes)

* AT Protocol: repo export (CAR); identity history (DID/handle changes); PDS migration notes.
* Complementary: WARC tooling for web UI/media capture; checksum utilities; description tools.

## 9) Decision Log

* **[YYYY-MM-DD]** Decision – Rationale – Impact – Owner.

## 10) Open Questions

* Include third-party replies? Under what conditions?
* Preferred access platform (replay vs. finding aid vs. dataset)?
* Frequency and retention periods for captures?

## 11) Next Actions

* List initial target accounts and date ranges.
* Choose inclusion policy for replies/quotes/reposts.
* Set capture schedule (CAR/WARC) and storage paths.
* Draft detailed metadata application profile (MAP).

## 12) References (to be populated)

* Standards (ISAD(G), DACS, RiC-CM, OAIS, PREMIS), community guidance (DocNow, IIPC/DPC), platform docs (ATProto).

## 13) Style & Taxonomy (working)

* Controlled tags/subjects; naming conventions for series/file/item identifiers.

**How to use this document** - Treat this as our shared “brief.” We will update sections as we decide policies and complete tasks. Add specifics (accounts, dates, policies) inline as they are confirmed.