🌸 Document Overview

flickr.org

Project Report

Project:	**Building a “Data Lifeboat” for Flickr** *To support the development of a mechanism called “Data Lifeboat” which gathers photograph collections hosted on Flickr.com for preservation elsewhere.*
Project Director:	George Oates Flickr Charitable Foundation
Funder:	Mellon Foundation
Grantmaking area:	Public Knowledge
Date of award:	1 May 2024
Published:	May 2025
Co-authors:	Fattori McKenna, Ewa Spohn, Alex Chan, Dan Catt, Jessamyn West, Joshua Christman, George Oates

Table of contents

Executive summary

This document contains findings from co-design workshops and in-depth interviews conducted with digital cultural heritage practitioners in Washington D.C. and London during October-November 2024. Funded by the Mellon Foundation Public Knowledge Grant, this research explored the development of the Data Lifeboat **tool for preserving networked image content from Flickr** and the speculative Safe Harbor Network of **trusted institutions for maintaining Data Lifeboats in the long-term.**

Our research revealed a strong institutional need for tools that preserve the valuable content and rich contextual information of networked images from social media platforms, such as Flickr. Practitioners identified several possible institutional use cases for Data Lifeboat, **from streamlining metadata collection to securing critically at-risk content.**

Ethical considerations also emerged as central to the networked image preservation process. Drawing from Indigenous data sovereignty frameworks like the C.A.R.E. principles, we've enhanced the Data Lifeboat tool with reflective README prompts that encourage creators to consider issues of **purpose, future access, storage, context, cultural sensitivities, privacy, and copyright**. Our research also established the viability of a Safe Harbor Network while identifying key **governance, policy, and resource challenges** that need addressing.

Based on these findings of this research we have adjusted our work on the Data Lifeboat tool and Safe Harbor Network development.

The final version of this report was written in March 2025.

Background

Data Lifeboat is a software application that creates a **downloadable, compressed file containing images and their metadata** – that means comments, tags, galleries; all the stuff that gives Flickr its uniquely social character. Data Lifeboats are designed to be self-contained, long-lasting, versatile, and readable. They are a **flexible container** allowing anyone with a Flickr account to gather networked images from the platform to archive elsewhere.

For the purposes of this phase of research, we were interested in the ways that the Data Lifeboat could be used in an **institutional archival context**, for the purpose of safeguarding the shared digital cultural heritage that lives on Flickr.com. In October and November 2024, the Flickr Foundation convened two co-design workshops with practitioners and advocates in the field of digital cultural heritage in Washington D.C., U.S.A and London, U.K. These workshops, funded by the Mellon Foundation Public Knowledge Grant, focused on the development of the Data Lifeboat tool, its real-world applicability and ethical implications, as well as the creation of a speculative Safe Harbor Network; a decentralized cohort of trusted institutions that would maintain Data Lifeboats for generations to come.

For those who were unable to participate in the workshops, we facilitated in-depth interviews (IDIs) with archivists and administrators within cultural heritage organizations to ensure their views were represented in the research. The majority of those interviewed were Flickr Commons members and therefore have a longstanding relationship with the Flickr platform. Our intention is that these workshops and interviews directly inform the next phase of development for Data Lifeboat (alpha release set for mid-2025) and the Safe Harbor Network.

This report outlines the key research findings suggesting recommendations and priorities for our work to come. These findings predominantly speak to the institutional use case for a Data Lifeboat and Safe Harbor Network: how these tools can assist existing missions in digital preservation and expand the repertoire of what can be collected and secured for the public good in the long-term. Future research, beyond the scope of this grant, will consider the Data Lifeboat from the perspective and needs of Flickr members (and other social media contexts).

Introducing the case for a Data Lifeboat in digital heritage

> *"We are still grappling with born digital… in fact,* ***we have never ingested born digital content"*** *— Flickr Commons member, National Library*

Persistent issues in social media preservation

Many of the issues around digital preservation have not improved since the advent of digital artifacts almost half a century ago. Despite advancements in hardware and software capabilities, greater social awareness about the importance of digital preservation and a richer understanding of online risks and harms, there is a lack of consensus on how to move forward with digital preservation, but nevertheless a strong desire to experiment and collaborate.

Social media preservation in particular, research participants claimed, is ad-hoc and incomplete at best, with no agreed-upon industry standard to date. Many attempts to systematically catalogue social media have been abandoned due to the sheer scale of the challenge, changes in platform policies, inadequacy of corporate cooperation, and lack of institutional awareness or value.

Across our workshops and interviews, four themes emerged as major challenges for archivists to systematically and confidently archiving social media content: technical challenges, stakeholder imbalances, resource management, and legal and ethical concerns. The context of these challenges is something we need to take into consideration if we propose introducing a new tool to the preservation workflow.

Technical challenges

The infrastructure required to properly archive and maintain social media presents numerous technical hurdles that lack clear solutions.

Issues include:

- Inconsistent storage formats or environments
- Nested content dependencies
- Lack of common metadata standards
- Keeping files connected to the metadata
- Intensive file fixity workflows
- Versioning control

- Bit rot

Stakeholder imbalances

The relationship between social media platforms and archival institutions is characterized by severe power asymmetries, where archivists must operate within constraints established by commercial entities that prioritize profit over preservation:

Issues include:

- Proprietary formats
- Platforms revoking access on a whim, or disappearing entirely
- Dependence on storage monopolies (often at the mercy of their pricing structures)
- Hostile platform management structures
- APIs designed for commercial, not archival, use
- Lack of platform recognition for archival duty

Resource management

Institutions tasked with preserving social media face constraints on financial, human and environmental resources. Frequently, social media preservation lands at the 'bottom of the pile' in terms of resourcing priorities, especially when a growing volume collides with diminishing capacity.

Issues include:

- Downward economic pressure
- Archivists required to advocate for value of social media collecting within institutions
- Staff turnover
- Overwhelmed by content scale
- Storage costs
- Stability and access of archival storage materials (e.g. motion picture film)
- Long-term climate sustainability

Legal and ethical concerns

The networked nature of social media content creates a complex web of stakeholders who must be considered in terms of rights, privacy, and ethical concerns — a level of complexity that traditional archival agreements rarely need to address.

Issues include:

- Ever-changing legal contexts for rights and permissions
- Cross-jurisdictional complications (when digital content is location-less)
- The emerging rights of the photography subjects
- Negotiating and managing Right-to-Remove claims (a “ticking time bomb”, claimed respondents)

Content and selection

Archivists face the monumental task of determining what social media content to preserve while simultaneously advocating for the value of preserving those materials in their institutions.

Issues include:

- “An ocean of images”
- Hard to have a handle on exactly what is available
- Content made private or taken offline
- Unclear selection criteria
- Lack of trained experts on social media curation
- Loss of original context and experiential data
- Minimal requests from researchers (”they say, ‘isn’t it all available online anyway?’”)

Data Lifeboat is **not a panacea for the persistent issues** in social media preservation. Instead, it attempts to provide a solution to a small slice of the puzzle by preserving user-selected collections from Flickr.com in a **simple to use, self-contained, long-lasting package that is conscious of the unique ethical and legal issues** that networked social images raise. We believe this packaging format could serve as a model or standard for archiving other types of social media content beyond Flickr.

Responses to the Data Lifeboat concept and prototype

Institutional use cases for a Data Lifeboat

The following possible use cases for a Data Lifeboat emerged from our workshops and interviews with archivists. These suggestions extended our initial ideas on how a Data Lifeboat could support archival activities, and we were pleased to hear the ways the tool might be used to fit institutional needs:

1. On-ramp to social media or born-digital collecting[1]

> *"More and more donors are asking us to take born digital records, and we don't have a way to hold it" - Flickr Commons member, Provincial Archivist*

For the reasons discussed in the previous section, few archives have successfully ingested social media content, yet there is an appetite to capture the rich content and conversations taking place on these platforms. Data Lifeboat takes Flickr.com as its starting point, an important locus for many of our Commons members who have seen their digital collections and engagement grow from strength to strength.

Often when archives (or individuals) request and download content from platforms, they receive material that is machine-readable but ultimately oblique, poorly structured, filled with unnecessary elements (e.g. font libraries or extraneous scripts). They end up largely unusable to people without specialized technical knowledge. This makes it difficult to integrate into archival systems — and even more challenging to make accessible and engaging for an audience.

Data Lifeboat, by contrast, provides a legible package in an accessible, explorable format that *only* contains the data you need: images and metadata. The HTML-based viewer we've designed proved popular and immediately intuitive and engaging, with participants suggesting this is what they would prioritize displaying for audiences:

> *"I like the viewer, it functions as a sort of codebook for collecting social media: 'Here's how you use it and here's how you look at it'" — Flickr Commons member, Military Archivist*

Data Lifeboat can serve as a stepping stone between archives securing digitized collections and opening up a world of contemporary networked image collecting. It functions as a tool for archives to discover and understand the wealth of cultural heritage that exists on Flickr, injecting a new energy into the site as archivists go digging for the artifacts they might want to bring into their collections and make accessible for the public:

---

1 There were also some suggestions of using Data Lifeboat as a bulk download tool from Flickr. With Flickr's current tools, much of the work of download, filtering and linkage takes place manually, whereas using Data Lifeboat could significantly streamline the process. However, this approach could be seen as extractive and against the ethos of conscious archiving (e.g. not using the ReadMe function towards its intended ends). Ultimately **we would like to discourage this practice** (so as not to simply create a Flickr v. 2.0) - and we will likely set a limit for the size of Data Lifeboats.

"We want to expand what's available to people… these collections are not intended to be under lock and key" — Flickr Commons member, University Librarian

2. Preserving and reintegrating social context

"For us the most valuable thing on Flickr is the social context… every day you get a story" — Flickr Commons member, Military Archivist

Many archivists recognise that an artifact's contextual information is just as valuable as the content itself — but this social layer is the most difficult to capture and preserve. Our Flickr Commons members expressed the desire to have a permanent record of engagement around their images (and others') on Flickr. In many cases, tags and comments have been invaluable in reshaping their internal catalogues, adding or correcting information. For example, in their first year of hosting on Flickr, The Library of Congress pictures received 67,000 tags and 7,000 comments. As a result thousands of Prints and Photographs Online Catalog (PPOC) records have been updated based on Flickr audience engagement. At other times, this social context can help situate historic collections in contemporary conversations (see example given in Ethical Considerations: Metadata below), thereby creating a snapshot in time.

Having a mechanism to seamlessly integrate this rich, community-generated knowledge back into institutional collections — whether for analysis, reflection, or simply safekeeping for posterity — would provide significant value to archives. A Data Lifeboat could potentially simplify the current extensive manual review process, streamlining the attaching and cataloguing of metadata, which is particularly important given the reality of limited time and resources.

Whilst many platforms offer users tools by which to export social media data, the richness of engagement is often lost in this process — either because it's not included in the package or it is near impossible to read or access. Data Lifeboat remedies this gap by structuring metadata in a light, machine-readable format, which includes a human-readable viewer, that can be transported between file management systems with relative ease. For smaller archives especially, which rely on engagement metrics for funding, Data Lifeboat provides a seamless way to demonstrate the impact and reach of their collections in a format that's both accessible to users and compatible with institutional systems.

3. A curatorial tool

In an era of overwhelming digital abundance, thoughtful selection is becoming more crucial than ever. With extensive backlogs for appraisal, archivists know that keeping everything is neither feasible nor desirable.

> *"Data Lifeboat is a good platform-based incentive to slim down archives, keeping on what is **needed and meaningful**" — Flickr Commons member*

Data Lifeboat presents an opportunity to decentralize curation through its straightforward workflow, designed to be equally accessible to professional archivists and community members. While institutions often lack time to sift through vast social media repositories, citizen-driven curation using Data Lifeboats could effectively fill this gap. This approach represents a radical departure from traditional archival deposit and appraisal systems, which have historically concentrated decision-making power in institutional hands, by diversifying *who* gets to determine what becomes part of our collective heritage.

This approach aligns with emerging best practices in participatory archiving (such as the National Museum of African American History and Culture's Community Curation Program or Collecting Social Photography Project) while providing the technical infrastructure to make such collaboration feasible. With the labor of selection (simple URL inputs) and description (keeping metadata attached to images) primarily borne by the Data Lifeboat tool, archives can focus on their role as safe-keepers, capturing and preserving only what has been deemed valuable to their audience rather than needing to sift through excess material.

Data Lifeboat is distinctive in its attempt to preserve the relationships between networked images— addressing the loss of original order that many archivists identified as a key concern with digital collections. One Municipal Archivist noted their institution's shift toward scanning entire album or scrapbook pages instead of individual images. The connections between images and their placement can create meaning greater than the sum of its parts, which the Data Lifeboat aims to maintain.

4. Securing orphaned or at-risk archives

> *"There's an old guy in our community and no one in the family wanted his pictures, so we've taken them in. Maybe a Data Lifeboat could be for something like this. **It makes a temporary home for it now keeps something available for the future when somebody might want it"** — Flickr Commons member, Community Archivist*

As we have described, digital content is uniquely vulnerable — accounts are abandoned, passwords lost, personnel move on. Data Lifeboat can offer a critical intervention tool for these situations. What distinguishes Data Lifeboat from other preservation approaches is its capacity to allow creators to save *other people's photos* ethically. Currently, rescuing someone else's digital content — whether from a recently deceased loved one or a dormant community project

— requires navigating lengthy proof-of-identity processes that often fail to accommodate the urgency of preservation needs.

Nevertheless, Data Lifeboat respects ethical standards through structured safeguards: a mandatory 14 day notification period, preservation of existing privacy restrictions in the Data Lifeboat, and README prompts that encourage reflective consideration of privacy and consent implications.

This use case becomes increasingly urgent as we witness the planned removal and disappearance of content from social media platforms — content that constitutes critical elements of our shared cultural heritage. For instance, we secured all USAID (United States Agency for International Development) Flickr content into (prototype) Data Lifeboats following the President's announced closure of the agency in February 2025.

> *"I've spotted another Congregation of Sisters that has a Flickr account but their owner seems to have vanished. We'd really like a way to be able to save that and bring it under our wing" — Flickr Commons member, Religious Archivist*

5. Additional archival storage

Perhaps closest to our initial intention when developing the concept: for archivists, the Data Lifeboat can function as a way to securely store networked images. The Data Lifeboat provides another secure back-up, aligning with the archival mantra, "Lots of Copies Keeps Stuff Safe" (LOCKSS). This proved particularly relevant for smaller archives and archives with fewer resources, for whom secure storage is a critical pressure point in long-term preservation:

> *"I'm literally storing these [images] on my own store-bought hard-drive" - Flickr Commons member, Community Archivist*

In this case, Data Lifeboat becomes another tool within a digital preservation ecosystem. It should be noted, however, that Data Lifeboats are less likely to be implemented in this way within larger, more established institutions. These organizations typically have their own systems, rules, and programmed conditions for managing and maintaining storage—meaning retrofitting a Data Lifeboat for its storage capabilities would be a hindrance to their existing processes. In both cases, however, this storage is rarely purpose-built for networked images or social media content, which leads us to our next use case.

Suggested Data Lifeboats

During our workshops and conversations some prevailing themes for Data Lifeboat contents emerged that could be valuable to institutional collections.

‘Time and Place’

A Data Lifeboat to capture a particular moment in time as experienced by a specific community. The format lends itself to saving a discrete collection (much like a Flickr Gallery), the value of which can be described by the creator. For example:

- The annual **Silver Bells Parade** in Lansing, Michigan, photographed by the local community over the last 50 years
- The **Notre Dame, Paris fire** and reconstruction as documented by city residents captures a specific event that shaped the city’s contemporary history.

Securing at-risk histories

A Data Lifeboat can be a lifeline for securing memories of sites, customs, communities at-risk of disappearance or erasure. Even when they have gone, the Data Lifeboat lives on to preserve this record. For example:

- The **Lobster Traps of Port Morien**: key to the community’s history, the knowledge and tradition of making lobster traps is fast disappearing as older residents pass on and the economy diversifies. Data Lifeboat can be a means to secure this element of intangible cultural heritage, capturing not just visual evidence of the practice but also instructions on how to enact it in future.

Crowdsourced collections

As libraries and archives increasingly outsource curation to capture a more diverse range of perspectives, Data Lifeboat can be host to intentionally created or commissioned collections, either drawing from existing collections on Flickr, or photographers uploading new content to the platform. For example:

- A **teenage girl from Whitechapel** curates a set of photos from Flickr that represent ‘her borough through her eyes’ — Brick Lane, where she goes shopping, Shadwell Basin where she swims with her friends, the East London Mosque, Stepney City Farm, the Elizabeth Line. These are donated to the **Museum of London** to showcase fragments of what it means to be a young person living in the city today.

- A photo from **every Boston city resident** on Flickr could be commissioned by Boston City Library to integrate local, contemporary perspectives into the collection.

Longitudinal collections

Given Flickr has been around for two decades, many photos replicate the same subject, often over time. Compiled and viewed in a Data Lifeboat, we could observe the changes in a natural or urban environment over an extended time period. Creating an invaluable record of (crowd-sourced) environmental change, ecological shifts, and biodiversity fluctuations, these collections become more historically significant with each passing year. For example:

- A longitudinal Data Lifeboat collection documenting specific sections of the **Great Barrier Reef** would create a visual chronicle of ecological transformation, capturing the stark contrast between vibrant coral ecosystems and the ghostly aftermath of mass bleaching events.
- A Data Lifeboat of the **rewilding of the Iberian Peninsula** would track the progressive restoration of native ecosystems, visually mapping how environmental interventions have transformed degraded landscapes into thriving habitats for returning flora and fauna, such as the Iberian lynx and Cinereous vultures.

The Web 2.0 Era

> *“I don't want to lose the first 10 years of my online life... All the commentary, tags and albums - messages from friends, that's important to me" — Flickr Commons member, District Archivist*

Data Lifeboats could be collecting tools for documenting the period of 2004-2014: the Web 2.0 era, Flickr’s most active period. There have been many laments that this period constitutes a digital dark-age, as camera and computer hardware developed rapidly and many websites went offline taking their content with them. Flickr is a rarity in still keeping this content available. For example:

- The **‘What’s in My Bag’ Flickr Tag** inadvertently preserved visual records of technology and personal ephemera from this period in time: camera equipment, iPod Nanos, flip phones, lipglosses, magazines. Once a social media trend, now this is a record of personal objects in the mid-2000s.
- The **‘Artifacts and Holdovers’ Flickr Group** documents “things nobody would have thought to photograph when the object was ‘in its prime’ or ‘of the time’ — but now they jump out at us as flashbacks”. As a Data Lifeboat, this group could showcase the urban fabric of the mid-2000s — payphones, decals, advertising and signage — which is hard to find collated at this volume elsewhere.

- A Data Lifeboat of Flickr's **'fav1000' tag**, acknowledging that the platform itself is a type of cultural artifact worthy of documentation. Preserving Flickr trends constitutes a record of how early social photography platforms shaped online visual culture and communities.

Besides these themes, there was also a significant appetite for Data Lifeboat as a personal archival tool. This is something we endeavour to explore in future research phases.

Enhancing the Data Lifeboat prototype

Building from the learnings from the workshops, conversations and consultations with legal and technical experts, we are working to implement the following additions to the Data Lifeboat's upcoming Alpha release:

Adjustments to Data Lifeboat contents:

- Video support
- Inline maps, with more accurate location information
- Inclusion of checksums
- Networked order preservation (e.g. for galleries, groups)
- A global list of tags and contributors

Data Lifeboat creation process:

- Preview / Review stages
- Support for including photos where the owner has disabled downloading
- Better error handling
- Design improvements
- Creator homepage with documentation and example Data Lifeboats for download
- Basic sharing instructions
- Limit to Data Lifeboat filesize

Legal & Ethical developments:

- Robust Notification and redaction procedure
- README flow based on C.A.R.E. principles
- Identifying the necessary changes to the Flickr.com Terms of Service and Privacy Policy
- A Terms of Service, Privacy Policy and Data Lifeboat Creator agreement that is consistent with privacy and data protection laws
- README flow determined by contents' permissions

Ethical considerations in networked image preservation

Archives are predominantly shaped by analogue-mindset, often retrofitting digital and networked acquisitions into pre-existing ethical frameworks. This is steadily changing thanks to the work of digital advocates, training and coalitions, but workshop participants raised the need for institutions to consider the unique issues raised by digital content, in particular networked images.

Ethics in the digital realm is an established field of critical inquiry and whilst we have already broached some of the legal implications of collecting social media content, we believe there is a responsibility to broaden our responsibilities to photo creators and subjects. Legal **compliance represents merely the baseline** of practice — a minimum threshold that fails to address the full spectrum of ethical considerations. Instead, we must ask: "How can we design a Data Lifeboat tool that transcends mere legality to become **ethically exemplary**, setting new standards for responsible networked image preservation?"

We asked participants to list their main concerns, the following issues were raised. Whilst some of these may overlap with offline artifacts, the digital nature of networked images introduces unique complexities that require **specialised and proactive** (instead of reactive) approaches. We need to be overly cautious in our handling of networked images, as Stuart Hall has argued, the future uses of archives "can never be foretold" (2001, 92).

Personally Identifiable Information (PII)

Networked images pose unique privacy challenges. Their contemporary, user-generated nature means that networked images often contain identifiable information of **living individuals**, at a scale far exceeding what traditional archives typically manage. And yet it is these contemporary subjects that constitute precisely what makes networked image preservation so important.

The disclosure of PII in the networked context may be deliberate on the part of the content's creator, but there are many cases where it may have been **unintentional**. An oversight or broad-brushing of privacy settings may result in images intended for private use being broadcasted to the public realm. Similarly, the metadata embedded within, or associated with, networked images — such as geolocation data, usernames, descriptions, tags or comments — can result in unintended information exposure. For example, in the metadata stimulus we shared at the workshops, an email address was displayed in the comments — once included in a Data Lifeboat this could potentially be stored for decades to come or accidentally made public.

Workshop participants also noted that our understanding of what constitutes personally identifiable information in networked images will likely evolve as technologies and legal frameworks develop, potentially **expanding future ethical obligations** for both image creators and subjects. Additionally, the interpretation of a right to privacy differs across geographies; under the E.U General Data Protection Regulation (GDPR), for instance, the representation of someone's likeness can be interpreted as PII, which significantly complicates subjects' rights in photographs. This regulatory landscape **varies considerably by jurisdiction**, as in the case of *Freedom of Panorama,* which restricts the photographing and sharing of public spaces.

Takedown requests within institutional contexts are both time intensive and require specialist expertise. While participants reported having experience with such requests, they noted that these have generally been few and far between. However, the abundance of potential personally-identifiable information embedded within networked social images poses a formidable and potentially overwhelming task that many archives at present actively avoid assuming.

What this means for Data Lifeboat:

- Clearly communicate the risks and considerations involved in unintended PII capture during the Data Lifeboat creation flow (see README questions)
- We need to adopt a selective approach to metadata preservation — keeping only what is valuable — rather than a capture-all approach
- Be pre-emptively cautious with metadata, recognizing that our understanding of PII in digital contexts may develop throughout the lifespan of a Data Lifeboat
- Consider implementing tools to scan metadata for PII in Data Lifeboats intended for public access (such as in the Safe Harbor Network)
- Develop a robust and supported ‘takedown’ policy for Data Lifeboats hosted in the Safe Harbor Network that minimizes resource demands on the archives themselves.

Cultural and contextual privacy

Participants recognized that whilst preservation of digital cultural heritage is important for fostering collective understanding, certain materials within networked image archives contain culturally significant or sensitive content that require **thoughtful access controls**. Some images may document cultural or spiritual ceremonies and thus should not be hosted digitally as doing so may risk diminishing their spiritual dimension — or should only be accessed following certain procedures or initiation rituals. In contrast, it may be important to store a sensitive historical image as a record of past injustices to ensure that they are not forgotten or repeated — but access to such materials may require extensive contextualisation to avoid misinterpretation or misuse. As one Women’s Studies librarian noted, regarding the misleading ethical compulsion of artifacts shared in the Public Domain:

> *“Not everything is meant for everyone” — Flickr Commons member, University Librarian*

The digital realm makes it difficult to enforce these careful provisions, as images frequently become detached from their original context and re-appropriated. Traditional archival safeguards such as having a trained, specialist librarian or curator present to mediate access are considerably more **difficult to replicate in the digital realm.**

However, the strides that have been made in traditional archives show that the simplistic binary approaches of complete access or total restriction need not be the only options for networked images. More **nuanced approaches include mediated access**, reflective compilation, and persistent privacy and sensitivity settings that endure over the long term. **Providing specialized tools** to community members most directly impacted by culturally sensitive content has emerged as an established best practice:

What this means for Data Lifeboat:

- Maintain privacy and sensitivity settings within the Data Lifeboat
- In the README, incorporate prompts for Data Lifeboat creators to think about the possible spectrum of cultural sensitivities that their content may raise
- In the README, provide a dedicated space for at-risk creators to qualitatively state their wishes for access, storage, and use of the Data Lifeboat contents
- Consider how the Safe Harbor Network, as a group of ‘trusted institutions’ (well-versed in negotiating cultural sensitivities in their existing archival content) can function as as arbiters and guardians of Data Lifeboat content in the future.

Authenticity

Issues of provenance take on new dimensions with networked images that resist simple transposition of procedures from traditional archival systems. The lack of clear provenance in many networked images complicates verification processes, raising questions about how to confirm an image matches not only its manifest but its true point of origin.

The preservation of networked images complicates traditional notions of authorship. Participants discussed how establishing provenance for social media images is frequently a challenge due to sharing, remixing, and platform-specific attribution issues.

Participants highlighted how networked images are particularly vulnerable to becoming separated from their original context and sequence. A networked image hosted on Flickr can exist many steps removed from its creator - for example, an image that has been scanned and edited - so reflecting (as far as possible) this chain of development is key. This has long-standing implications for copyright claims.

Perhaps more concerning is the possibility for preserved networked images to solidify *inaccurate* information, particularly as tampering with digital materials becomes increasingly prevalent in unstable political contexts. Implementing preventative measures that can flag manipulation or changes to preserved networked images is crucial.

What this means for Data Lifeboat:

- Include changelogs and manifests in a Data Lifeboat that provide a record of the original, intended content
- Host and maintain a ledger of Data Lifeboats created
- Determine what is an appropriate amount of data to store in the ledger that maintains fidelity of content verification without compromising privacy
- Consider how a Safe Harbor Network of 'trusted institutions' may be able to enhance and validate this authenticity.

Metadata

The rich contextual information embedded within and surrounding networked images is **precisely what makes them so valuable** in archival contexts. Because of this, networked image preservation warrants the prioritization of the fidelity and accessibility of this metadata for future audiences. Unfortunately, this metadata is often buried in networked images or made overly complicated to access by platform hosts — rendering it difficult for smaller archives to access, parse, store, and eventually display this trove of information. Beyond the technical challenges of metadata preservation, we must be mindful of its dual capacity to both enrich our understanding of preserved materials and introduce ethical complexities that require careful treatment.

Metadata can provide essential contextual framing for images, **offering unique insights into contemporary perspectives** as was the case of the photo stimulus of the Library of Congress' image of *Negro Boy near Cincinnati, Ohio.* Taken in c. 1942, the image title as presented by the Library of Congress was fervently debated in the comments on Flickr. By preserving these discussions alongside the image itself, we are able to document evolving sentiment around racial nomenclature and representation in the 2010s, contributing to a greater understanding of the topic.

When examined alongside its associated metadata, an image's perceived meaning may transform dramatically. As one participant responded to the photo stimulus of two women laughing and eating cake at a work party:

> *"The metadata made me interpret [the photo] completely differently" — Mellon workshop participant, Legal Scholar*

In this case, the comments transformed an innocent photo of women having fun into a potential site for misogynistic treatment of the subjects, as commenters highlighted potential lurid undertones and requested the photo be included in Groups such as 'food and chics *[sic]*'. This raises important questions about how metadata can complicate the original intentions and how storing it may potentially harm the image creator or subjects.

While metadata can add valuable context, it can also be misleading if not reviewed. For instance, EXIF data from networked images can be inaccurate, especially when older images are digitized and reshared.

What this means for Data Lifeboat:

- Data Lifeboat presents an opportunity for Data Lifeboat creators to preserve images along with their metadata with relative ease and in a readable format
- Data Lifeboat creators must be made aware of the potentially erroneous, misleading or harmful nature of metadata
- As such Data Lifeboat creators will be given space to review and reflect upon the metadata included — through the preview function — and presented with opportunities to annotate or add context to this — through the README.

Given these unique considerations, it is critical that Data Lifeboat accounts for the legal and ethical sensitivities that networked images present. In seeking to address these concerns, we have drawn upon the groundbreaking work of indigenous activists who have long championed more equitable, care-informed, and community-centered approaches to archival contexts.

Learning from Indigenous data care practices

> *"There's certainly a lot of work that goes into negotiating take-down requests but it's important to do right by the community… **we serve the needs of the community and how they want to be described and depicted**" — Flickr Commons member, Provincial Archivist*

To carefully handle the myriad of complexities inherent within a huge-scale digital archive of networked images, we recognized the need to **learn from substantive dialogues within the field of reconciliatory justice in archival practice**. Archival institutions have long functioned as both practical and symbolic containers of nationhood and identity, placing them at the center of critical discussions about ownership, access and representation. Many of our workshop participants were actively engaged in reconciliatory work, including negotiating "take-down" requests for contested materials in their collections, so these are **concerns we need to be responsive** to in building Data Lifeboats.

The true leaders in this field are Indigenous activists who have developed frameworks that fundamentally challenge both what is preserved and how preservation systems operate. From NAGPRA (Native American Graves Protection and Repatriation Act), which enables tribes to reclaim human remains and sacred objects through consultation with museums, to Traditional Knowledge labels that identify Indigenous protocols for accessing knowledge within existing collections, these efforts have contested not only the contents held within institutions but also the systems through which they are catalogued, maintained, and accessed.

In recent years this work has expanded beyond traditional, archival institutions to address contemporary challenges posed by Big Data and Machine Learning — from entrenched bias in datasets to algorithmic opacity. Indigenous groups such as the Te Mana Raraunga Māori Data Sovereignty Network, the US Indigenous Data Sovereignty Network, and the Maiam nayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Collective have led efforts to articulate ethical frameworks for data governance that center community needs and values. These collaborative efforts culminated in a global, inter-tribal workshop in 2018, which formalized the C.A.R.E. principles. Published by the Global Indigenous Data Alliance, these principles propose a governance framework with **people and purpose at its core.**

The **C.A.R.E. principles** bring to the fore four essential values around data:

1. **Collective Benefit**: Data must enhance collective well-being and serve the communities to which it pertains.
2. **Authority to Control**: Communities must retain governance over their data and decide how it is accessed, used, and shared.
3. **Responsibility**: Data handlers must minimise harm and ensure alignment with community values.
4. **Ethics**: Ethical considerations rooted in cultural values and collective rights must guide all stages of the data lifecycle.

As Data Lifeboat sits at the intersection of technology and cultural heritage, we recognized the importance of building upon this foundation rather than reinventing ethical frameworks. While Indigenous cultural heritage inevitably exists within Flickr's collections — particularly among Flickr Commons members pursuing their own reconciliation initiatives — **the value of these principles extends beyond Indigenous cultural heritage**, serving as a foundation for ethical data practices that **benefit *all* data subjects** in the age of Big Data.

README as a C.A.R.E.-full response

> *"The README acknowledges the **impossibility of objectivity**, which is often the archivist's pitfall" — Flickr Commons member, University Archivist*

> *"Whilst you can't guarantee it'll be carried out, if you put it in the README, it at least shows you're **thinking about the collection in the long-term**" — Flickr Commons member, Military Archivist*

In order to not replicate the mindset of perpetual accumulation that plagues both museums and Big Tech alike, and to be responsible to current and future data subjects, one of the possible ways we believe we can engender conscientious and careful collecting is through the inclusion of a README in a Data Lifeboat — a series of guided questions with free-text input, the responses to which will be displayed at the top of Data Lifeboat when opened, thus operating as a sort of 'Note to the Future'

READMEs are files traditionally used in software development and distribution that contain information about files within the directory. We aim to extend the README's purpose beyond a description of the files, giving Data Lifeboat creators the space to **add detail, nuance, and context** by reflecting on and writing about the collection as a whole, or the specific images it contains. This is particularly important for digital cultural heritage datasets, as it is frequently an issue that images arrive in archival collections without context or may be digitised and uploaded without adequate or consistent metadata (see Gebru et al., 2021 and Alkemade et al., 2023). At the very least, the inclusion of a README in the workflow encourages Data Lifeboat creators to slow down and think critically and carefully about the contents they are saving.

Through the README, we are seeking a means to **capture intention without promising action** (for we are no longer in control)**.** We ought to ask, had early collectors had recorded the true will and intention of the original communities they took from, would the same archival injustices have been carried out? Perhaps the answer is still 'yes', but at the very least there would be a record of diverging from those intentions.

README prompts for Data Lifeboat creators

> *"The README functions as a sort of **cipher**, it's a possible way to reconstruct the Data Lifeboat in the future if it becomes unmoored." — Flickr Commons member, City Archivist*

In the prototype shared at the workshop, the prompts for the README were as follows:

- Tell the future why you are making this Data Lifeboat.
- Is there anything special you'd like future viewers to know about the contents? Anything to be careful about?

While these questions were a good start for getting potential Data Lifeboat creators to think about the intentions and future reception of the contents or collection, we wanted to apply the specificity of the considerations raised above vis-a-vis networked image preservation.

The result of our question-plotting exercise was a list of topics that will be surfaced during the Data Lifeboat creation process, collated from the responses of workshop participants. Certain questions, we decided, ought to be mandatory for all Data Lifeboats—because we want all creators to be thinking about the afterlife of their images. Other questions will be surfaced and strongly suggested to creators when including content that is not their own or contains specific rights or content restrictions.

In the creation flow, Data Lifeboat creators will be presented with two possible moments to write and reflect on their responses. First, when they see an overview of permissions of their Data Lifeboat content, and second, when they see a preview of their Data Lifeboat before 'baking'.

Mandatory questions in the README[2]

A. Purpose & Compilation

Clearly defining the purpose and methods for compiling the photos in the Data Lifeboat prompts creators to reflect on their motivations and intentions.

Sample questions:

- What is the purpose of this Data Lifeboat?
- What's important to you about these photos?
- Can you explain how you assembled these particular pictures?

B. Future Access & Use

Outlining conditions or requests for future access and use of the Data Lifeboat collection, whilst this cannot be secured, can at least serve as guardrails.

Sample questions:

- Is this Data Lifeboat just for you, or do you want it to be made public one day?
    - If so, when?
- Who will you be sharing this Data Lifeboat with?

2 Specific questions are still to be refined throughout the Alpha development in Spring 2025. The current sub-questions displayed here are simply indicative of scope and priorities.

- What should or should not be done with this Data Lifeboat and its contents?

C. Storage (& Safe Harbors)

Recording (or suggesting) where the Data Lifeboat could end up prompts creators to think about future viewers or stakeholders. In future, we hope that Data Lifeboat creators could designate (or refuse) a Safe Harbor dock and its (desired) conditions for storage.

Sample questions:

- Where do you intend for this Data Lifeboat to go once downloaded?
- Is there anyone you'd like to notify about the creation of this Data Lifeboat? Is there anyone who should receive a copy?
- Where would you like this Data Lifeboat ideally to be 'docked'?

Conditional questions in the README

D. Context & Description

Providing rich, contextual information (which the free text input allows for) can help supplement existing collections with missing information, as well as helping to avoid misinterpretation or detachment from origins.

Sample questions:

- Would you like to add context or description to any image(s) in this Data Lifeboat?
- Would you like to add context or description to any comments or tags in this Data Lifeboat?
- Is there any important context about this collection that you want a future viewer of the Data Lifeboat to be aware of?

E. Ethical & Cultural Sensitivities

We have the opportunity to append ethics to historically unjust collections by giving space for Data Lifeboat creators to write how the images should be viewed, understood and should. These questions are often deprioritised in technical products of archiving, so we felt it imperative to surface them here.

Sample questions:

- Are you part of the community shown in this Data Lifeboat?
    - If not, have you considered how this might impact that community?

- Is there any sensitive information in this Data Lifeboat future viewers should know about?
- Does this material depict historical or current harms you can explain or draw attention to?
- Could bad actors misuse any of this content?
    - If so, should it be excluded?

F. Privacy & Consent

Respecting privacy and obtaining consent (where possible) are critical safeguards for the dignity and rights of the represented, particularly important for sensitive content or at-risk communities.

Sample questions:

- Could someone (living) be identified from this Data Lifeboat?
    - If so, has their consent for inclusion in this Data Lifeboat been reasonably sought?
    - If not, please explain why and any steps you have taken to keep their privacy in mind.

G. Ownership & Copyright[3]

Clear documentation of authorship and ownership (wherever possible) can protect creators' rights, or attribution at a minimum.

Sample questions:

- Do you own all the rights to the images included in this Data Lifeboat?
    - If not, can you explain why it is important they are included?
- Are there any other creators involved in these images to whom ownership should be attributed?
- Did the images in this Data Lifeboat have a different 'owner' before Flickr?

3 A record of licenses are already embedded within the Data Lifeboat. This line of questioning allows for creators to add qualitative annotation to the conditions of copyright attached to the photos. Copyright of Flickr photos primarily is intended to be machine-readable and reduces the complexity of human dynamics of ownership e.g. multiple creators, unfair copyright attribution, intended revoking date of copyright.

Exploring a viable Safe Harbor Network

> *"A Safe Harbor Network fits within our mission as a national library, it's something **we should be doing** …. It should be the responsibility of national libraries to be **responsible stewards** of individuals' and communities' social media data" — Flickr Commons member, National Librarian*

The long-term preservation of Flickr content in Data Lifeboats depends fundamentally on robust storage infrastructure and sustained maintenance. While we imagine the majority of Data Lifeboats will be created solely for personal or private usage — ultimately residing on what are calling a "beach", such as a personal computer, hard drive, or cloud storage — others will be **deliberately created to serve the broader archival mission** of stewarding our shared digital cultural heritage.

Throughout our workshops and conversations, we evaluated the viability of developing a Safe Harbor Network. Our aims were to understand what already exists in terms of archival network models, identify common challenges faced by similar initiatives, and design membership conditions grounded in these practical realities. We drew valuable insights on what this would take from archivists engaged in existing networks that operate on many scales– from multinational (Europeana, UNESCO Memory of the World), to regional (Michigan Digital Preservation Network, Digital Commonwealth Massachusetts), and cross-community (CLOCKSS).

Overall, our research revealed a **strong appetite among institutions** for a Safe Harbor Network that is specifically designed for hosting Data Lifeboats. Participants noted the conspicuous current absence of a coordinated network for social media preservation and expressed enthusiasm for a structure that could provide **practical guidance** for networked image collecting while **distributing the associated costs and risks.** Nevertheless, it became evident that any future network must thoughtfully navigate complex issues of governance, resource allocation, and institutional autonomy. These can be categorized as follows:

Opportunities for the Safe Harbor Network

- **Alignment with institutional missions:** A Safe Harbor Network aligns with the core purpose of memory institutions, operating outside the profit-driven motives of commercial platforms that currently manage the majority of social media content. While platforms operate on quarterly timelines, archives work on generational scales. A Safe Harbor Network supports these institutions in fulfilling their mandate of keeping shared cultural heritage available and accessible to the public for the long-term.

- **Advocates for importance of networked image preservation:** The very existence of a Safe Harbor Network serves as evidence of the value of networked image and social media preservation. Institutions can point to the network when justifying the value of preserving networked images to internal stakeholders (according to participants, it currently ranks quite lowly in priorities). Achieving a critical mass, Safe Harbor institutions can advocate for the value of these collections to external audiences and policymakers.
- **Distributed responsibility:** Given the risks and complexities of storing networked images (as discussed above) — factors that discourage many institutions from engaging with these materials at all — a Safe Harbor Network helps distribute these risks and responsibilities across organizations. The network can facilitate trouble-shooting through shared experiences, building a repository of institutional knowledge. Even with inevitable staff turnover over time within individual institutions, the distributed nature of the network increases the likelihood that Data Lifeboats will persist, as institutional knowledge remains preserved across the group.
- **Large archives supporting smaller ones:** Small-scale archives consistently expressed their enthusiasm for a Safe Harbor Network, particularly as they perceive their content to be the most vulnerable compared to national collections. It can also be mutually beneficial, with the network providing smaller institutions with supporting technical infrastructure (in particular storage and up-to-date policies), while simultaneously offering larger archives opportunities to expand community outreach and diversify their collections.

Challenges for the Safe Harbor Network

- **Governance concerns:** The constantly shifting landscape of digital legislation raised concerns among participants about the ultimate responsibility for content within Data Lifeboats that they are expected to store. Is liability held by individual institutions, the network as a whole, or the Flickr Foundation? Practical implementation questions arise around processes like responding to takedown requests or addressing the discovery of illegal content within a Data Lifeboat. Additionally, the network must determine how to honor the requests of Data Lifeboat creators (as recorded in the README), particularly when these wishes involve complex ethical considerations, such as excluding specific institutions due to historical injustice. Fundamental questions remain about what constitutes a "trusted institution," what conditions must be met for inclusion, how these standards are established, and what happens to the Data Lifeboats it holds if an organization violates these terms.
- **Collecting policies and priorities:** The network raises questions about how Data Lifeboats fit into existing collecting responsibilities and requirements? Participating institutions will need clarity regarding mandatory holdings (amount of data, contents of

Data Lifeboats, expectations for maintenance) before agreeing to participate. Many archivists emphasized that institutions often operate under strict collecting policies or programming priorities—for example, the National Archives of Sweden cannot hold materials from Finland—creating potential future conflicts with network participation requirements.

- **Resource limitations:** Many archives are already operating at capacity, meaning the introduction of any new initiative must be mindful of present circumstances. Small archives in particular typically lack dedicated technical staff. The labor of additional appraisal (or curation) for Data Lifeboats should also be considered. Streamlining onboarding processes, minimal participation requirements and clear troubleshooting pathways can mitigate this. Questions of economic and storage resources (and contingencies) persist and ought to be resolved before moving ahead.

Despite these challenges, the potential benefits of a Safe Harbor Network for preserving Data Lifeboats remain compelling. Looking ahead, we are committed to developing this technical and social infrastructure collaboratively with the very trusted institutions who would ultimately comprise the network.

Appendix: Definition of terms

We believe it is helpful to define the terms that are used throughout this report. Given many of these terms are used interchangeably in the field of digital cultural heritage, it is worthwhile to pinpoint specific content types and priorities in order to better design for the intended audience.

Please note: this list of terms is non-exhaustive and may change as new research surfaces.

Digital preservation: An umbrella term to describe the preservation of digital materials to ensure their long-term accessibility, authenticity, and usability. Digital materials can include documents, datasets, images, audio, software, and video games. Examples include: Europeana, South African History Archive (SAHA), Digital Library of the Caribbean (dLOC).

Web archiving: The process of capture, documentation, and storage of portions of the Internet to ensure continued access to information that might otherwise be lost. This includes web crawling, API-based archiving, and screenshot preservation. Examples include: Internet Archive's Wayback Machine, UK Web Archive, Archive-It.

Social media preservation: The systematic capture, documentation, and storage of content created on social media platforms, including posts, comments, images, videos, and associated metadata. There are not yet industry standards for social media preservation. Some institutions use web archiving tools that are designed for the broader web and narrow their focus to social networking services, but there are relatively few tools dedicated specifically to social media preservation. Examples include: Documenting the Now, Library of Congress Twitter Archive, Vox Populi Tahrir Square Archives.

Networked Images or Networked Social Photography[4]: Images that are born-digital and born-social (see below). The digital social context, typically hosted or facilitated by a third-party social media platform, is an integral part of what makes a networked image distinct from a standalone image. This definition includes photos that may have once been analog but were later digitized and uploaded. Examples include: Flickr photographs, Instagram stories, TikTok slideshows.

4 To arrive at this definition we build on the discussion outlined in Collecting Social Photography's report. This term is not to be conflated with the established fields of social photography or vernacular photography, though these may dovetail with networked social photography

Networked image preservation: The capture, documentation, and storage of networked images — see above for definition.

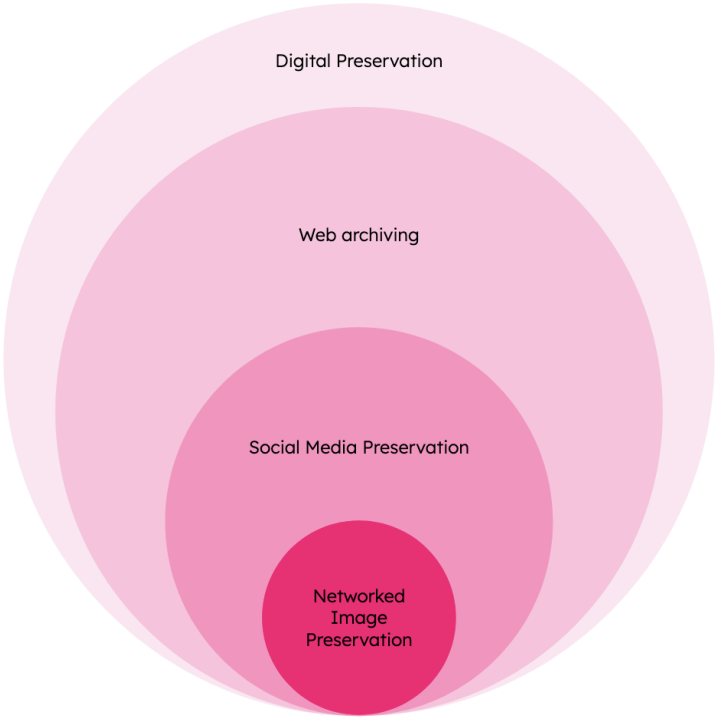

Figure 1: the nestled definitions of networked image preservation

Social context: The surrounding elements of a networked image that establish its social nature and relationships within the platform ecosystem. Examples include: comments on Flickr, hashtags on Twitter/X, playlists on YouTube.

Metadata: A set of data that is connected to and gives information about other data. The capture, composition and availability of metadata is platform-specific. Metadata includes both human-readable elements, such as the title of a photo shown on a web page, and machine-readable elements, like longitude/latitude coordinates. It is important to preserve metadata in networked image collecting because these elements are crucial to a holistic understanding of the digital object and its contextual interpretation in the long-term. However, which metadata and how much to preserve is up for debate. Examples include: comments, tags, the date a photo was taken, dimension and orientation, geolocation, device identifiers, internal content moderation flags.

Born-digital[5]**:** An artifact that originates in a digital form and cannot be reproduced by digitising a physical backup. Examples include: photographs taken on smartphones; computer-generated imagery; digital texts.

Born-social[6]**:** An artifact created predominantly for, or in the context of, social media platforms. Its conception, production, treatment and distributions may be shaped by the social relations and settings native to the platform. Examples include: Tiktok reels, 365-challenges, selfies.

5 Not everything on Flickr is born-digital but in the act of uploading it does become a unique digital object with its own properties (even if the physical original still exists). For more, see Lev Manovich on 'database identity'.

6 Not everything on Flickr is born-social either, but it does become a uniquely digital-social object by its inclusion on Flickr.com. Even if an image is set to Private, its creator is still operating within a nexus of relations on a fundamentally social platform.