

Embedding Preservability for New Forms of Scholarship

Jonathan Greenberg, Digital Scholarly Publishing Specialist,
NYU Division of Libraries

Karen Hanson, Lead Research Developer, Portico

January 27th, 2026 - DPC Event - Preserving eBooks: where are we now?

Enhancing Services to Preserve New Forms of Scholarship (2018-2020)

- Looked at complex publications already published
- Determined if they could be preservable *at scale*
- Resulted in 68 guidelines to improve preservability of publications

Funded by:

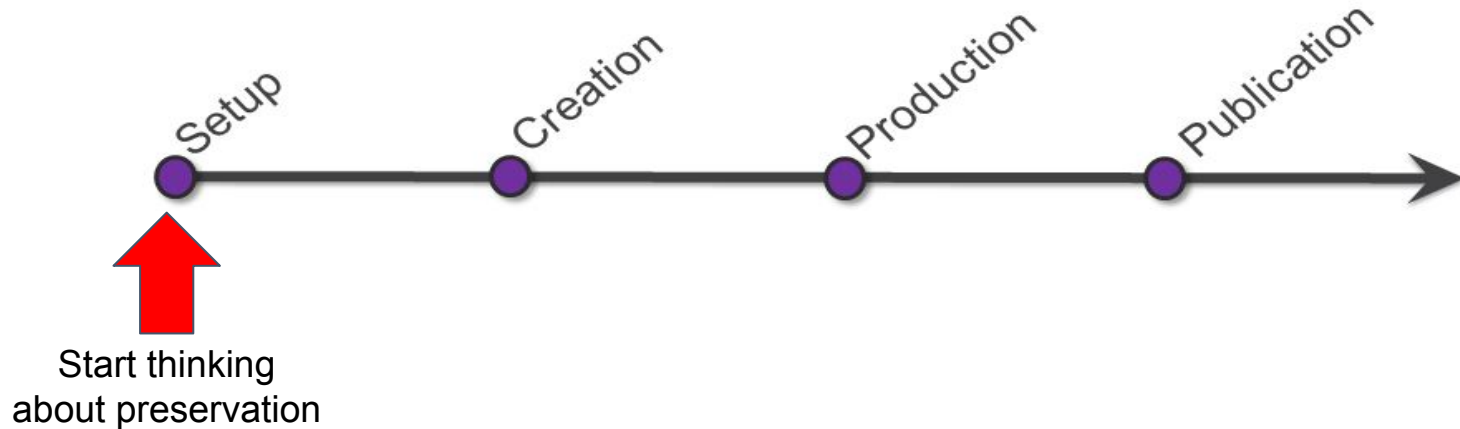


Led by:



Early decisions promote preservability

Many choices that affect preservability are made early on. Authors, publishers are most flexible at the start.



Guideline Example: Caption

16. Create a meaningful caption for all non-text features embedded in a publication.

THIRD-PARTY DEPENDENCIES

EMBEDDED RESOURCES

EPUB

WEB-BASED PUBLICATIONS

Embedded enhanced features, especially those that link to resources outside of the publication or use an unusual format, are at the highest risk of failing in the future. For this reason, a meaningful caption is vital for providing clues to future readers about what they should expect to find in that location in the text, and preferably some means of finding it and accessing it. Ideally, this caption would include a title, source, unique persistent identifier (e.g. DOI, ARK ID, or Handle), and a link to an archived copy if different from the identifier. Though any link could ultimately fail, this information would at least provide clues to where the user might find an archived copy. When creating captions, apply the standards available within the format you are using to support automated parsing. For example, use HTML5 `<figure>` and `<figcaption>` elements. Alt attributes are also widely used to supply a description in case a feature cannot be viewed. In this respect, a meaningful caption may also meet standards for digital accessibility. For a fuller treatment of this topic, see the poster presentation [Embedding Preservability: IFrames in Complex Scholarly Publications](#), presented at the IPRES 2023 conference.

Where non-text features are supplied as separate publication resources, this guideline may also be relevant:

24. [Create metadata for each publication resource](#)

▶ Example

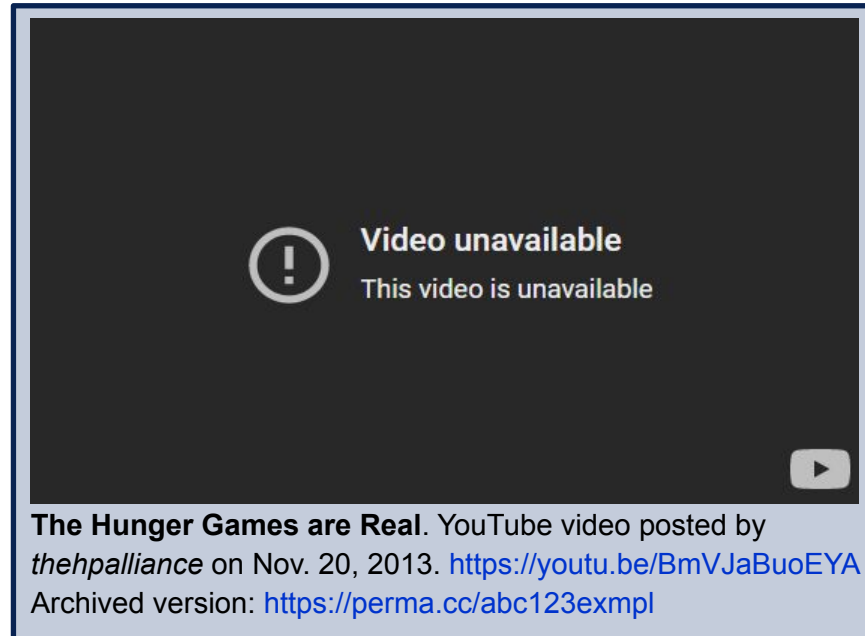
Positive example
from one of our
projects

Keywords to
help filter by
topics

References
to related
guidelines

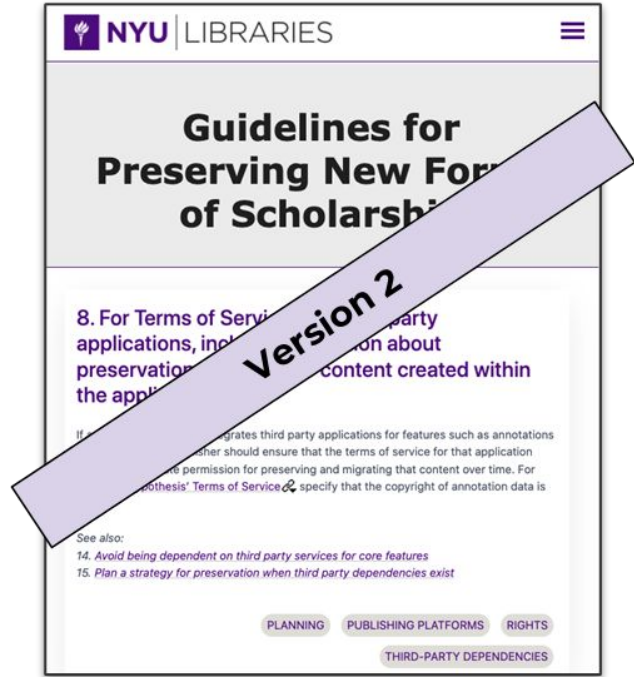
Researchers will innovate

Unrelenting reliance on third party platforms that have functionality they need (e.g. Tableau, Flourish, YouTube, SoundCloud); customize code to accomplish goals. This isn't likely to change so we need to soften the damage this can cause.



Embedding Preservability for New Forms of Scholarship (2021-2024)

- Test assumptions of the guidelines (were based on things already published)
- Embed preservation experts with publishers and their platforms
- Use guidelines to provide advice about improving preservability and track what is implemented
- Use evidence gathered from process to iterate on guidelines



Embedding process

- With publisher, identify suitable publication
- Structured, context-setting interview with publisher
- Deep dive into publication and goals for preservation
- Identify major risks
- Check in with publisher at key inflection points
- Assess results after publication

Embedding Team



LIBRARIES



LOCKSS



PORTICO

Observations: Community drives prioritization

Platform development cycles prioritize expanding and supporting user base (unsurprisingly). May need users to ask for preservation in order to make it a priority or otherwise help secure funding earmarked for preservation improvements.



Observations: Not everything needs preservation

Not everyone wants preservation, some things meant to be ephemeral. Platforms are reluctant to burden all users with preservation-related features, prefer best practice documentation.



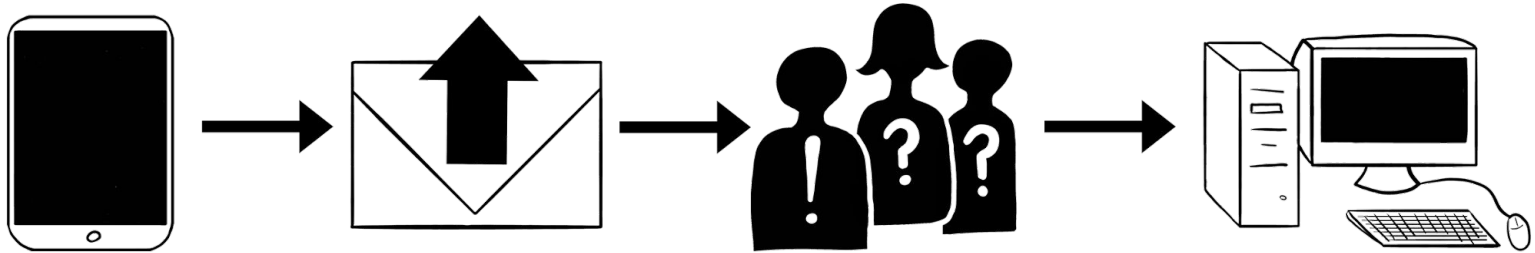
Observations:

Platform and publication are sometimes in tension

Even the most flexible, powerful publication platforms have limits on their feature set. Creative scholars and publishers are limited by these restrictions, including restrictions that might hamper preservation.

Observations: Preservability depends on processes

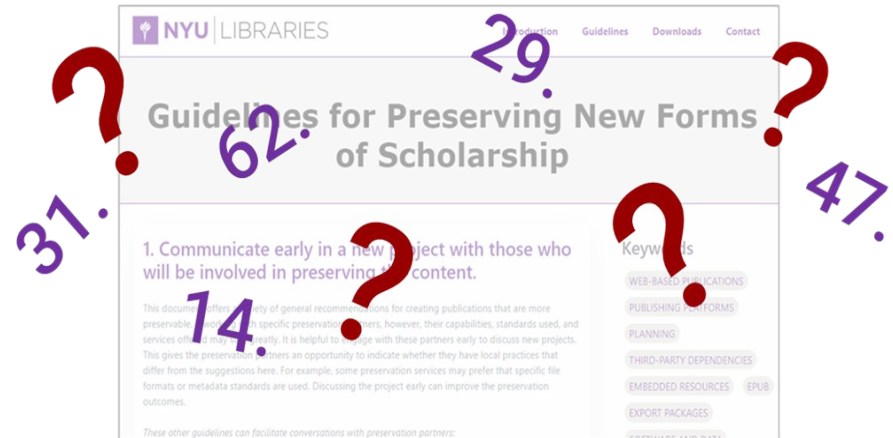
Making new forms of digital scholarship preservable requires building best practices into design and production processes.



Embedding Preservability for New Forms of Scholarship (2021-2024)

Finding: Publishers, platforms willing to take some action, BUT guidelines overwhelming without us as intermediaries.

New deliverable: Preservability Self-Assessment Tool to help assess gaps in preservation, lead you to relevant guidelines



Outcomes:
*Preservability Self Assessment
Tool for Complex Works*

Presented by Karen Hanson

Outcomes - Preservability Self-Assessment Tool

Goals

- Facilitate a conversation about preservation, emulate our embedding process
- Identify preservation risks within complex works
- Guide people to suggestions for improving preservability of their work, reduce risks

Who is it for?

- publishers, authors
- publishing platform developers who host/produce complex works
- those who *facilitate* preservation of complex works?

Four parts of the Self-Assessment

Part 1: Creates context. Collects foundational information about the publication and current preservation approach

Part 2: Defines core intellectual components that should be preserved, identifies their **Risk Category**.

Part 3: Explores each Risk Category via yes/no questions to **identify relevant Guidelines**.

Part 4: Uses relevant guidelines to **form an action plan** for improving preservability.

Self Assessment Tool Case Study: Owning My Masters (Mastered) by A.D. Carson

 UNIVERSITY OF MICHIGAN PRESS
EBOOK COLLECTION

Search and discover books 

BROWSE BOOKS HELP

Get access to more books. [Log in with your Institution.](#)

[Home](#) / [Books](#) / Owning My Masters (Mastered): The Rhetorics of Rhymes & Revolutions



Owning My Masters (Mastered): The Rhetorics of Rhymes & Revolutions
A.D. Carson

 Open Access

[Read Book](#) [Download](#)

 Some of the videos and audio tracks included in this project contain potentially offensive terminology or images. Readers should bear this in mind when accessing this project.

Citable Link
<https://doi.org/10.3998/mpub.12584348>


Published: 2024
Publisher: University of Michigan Press

[Overview](#) [Contents](#) [Accessibility Claims](#) [Stats](#)

Resources

Search and Filter Resources

« Previous | 1 - 20 of 72 | [Next](#) »




Keyword >


Creator >

Format v

audio	34
image	19
video	14
text	4
dataset	1



[Owning My Masters: The Rhetorics of Rhymes & Revolutions digital book](#)



[Intro \[Dixie Remixed, or](#)

Owning My Masters (Mastered) by A.D. Carson
<https://doi.org/10.3998/mpub.12584348>

Identifying Core Intellectual Components

What do you want to preserve?

1. Preserving the EPUB is paramount
- 2. Preserve the resource files themselves**, whether embedded in the EPUB or not. Specifically: **music tracks** and liner notes, still images, videos, playlists.
3. Preserve the resource metadata. Make sure relationships to projects are expressed.
4. We are not interested in preserving resources linked outside of the EPUB on other platforms, *except* the interactive timeline - it would be nice to have a representation of this

Identifying Core Intellectual Components / Preservation Requirements

Table 1: Populate features, what to preserve

PUBLICATION FEATURE	REQUIRED FOR PRESERVATION?	RISK CATEGORY (from TABLE 2)
EPUB of text (liner notes	REQUIRED ▾	Select Risk Category ▾
Resource files associated with the project (audio, images, playlists)	REQUIRED ▾	Select Risk Category ▾
Resource metadata including relationships to project	REQUIRED ▾	Select Risk Category ▾
Interactive timeline hosted outside of publication	PREFERRED ▾	Select Risk Category ▾
Other resource linked outside of the EPUB kept on other platforms	DO NOT PRESERVE ▾	Select Risk Category ▾

Mapping Core Intellectual Components to Risk Categories

Table 2: Risk Category Lookup

RISK CATEGORY KEY	RISK CATEGORY NAME	DESCRIPTION	WEB ARCHIVING TYPICALLY SUPPORTS
1. Basic Formatting			
A	Formatted text	Formatted, reflowable text e.g. HTML, XHTML, or other markup/markdown. This maybe on a webpage or in an EPUB. It could be supported by stylesheets and/or custom or rare fonts for correct presentation. This does not include JavaScript	Yes
B	Linked External Resources	Hyperlinks to third party resources (not embedded) e.g. articles, datasets	Yes
C	Publication as EPUB	Main text can be downloaded or exported as an EPUB	Yes
2. Locally Hosted Supplements			
D	Locally hosted text, image, audio, or video supplement	Supplemental text (e.g. PDF), image, audio or video as downloadable file(s). The files are hosted on the publication platform.	Yes
E	Locally hosted data/code supplement	Supplemental dataset or code as downloadable file(s). The files are hosted on the publication platform.	Yes

Identifying Relevant Guidelines

Table 3: Look up the risks to view related questions

RISK CATEGORY	PRESERVATION APPROACH	RISK CATEGORY QUESTIONS	RESPONSE	IF "NO" READ GUIDELINE(S) OR FOLLOW
D	Locally hosted text, image, audio, or video supplement			
D	ALL	Are you using non-proprietary, open formats for these supplemental files?	N/A	Guideline 11
D	ALL	Have you discussed delivery and preservation of these supplements with the author?	N/A	Guideline 12 Guideline 13
D	EXPORT	Are these supplements included in the export package?	N/A	Guideline 20
D	ALL	Does the platform support the creation of unique persistent identifiers for each resource?	N/A	Guideline 27
D	ALL	Does the platform support structured, descriptive metadata for each resource in a publication? <i>If yes, consider completing Risk Category (0)</i>	N/A	Guideline 24
D	WEB ARCHIVING	Does your web publication have a sitemap.xml that references all supplements that are a part of the publication? ***	N/A	Guideline 43

Other Notes

- Facilitated workshops were helpful for people to work through the tool
- Not necessary to do all stages every time
- Something like this is never complete - we've proposed a framework, time will tell if it's a useful one to build on!

Outcomes: *Internal to Portico*

Challenge as a preservation service for e-books

- Portico workflow **optimized for traditional e-books**
- These are **mostly uniform**, follow standards, and still the overwhelming majority
- The uniformity allows us to take advantage of **economies of scale**.
- **Access** is important for our service.

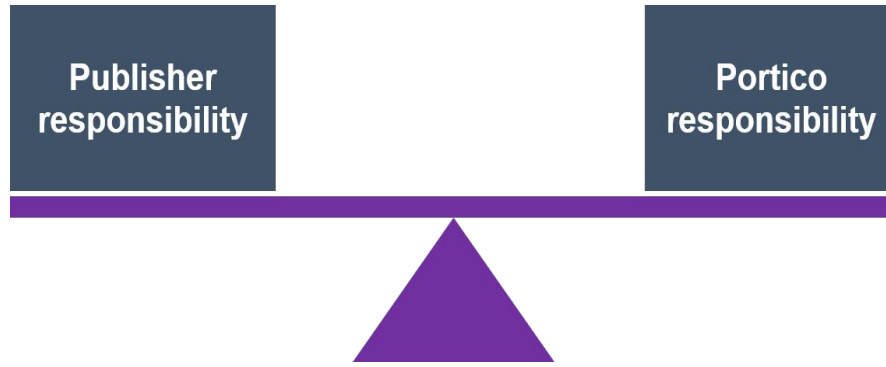
Can we still preserve these *new forms of scholarship* at scale and support access?

It depends...!



Depends on...?

- **Degree of consistency.** E-books on publishing software build to host multiple publications are most amenable.
- Scalable preservation of complex works best with **effort on both sides.**



Portico enhancements to support new forms

Project identified **enhancements**, some already implemented:

- EPUB validation
- Web archive tool
- File format tool upgrades
- **“e-projects”**

Fulcrum – Resources with unique identifiers

Resources

Search and Filter Resources


Keyword >

Creator >


Format ▾

- audio 34
- image 19
- video 14
- text 4
- dataset 1

« Previous | 1 - 20 of 72 | Next »



[Owning My Masters Revolutions digital book](#)



[Intro \[Dixie Remixed\]](#)

The Song I Should Have Written

From [Owning My Masters \(Mastered\): The Rhetorics of Rhymes & Revolutions](#)
A.D. Carson

0:00 / 4:46 Speed: 1x Paused

Auto scroll

[A.D. Carson plays "The Song I Should Have Written"]

[applause]

[r] (James Baldwin) I'm going to improvise

[r] like a writer, on some assumptions.

Citable Link

<https://doi.org/10.3998/mpub.12584348.cmp.13>

Download WAV (48.5 MB) Share

Creator Role: author

Subjects: American Studies, African American Studies, Music:Musicology

Keywords: Track 10, Saas-Fee, Song I Should Have Written, Preme, James Baldwin, summer session

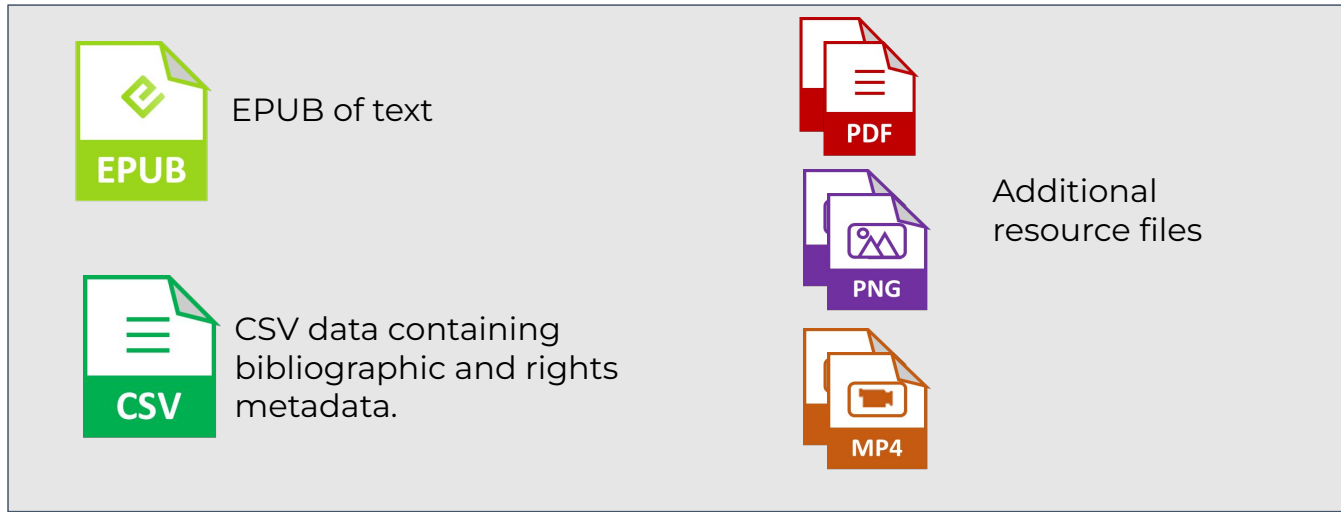
Language: English

Citable Link: <https://doi.org/10.3998/mpub.12584348.cmp.13>

A. D. Carson. (2024) *The Song I Should Have Written*.
<https://doi.org/10.3998/mpub.12584348.cmp.13>

Fulcrum example - Export

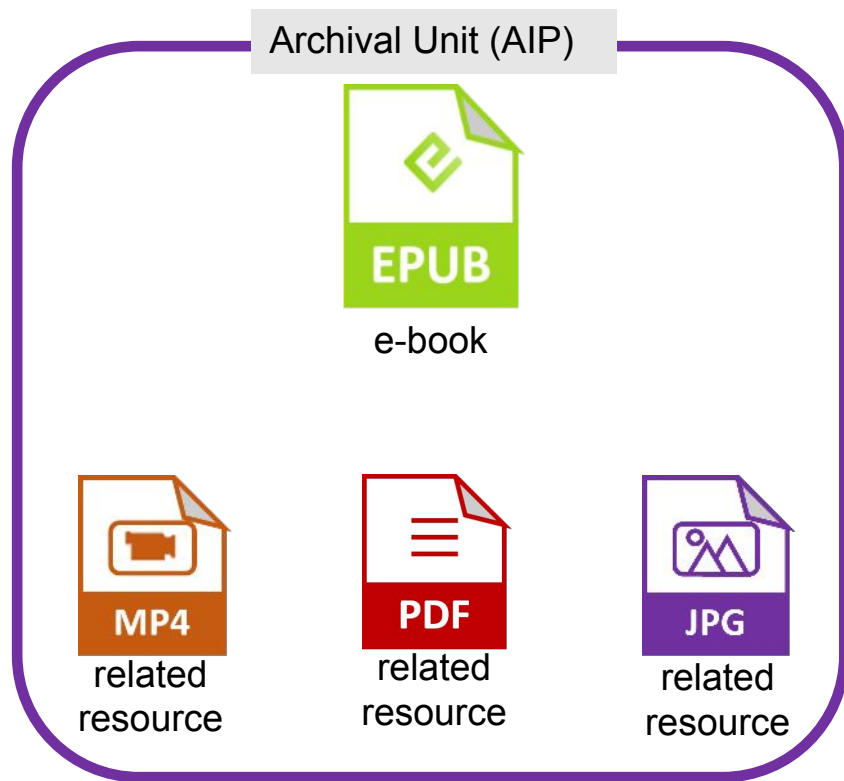
Fulcrum uses an **export** process to create packages for preservation. It is a zip-file (a Bag-It bag) containing the following data:



Before: 1 publication = 1 archival unit

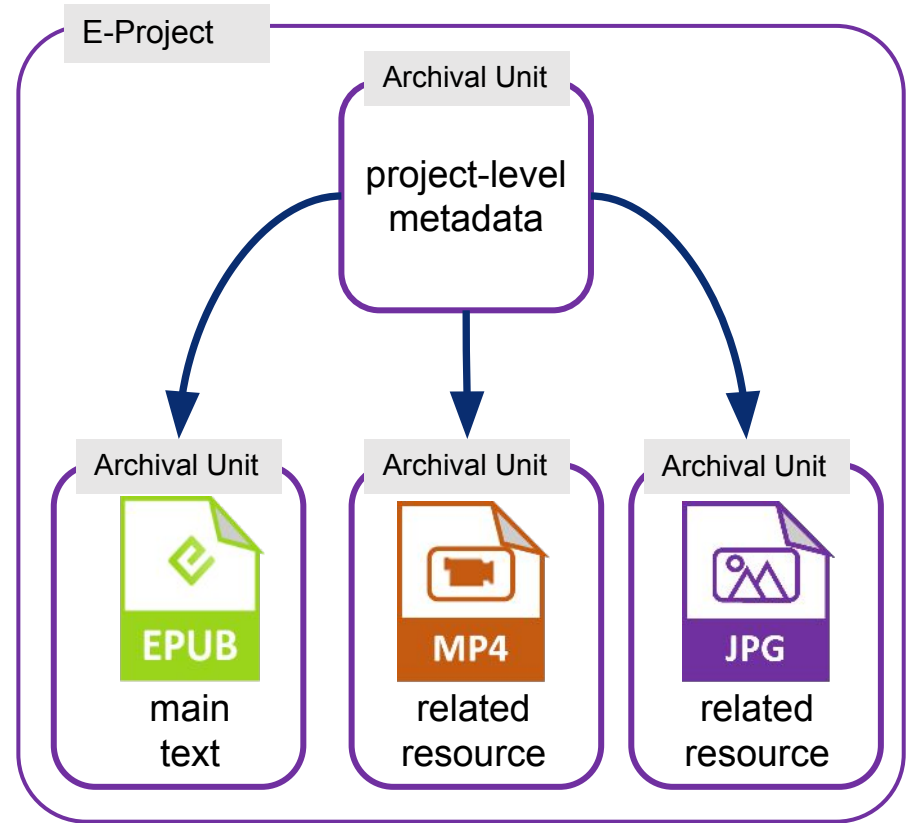
Resources treated as supplements, which meant:

- Descriptive metadata at book level, resource-level metadata not managed
- DOI resolves to book level only
- If one file changes, it affects the whole package



E-projects: Interlinked Archival Units

- An *e-project* can be composed of many linked resources
- Individual resources are discoverable, resolvable (can map DOI)
- Smaller packages, easier to manage, version
- More closely reflects treatment on original platform
 - items elevated to first class objects



Closing

- Research incredibly useful for understanding what role publishers/platforms can play in preservation
- It emphasized the importance of our relationship with the publishers and platforms in preserving complex works
- We have proposed a framework for this through our research, but would love feedback on this

Questions

jonathan.greenberg@nyu.edu

karen.hanson@ithaka.org

Project deliverables: <https://archive.nyu.edu/handle/2451/63332>

Guidelines website: <https://preservingnewforms.dlib.nyu.edu>