Enhancing Services to Preserve New Forms of Scholarship

The Charleston Conference, November 8, 2019

Jonathan Greenberg, NYU Libraries & NYU Press
Charles Watkinson, Michigan Publishing
Karen Hanson, Portico
Susan Doerr, University of Minnesota Press
Project Goals

Led by NYU Libraries and funded by the Andrew W. Mellon Foundation, the project is a collaboration of libraries, preservation organizations, and university presses with 2 goals:

**Goal 1:** Given samples of the works the preservation service providers will identify what aspects can be preserved, and how scalable/resource-intensive the service would be.

**Goal 2:** Based on results, and on feedback from publishers and experts in preservation, copyright, and scholarly communications, develop guidelines and best practices that publishers and authors can use to improve preservability of the monographs.
New Forms of Scholarship

... push the boundaries of how scholars communicate and publishers publish.

... include **dynamic features** that cannot be represented in print such as:

- embedded audio / video / high-resolution images
- non-linear navigation
- data visualizations
- interactive maps
- supporting software, code, data
- annotations and highlights

... are created **using flexible formats** such as EPUB3 and HTML5. Each depend on **web languages**.

... are a **preservation challenge**!
Workflow

Select publication within format being analyzed.

Preservation partners work with publisher to define acceptable preservation.

Publication transferred to, or collected by, preservation partner(s).

Design preservation workflow, prepare sample rendition if relevant.

Publisher evaluates preservation plan against agreed criteria.

Document issues, create guidelines.

Repeat until good coverage of preservation challenges.
Open Square

• In-house platform for OA books

• Enhanced Ebooks (in collaboration with NYU Digital Library Technology Services)

• Readium browser-based EPUB reader

• New features and design for a rebuilt Readium reader to come
Enhanced Ebooks from NYU Press

BY ANY MEDIA NECESSARY
THE NEW YOUTH ACTIVISM
HENRY JENKINS, SANGITA SHRESTHAVA, LIANA GAMBER-THOMPSON, NETA KUGLER-VILENCHIK, and ARELY M. ZIMMERMAN

SHOW SOLD SEPARATELY
PROMOS, SPOILERS, AND OTHER MEDIA PARATEXTS
JONATHAN GRAY

NYU LIBRARIES
Fulcrum is an open source publishing platform for long-form digital scholarship. It responds particularly to the increasing need humanists have to integrate their “data” into their “narratives.”

Four design principles

1. Flexible
2. Durable
3. Accessible
4. Discoverable
**Enhanced Ebook**

**Reading layer:**
- EPUB3 opened in EPUB.js reader
- Embedded AblePlayer renders video resource

**Repository layer:**
- Resource managed in Hyrax repository
- Comprehensive metadata and accessibility
- Preservation pathways for EPUBs/PDFs (to CLOCKSS) and all resources (to APTrust)
“Interactive scholarly work”

- EPUB3 in EPUB.js
- Unity 3D Model in WebGL
- Media resources managed in Hyrax
- Link out to SQLite database
- Deep two-way linking

Stewardship commitment from U-M Library. Components can be preserved, but preserving interactivity between them is challenging.
What is Portico?

• Community-supported digital preservation archive
• Preserves scholarly publications
• If content no longer hosted online, it becomes available through Portico
Some EPUB3 tidbits

remote resource

visually embedded
E-books in Portico

If valid, keep accessible

Ensure bytes don’t change
New EPUB JHOVE module

Well formed and valid
EPUB rendition
Examples of EPUB challenges
Challenge 1: Embedded YouTube videos

The video starts with a quote from Haymitch, who advises Katniss and Peeta, contestants in the fictional Hunger Games: "From now on, your job is to be a distraction, so people forget what the real problems are." The video then shows some of the marketing efforts for Catching Fire, highlighting the "hot guys" in the cast, only to interrupt the scene with static and a disruptive black-and-white "rebel message." Lauren Bird, the spokesperson for the HPA, dressed in an austere black outfit like that of a rebel leader in the film, proclaims:

Enough with the distraction, the Hunger Games are real.
Check it out: In The Hunger Games a small portion of the population controls a majority of the wealth. People have full time jobs and still go hungry.
Think it's fiction? Think again.

While the audio details the economic disparities in Hunger Games' dystopian country of Panem, the visual shows statistics of real-world economic inequality. For example, the death of Katniss' father in a mining accident is linked to the decline of unionization in the United States; the corrupt justice system in Panem is connected to the fact that African Americans are incarcerated at six times the rate of whites in the U.S. Bird concludes:

Does the Hunger Games have hot guys in it? Big fucking deal. It also has something else. Us. People who want justice. President Snow of the world, your reign is coming.
Approach 1: Keep the EPUB as-is

The video starts with a quote from Haymitch, who advises Katniss and Peeta, contestants in the fictional Hunger Games: “From now on, your job is to be a distraction, so people forget what the real problems are.” The video then shows some of the marketing efforts for Catching Fire, highlighting the “hot guys” in the cast, only to interrupt the scene with static and a disruptive black-and-white “rebel message.” Lauren Bird, the spokesperson for the HPA, dressed in an austere black outfit like that of a rebel leader in the film, proclaims:

Enough with the distraction, the Hunger Games are real.

Check it out: In The Hunger Games a small portion of the population controls a majority of the wealth. People have full time jobs and still go hungry.

Think it’s fiction? Think again.

While the audio details the economic disparities in Hunger Games’ dystopian country of Panem, the visual shows statistics of real-world economic inequality. For example, the death of Katniss’ father in a mining accident is linked to the decline of unionization in the United States; the corrupt justice system in Panem is connected to the fact that African Americans are incarcerated at six times the rate of whites in the U.S. Bird concludes:

Does the Hunger Games have hot guys in it? Big fucking deal. It also has something else. Us. People who want justice. President Snow... of the world, your reign is coming.
Approach 2: Automatically save video and link

Video not working? Find an archived copy in the Wayback Machine
Copy of video in the Wayback Machine
Problems with *Portico* taking this approach

- Video may already be gone
- Not always successful, ideally there would be manual QC
- Prefer not to modify content we receive

*Ideally* this would take place during production of the e-book

*Video not working? Find an archived copy in the Wayback Machine*
Challenge 2: Embedded multimedia viewers
Challenge 2: Embedded multimedia widgets

The Extra Texts, Bonus Texts, and Ideal Texts of DVDs

If paratexts can brand and recode reality, fictional universes prove an even easier target for branding and recoding. And while fictional films and television shows frequently boast many of the same types of paratexts that makeover shows have, a particularly strong paratext has been the DVD, complete with them, exclusively for the audience. The DVD can turn that paratext into a multimedia experience. Thus, for a marketing strategy, the producers of the show like The Biggest Loser use the DVD package, the so-called "Blu-ray" edition, to write that the show's producers have "an exclusive sneak preview of the season's most exciting new content, meaning a whole new level of entertainment..."

The media-industry frenzy surrounding the DVD release has led to the creation of a new kind of "bonus material" that is linked to the show in some way. The bonus material is presented in a somewhat different form, as it is not meant to be a complete re-creation of the show. While the bonus material is not necessarily intended for audiences that are not involved in the show, it is presented in a way that is very specific to the audience.

Moreover, the bonus material is presented in a way that is very specific to the audience. The DVD is intended for a specific audience, and the bonus material is created to engage that audience in a particular way. The bonus material is designed to appeal to the audience, and it is presented in a way that is very specific to the audience. The bonus material is presented in a way that is very specific to the audience, and it is presented in a way that is very specific to the audience.

Fig. 3.1: A page from The Biggest Loser's website.
Fulcrum package received by Portico

• The EPUB3 (with remote multimedia viewers)
• Multimedia files, separate from EPUB
• Metadata, including mapping between file names and fulcrum.org web page displayed in EPUB

<table>
<thead>
<tr>
<th>File Name</th>
<th>Link</th>
<th>Embed Code</th>
<th>Title</th>
<th>Resource Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_4.mp4</td>
<td><a href="https://www.fulcrum.o">https://www.fulcrum.o</a>&lt;div style='width:auto;</td>
<td><em>Neverquest</em> adv video</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1_5.mp4</td>
<td><a href="https://www.fulcrum.o">https://www.fulcrum.o</a>&lt;div style='width:auto;</td>
<td>_Super Happy Fun video</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2_1.jpg</td>
<td><a href="https://www.fulcrum.o">https://www.fulcrum.o</a>&lt;div style='width:auto;</td>
<td><em>Walking Tall</em> adv image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2_2.jpg</td>
<td><a href="https://www.fulcrum.o">https://www.fulcrum.o</a>&lt;div style='width:auto;</td>
<td>Baby Mama adv image</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Approach 1: Leave EPUB as is

Not surprisingly, then, *The Blair Witch Project* has arguably remained as famous (if not more so) for its creative and masterful promotion as for the film itself. Since in many ways, the horror began online and in front of the television, not simply in the movie theater.

Can't reach this page

- Make sure the web address https://www.fulcrum.org is correct
- Search for this site on Bing
- Refresh the page

Fig. 2.9. *The Blair Witch Project* poster.
Leaving as-is would require user assembly

https://www.flickr.com/photos/mr_t_in_dc/1547501637
Approach 2: Create a large self-contained EPUB

Move remote resources inside EPUB
Move the remote resources inside
Move the remote resources inside
Approach 3: It’s behaving like a website, preserve it as a website?

- Would preserve Fulcrum look and feel, including accessibility features

BUT

- It’s not an EPUB any more!
- Playback may be more challenging

Harvest website (web archive file)
All 3 approaches combined?

Original + supplements

Self-contained “EPUB/A”??

Web archive of book on fulcrum.org
Introducing Manifold

Manifold is an intuitive, collaborative, open source web platform for scholarly publishing. With iterative texts, powerful annotation tools, rich media support, and robust community dialogue, Manifold seeks to transform scholarly publications into living digital works.
preservation Format:

• EPUB3

ePub is a standard and there is interest in building preservation specifications into the future versions.

ePub can contain most of/all of the elements of a Manifold project.

Groups that are unable to produce ePub would be able to create them with Manifold (and potentially sell them).

This is potentially a mechanism to serve a complete Manifold edition for an offline reading experience.
REUSE Your Metadata!

- Metadata created for Accessibility is useful for Preservation:
  - Long descriptions
  - Keywords
  - alt-text
  - abstracts
But wAit! There’s MOre!

What about projects that do not export neatly into ePub?

_Cut/Copy/Paste: Fragments of History_
by Whitney Treitten

https://manifold.umn.edu/projects/cut-copy-paste
Website preservation approaches
Standard website harvester

• Good for simpler websites.
• May struggle with sites that interact frequently with the server based on user interactions
anamorphic distortion. From a randomly generated map with multiple levels of digital distortion, the puzzle may have been made but it is not for us. It is picture planes all the way down. To see 
Memento Mortem Mortis, go to http://manifold.umn.edu/mo
Something in-between?

- Webrecorder **Autopilot** (Rhizome)
- Memento Tracer (Los Alamos National Lab)
Emulation as a Service Infrastructure (EaaS)I

https://www.softwarepreservationnetwork.org/eaasi/
Emulation as a Service Infrastructure (EaaSI)

Live site @ manifold.umn.edu

Emulated Manifold using EaaSI

e-book data
EaaS for EPUBs

Like all people with a common occupation, western Lake Erie fishermen recognize their fellow workers by a shared knowledge of work technique. The process of becoming a fisherman is the process of learning the fisherman’s work through day-to-day, trial-and-error experience. It is accomplished by following orders, by practicing actions demonstrated by others, by referring to accepted ways of doing everyday work as they have been explained by more experienced fishermen, and by improvising solutions to new problems if traditional ways fail short. While this common knowledge is necessary for successful work to be done, it goes beyond matters of practical fishing procedure. It also has an evocative and aesthetic side, through which the quality of work can be judged and the pleasure in cooperation and good work can be expressed. Bob Bozda and Martin Hoaden, who worked together for many years out of Toledo, recall their subtle cooperation.

BB: We got to know each other so good that Martin didn’t have to tell me what he was going to do. I knew what he was going to do, and he knew what I was going to do.
MH: I make a little motion, and he knew what that motion meant. Just a little nod with the head, and he knew exactly what to do. Thank you, Bob.

In this chapter, we describe the work processes of the three types of fishing done in western Lake Erie, based on our observation of and our interviews with fishermen, and then examine the ways fishermen describe their knowledge of work techniques and customs as part of their occupational culture and identity.
Thank You & Questions