

# *Building on Hyrax and Avalon for the*



OF PUBLIC BROADCASTING

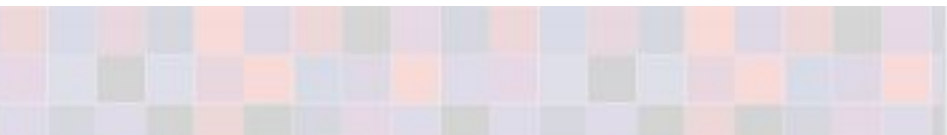
Samvera Connect 2018  
October 11, 2018

---

**Casey Davis Kaufman**, *WGBH Media Library and Archives Associate Director / AAPB Project Manager*  
**Sadie Roosa**, *WGBH Media Library and Archives Technical Project Manager / AMS 2.0 Product Owner*  
**Andrew Myers**, *WGBH Media Library and Archives Supervising Developer*  
**Jason Corum**, *WGBH Media Library and Archives Web Developer*

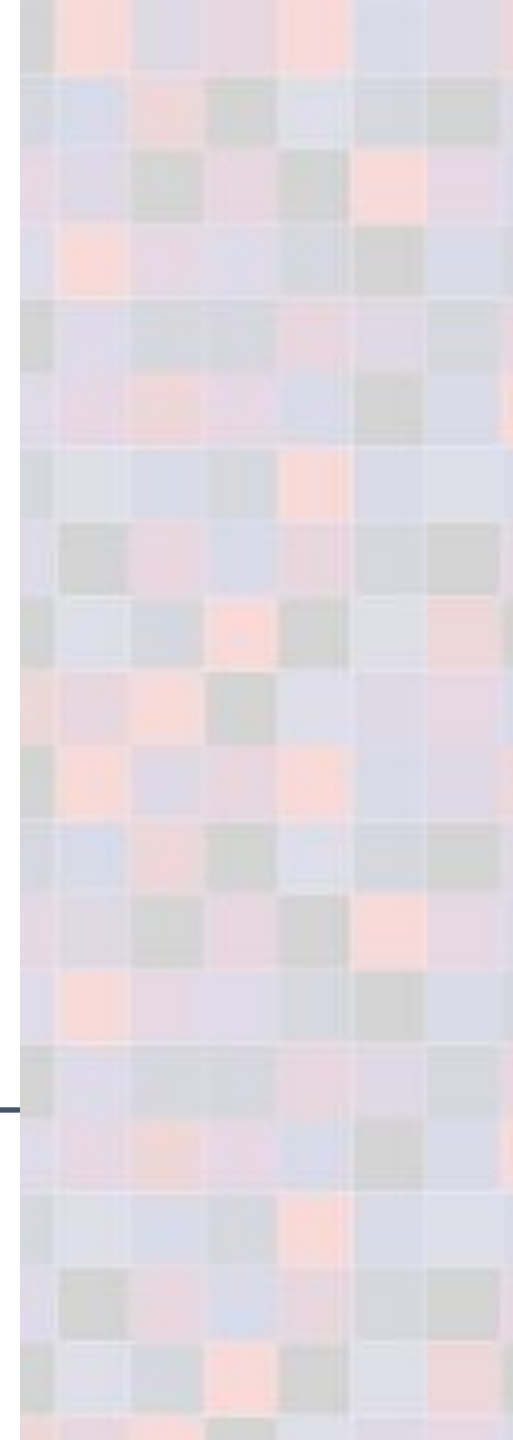
# *Outline*

- AAPB Background
- Plans for the Repository
- PBCore Metadata
- Advantages of Hyrax
- Challenges so far
- The Future



# *AAPB BACKGROUND*

---



# American Archive

---

OF PUBLIC BROADCASTING

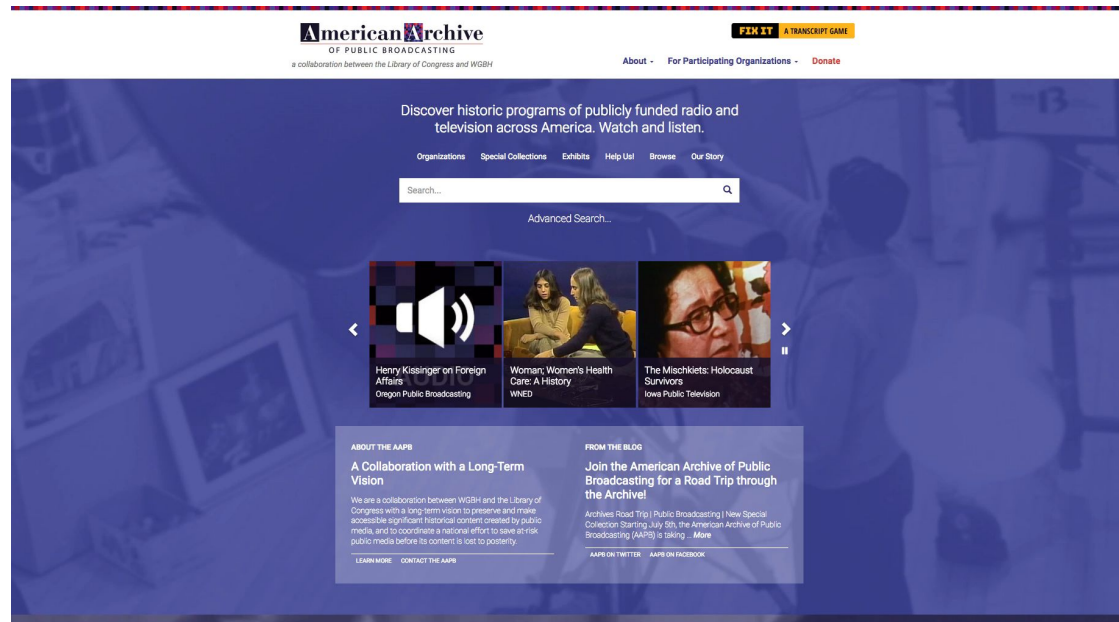
a collaboration between

WGBH and the Library of Congress

seeking to preserve and make accessible significant historical  
content created by public media, and to coordinate a  
national effort to save at-risk public media before its  
content is lost to posterity

# *The AAPB Collection*

[www.americanarchive.org](http://www.americanarchive.org)



More than **50,000 hours of digitized** and born digital material from over 100 public broadcasting stations and organizations

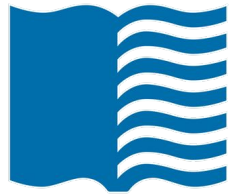
Website launched October 2015

**>35,000 streaming video and audio** files in an Online Reading Room (39% of full collection)

**Public access** to the full collection of video and audio **on-site** at WGBH and the Library of Congress

**>2.5 million inventory records from 120 stations**, in addition to the digitized items, are available for research

# *AAPB Collaboration*



LIBRARY OF  
CONGRESS

Preservation



Access & Outreach

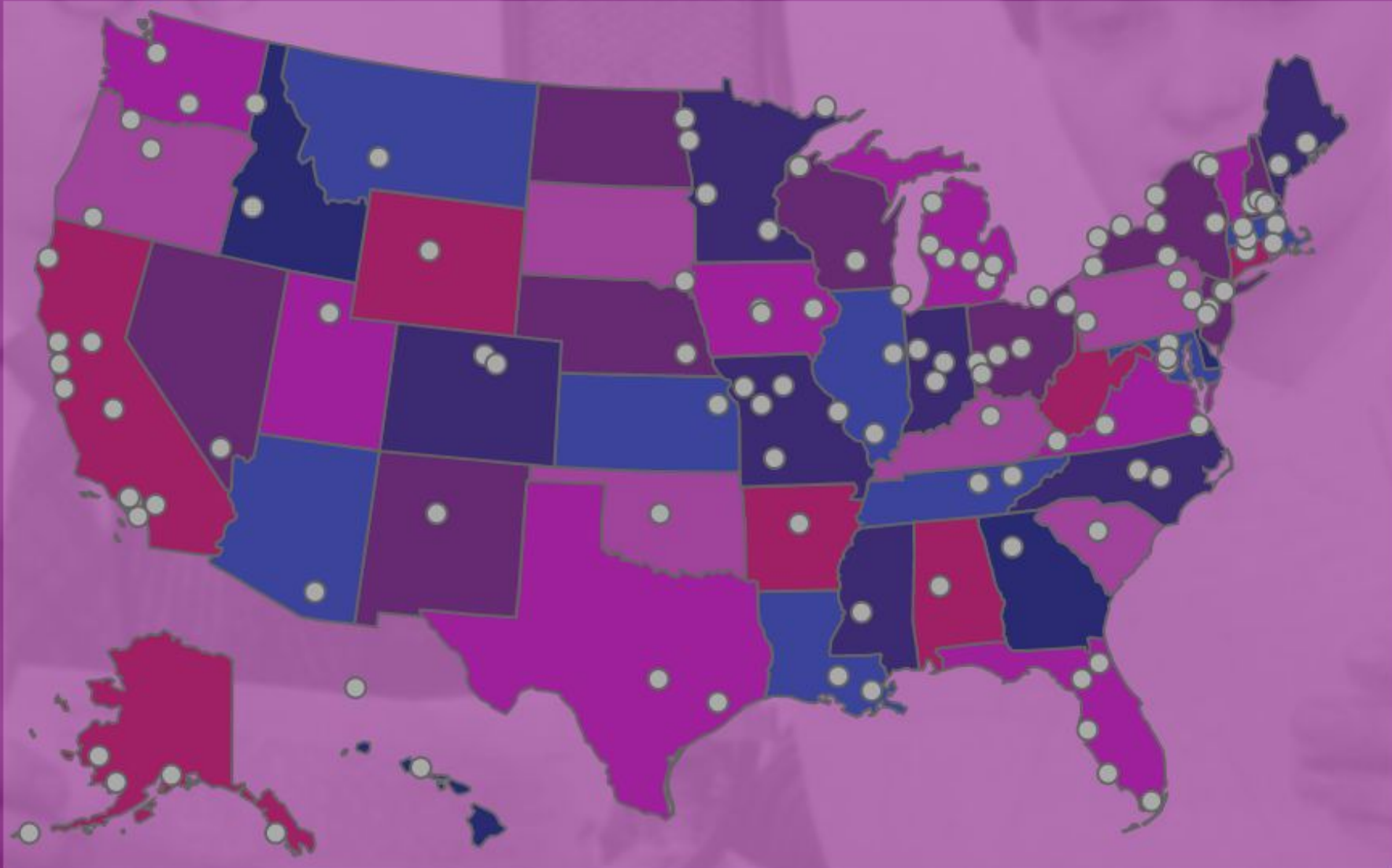


## Shared Responsibilities

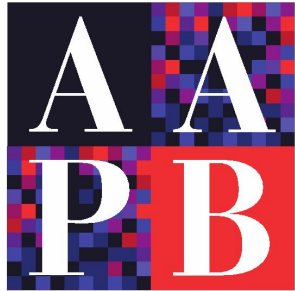
- Overall governance
- Policy
- Collection development
- Ingest
- Rights decisions

# Participating Organizations

Learn about the history, significant productions, and archive contributions of more than 120 public media stations and archives. [View all participating organizations.](#)







## *Goals*

Coordinate a **national effort** to preserve and make accessible as much significant public broadcasting materials as possible

Become a focal point for **discoverability**

**Provide standards and best practices** for storing, processing, preserving, and making accessible historical content

**Facilitate the use of archival content** by scholars, educators, students, journalists, media producers, researchers, and the public

**Increase public awareness** of the significance of historical public media and the need to preserve it and make it accessible



# *Commitment to Growth*

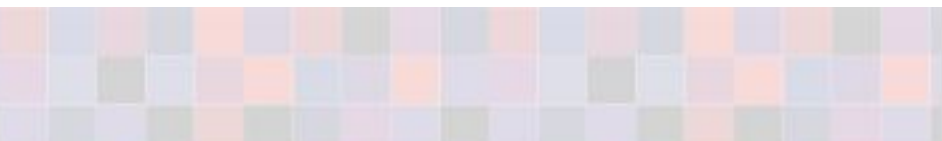
The Library and WGBH are committed to **growing the AAPB collection by up to 25,000 hours** of digitized, or “born digital” content per year

This year, we have targeted outreach to stations in states, regions and communities currently **underrepresented in the AAPB**.

We are currently lacking any content in the archive from 12 states and the territories (excluding Guam)

We are **providing grant writing assistance** to organizations submitting applications for digitization grant programs

We have created a **grant proposal document package** for organizations that want to collaborate with us on proposals

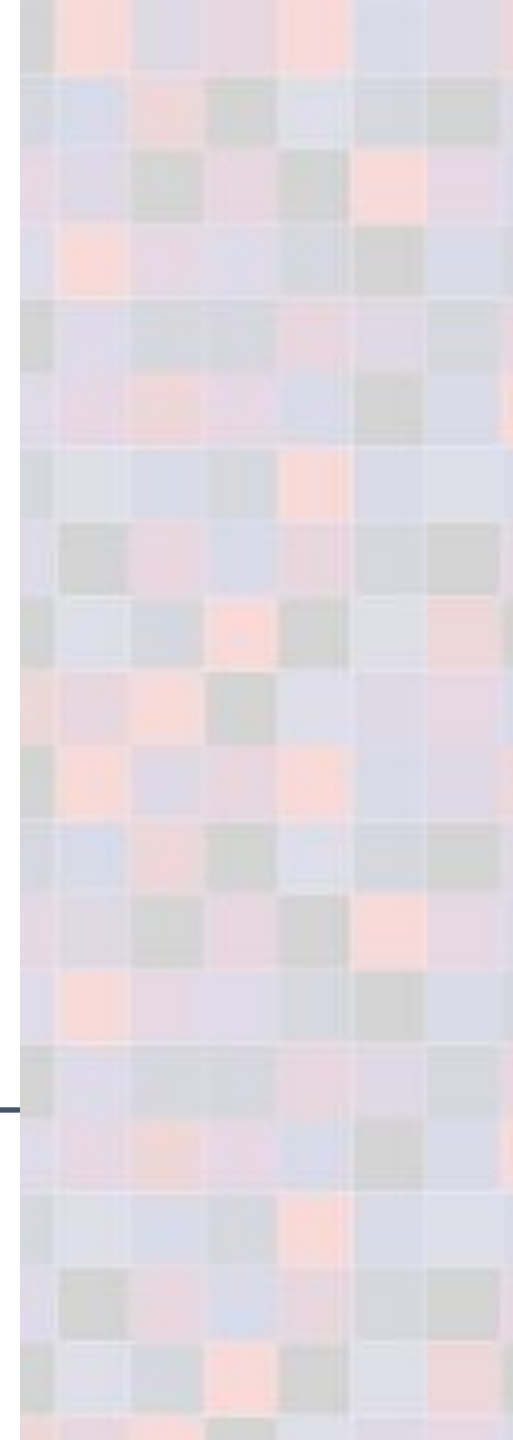


# *The Value of AAPB Participation*

- Copies preserved long-term at the Library of Congress
- AAPB manages the access platform and media servers
- Their collection becomes part of a national initiative and has broader reach
- AAPB provides guidance, fundraising assistance, and digitization project management support
- Materials donated are included in AAPB initiatives such as transcript creation, crowdsourcing, automated metadata creation
- Organizations have access to the AAPB's Archival Management System (metadata repository) where they can search, manage, update, and access their records and media

*AAPB  
REPOSITORY  
PLANS*

---



# Archival Management System

Participating organizations have access to the Archival Management System (AMS) where station administrators can:

- Search and access their metadata records
- Export their records
- Import additional records
- Edit their records
- Watch/listen to their low-res proxy files
- Download their low-res proxy files

AMS

DashboardRecordsStationsReports

Program : Hazel Dickens: It's Hard to Tell the Singer from the Song

Raw Footage: An Interview with Hazel Dickens

Return<< PreviousNext >>

Asset Information

INSTANTIATION

Original

00000793 (Appalshop Barcode)

Physical Format: 3/4 inch videotape: U-matic

Projected Duration: 00:20:00

INSTANTIATION

Mezzanine

cpb-aacip-138-29b5ms90.mpeg2.mxf (mediainfo)

Digital Format: application/mxf

Actual Duration: 00:22:20

INSTANTIATION

Proxy

cpb-aacip-138-29b5ms90.h264.mov (mediainfo)

Digital Format: video/mp4

Actual Duration: 00:22:19

INSTANTIATION

Preservation Master

cpb-aacip-138-29b5ms90.j2k.mxf (mediainfo)

Digital Format: application/mxf

Actual Duration: 00:22:19

flowplayer

00:0122:20

Open Proxy file

Edit AssetAdd Instantiation

\* Organization:

Appalshop, Inc. (WMMT and Appalshop Films)

Asset Type:

Raw Footage

\* Title(s):

Program : Hazel Dickens: It's Hard to Tell the Singer from the Song  
Raw Footage: An Interview with Hazel Dickens

\* Description:

Raw Footage Description: This portion of an interview with bluegrass singer and songwriter Hazel Dickens focuses on problems with an aging population, traveling and playing music overseas, and the differences between traditional and modern music.

Genres:

Music (AAPB Topical Genre)  
Social Issues  
Travel

Contributor:

Dickens, Hazel, Interviewee

Rights:

No copyright statement in content

Audience Level:

Audience Rating:

# *Unique Features of AMS*



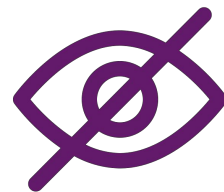
Does not manage files



Hierarchical PBCore  
data model



A lot of batch operations

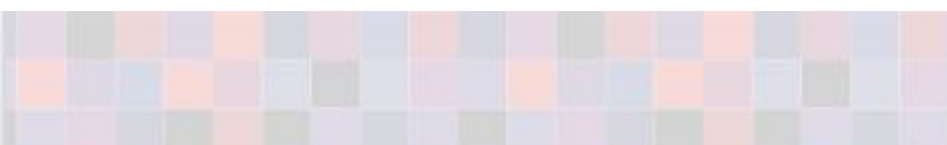


Not for general access  
purposes

# *How to move forward?*

Compared 3 options against our updated system requirements/desires

1. Extending AMS 1.0
2. Building a new system from scratch (pretty much immediately discarded this one)
3. Building something Samvera-based
  - Avalon
  - Hyrax



# ***Change 1: Update PBCore data model***

## **Current AMS**

Centers on the Contributing Organization

Built based on physical inventories

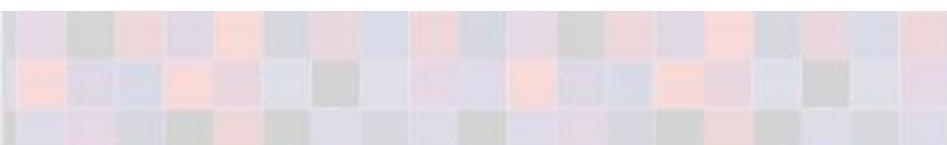
Designed primarily to manage digitization projects

## **Updated AMS**

Centers on the Asset

Only Instantiations "belong" to an organization

Multiple organizations can add child Instantiations to a central Asset





## Current AMS

Uses MINT 1 plugin for data mapping from XML and CSVs

MINT 1 is no longer supported

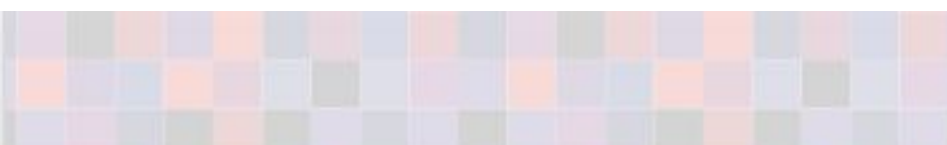
MINT only works in 1 browser right now

## Updated AMS

Samvera works to actively maintain core components

WGBH is part of Samvera and can assist with continued maintenance

***Change 2:***  
*Use actively  
maintained  
plug-ins*



***Change 3:***  
*Build a system  
we can continue  
to develop as  
needs change*

**Current AMS**

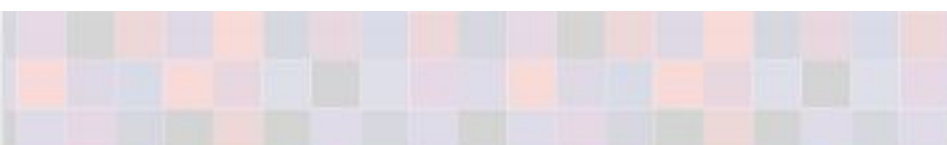
Built at the beginning of a project  
that has changed a lot

Built completely by outside  
contractors in PHP

**Updated AMS**

Built with a better idea of a more  
stable project's needs

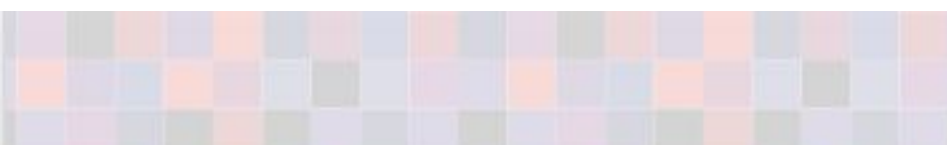
Built in a robust community in which  
our developers have expertise



## *Initial Plan*

Building on Avalon gave us the most to start working with, to let us focus on fine tuning the AAPB specific features.

The Andrew W. Mellon Foundation awarded AAPB a grant in part to carry out this work, with WGBH and AVP developers and Indiana University Avalon team advisors.

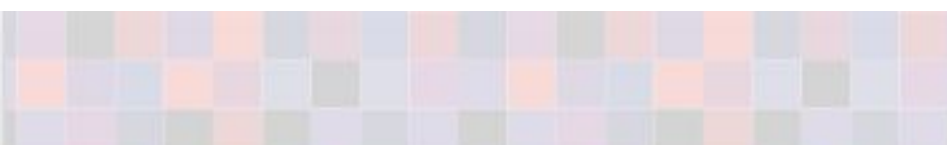


# *Hiccups*

By the time the AMS development project launched, Avalon had decided to move to Hyrax for Avalon 7.

Avalon would be moving to Hyrax during the same time period that AMS would be being developed.

We couldn't wait for Avalon to finish moving to Hyrax before kicking off development because of grant deadlines and project needs.



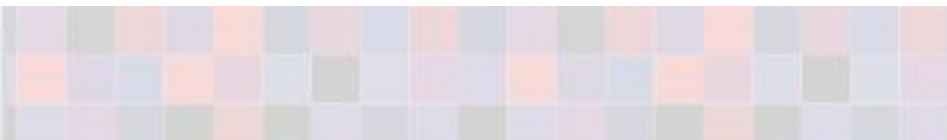
## *New Plan*

Build AMS on Hyrax

Build new AAPB specific features, some baked into the application and some as components or flip-flop features.

Use Avalon components where we can, like the Avalon Media Player.

Build components with the Avalon team to address common needs, like batch import.



*AMS 2.0*  
*Core Team*

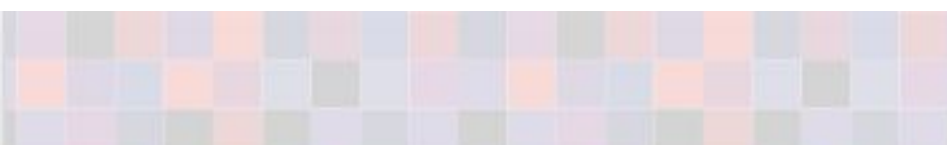
**Sadie Roosa** (WGBH) - Product Owner

**Adeel Ahmad** (AVP) – Developer

**Andrew Myers** (WGBH) – Developer

**Jason Corum** (WGBH) – Developer

**Kara Van Malssen** (AVP) – Scrum Master



# *AMS 2.0 Roadmap*



**Milestone 1 – June 19, 2018**  
**Goal: Create/Edit Records in the UI**

**Milestone 2 – September 11, 2018**  
**Goal: Search/Discovery and Record Page Design**

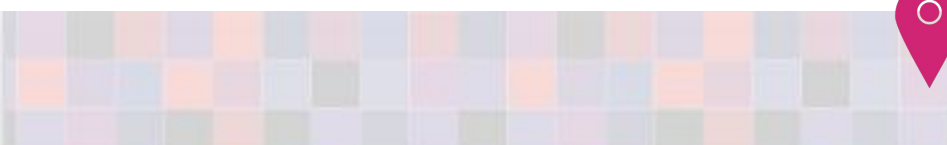
**Milestone 3 – November 6, 2018**  
**Goal: Media Display**

**Milestone 4 – November 20, 2018**  
**Goal: Batch Import/Edit**

**Milestone 5 – December 4, 2018**  
**Goal: Export and Reporting**

**Milestone 6 – December 18, 2018**  
**Goal: Batch Edit**

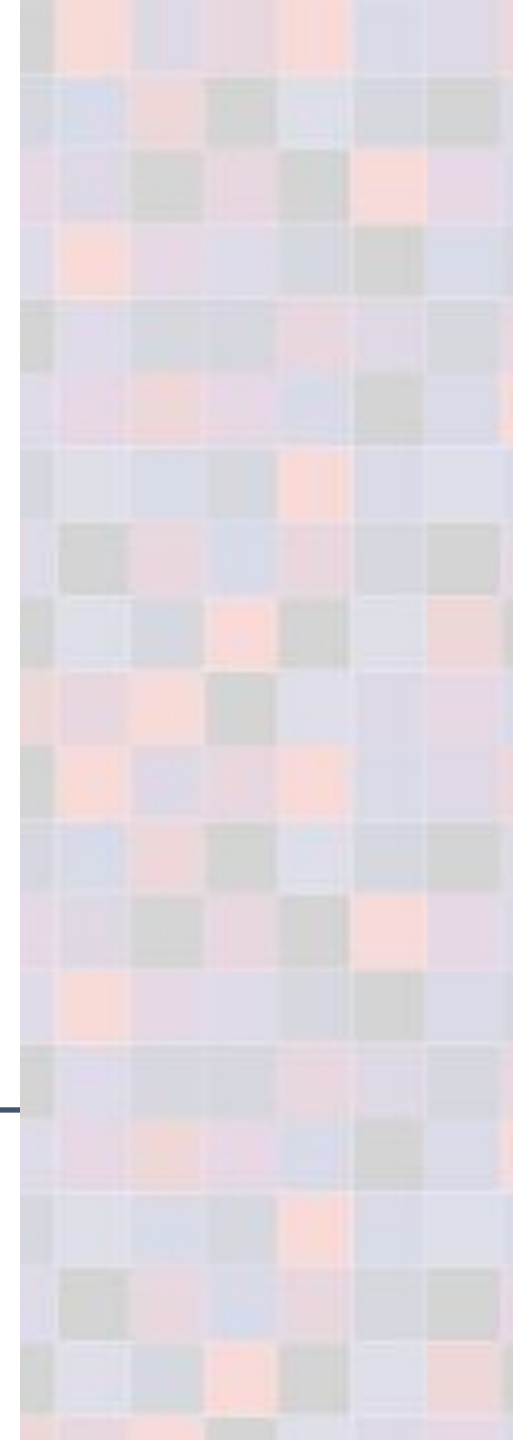
**Milestone 7 – February 12, 2019**  
**Goal: Migration**





# *AAPB METADATA*

---



# *PBCore Description Document*

## **Asset**

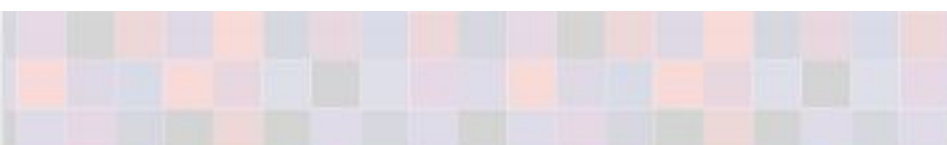
Title, Description, Subject, Contributors, Rights, etc.

## **Instantiations**

Physical Tape Format or Digital File Format, Duration, Media Type, Generation, Color, etc.

## **Essence Tracks**

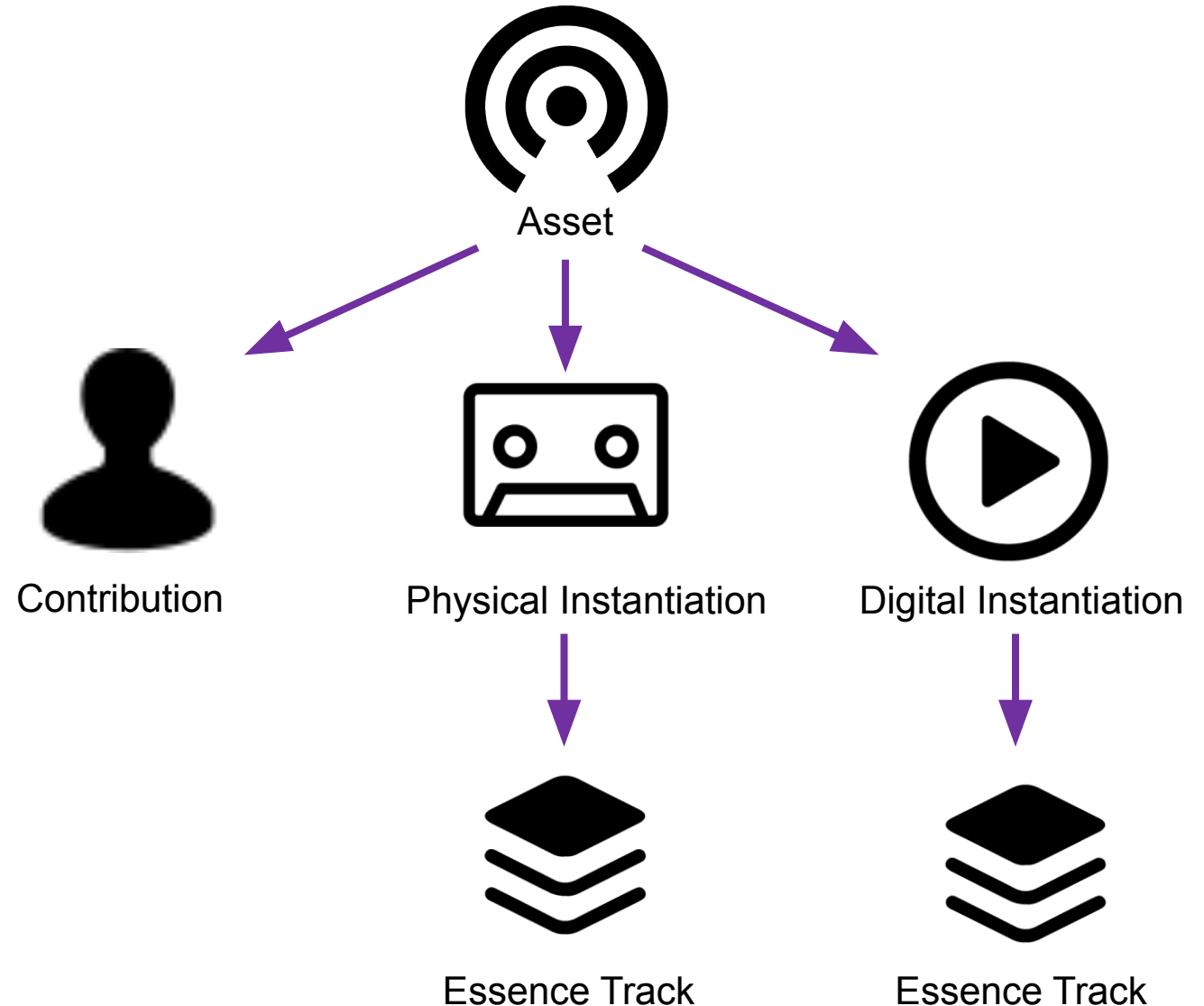
Data Rate, Encoding, Sampling Rate, Aspect Ratio, etc.



# *PBCore Work Types in Hyrax*

1 Description Document  
=/  
1 Work

Each repeatable section gets its own work type related to the others.



Models the idea of a person/org contributing to an Asset in a specific way.

Too many roles to have each one as a predicate


*What is a  
Contribution  
work?*



**Contributions** consist of

- Name (required)
- Role
- Affiliation
- Portrayal

# Simple Cataloging Workflow

 **Edit Work**

Descriptions

Files

Relationships

Sharing

Identifying Information

Title required

A name given to an asset.

Woman

Series

✕ Remove

Women's Health Care

Episode

✕ Remove

✕ Remove

+ Add another Title

Producing organization

The name of the organization(s) that produced the asset.

WNED

✕ Remove

✕ Remove

+ Add another Producing organization

Local identifier

An identifier used by a local holding institution.

+ Add another Local identifier

Save Work

Requirements

✓ Describe your work

Visibility

☐ Public

Make available to all.

☐ Institution

Restrict access to Institution.

☐ Embargo

Set date for future release.

☐ Lease

Set date for future reduced access.

☒ Private

Keep to myself with option to share.

Save & Create Digital Instantiation

Save & Create Physical Instantiation

By saving this work I agree to the

[Deposit Agreement](#)

Cancel

Save changes

*Embedding  
Smaller  
Child Works*

## Add New Asset

Descriptions

Files

Relationships

Sharing

To create a separate work for each of the files, go to [Batch upload](#)

### Identifying Information



### Subject Information



### Rights



### Credits



#### Contributor

Narrator



Price, Vincent

Affiliation

Portrayal

[+ Add another Contributor](#)

### AAPB Admin Data



## Save Work

### Requirements

Describe your work

### Visibility

- ☐ **Public**  
Make available to all.
- ☐ **Institution**  
Restrict access to Institution.
- ☐ **Embargo**  
Set date for future release.
- ☐ **Lease**  
Set date for future reduced access.
- ☒ **Private**  
Keep to myself with option to share.

Save & Create Digital Instantiation

Save & Create Physical Instantiation

By saving this work I agree to the

[Deposit Agreement](#)

Cancel

Save

# Embedding Smaller Child Works

## Relationships


In Administrative Set:  
[Default Admin Set](#)

## Asset Details

Producing organization	<a href="#">Arkansas Educational Television Network</a>
Program title	<a href="#">Arkansas School for the Deaf</a>
Program description	This documentary describes the courses and programs at the Arkansas School for the Deaf. In addition to standard curricula, students can take vocationally focuses courses. The school also uses computers in the classroom and for particular applications like speech classes. Other topics include field trips, dormitory living, and extra curricular activities. The documentary is composed of interviews with school administrators and teachers, along with footage and photographs of students in classrooms, around campus, and at special events.
Asset types	<a href="#">Program</a>
Genre	<a href="#">Documentary</a>
Broadcast date	1994-06-01
Rights summary	Copyright 1994 AETN. All Rights Reserved.
Topics	<a href="#">Technology</a>
GUID	cpb-aacip_600-7p8tb0xt22

Credits			
Role	name	Affiliation	portrayal
Producers	Sandage, Charlie		
AAPB Admin			

## Instantiations

Thumbnail	Details	Date Added	Holding Institute
	Format: <a href="#">application/mxf</a> Generation: <a href="#">Preservation Master</a> Duration: 00:29:53	2018-09-05	Arkansas Educational TV Network (AETN)
	Format: <a href="#">3/4 inch videotape</a> Generation: <a href="#">Master</a>	2018-09-05	Arkansas Educational TV Network (AETN)



# Asset-centric Search

## Limit your search

- Date >
- Asset Type >
- Topic >
- Genre >
- Producing Organization >
- Media Type >
- Physical Format >
- Holding Organization >
- Language >
- Level of user access >
- Transcript Status >
- Minimally Cataloged >
- Outside URL >
- Digitized Copy in AAPB >

« Previous | 1 - 10 of 14 | Next »

Sort by relevance ▾

10 per page ▾

    Export ▾

### Woman; Women's Health Care; Hello



**Admin Set:** Default Admin Set  
**Created Date:** 1975-10-21  
**GUID:** cpb-aacip\_600-251fj2998x  
**Description:** This episode features a conversation with Virginia Drachman. She is a Smithsonian Fellow, a...

### Voter's Pipeline; ERA: Is It Necessary



**Admin Set:** Default Admin Set  
**Created Date:** 1982-06-17  
**GUID:** cpb-aacip\_600-kp7tm7246m  
**Description:** This episode of Voter's Pipeline is a discussion about whether the Equal Rights Amendment should...

### Arkansas School for the Deaf



**Admin Set:** Default Admin Set  
**Broadcast Date:** 1994-06-01  
**GUID:** cpb-aacip\_600-7p8tb0xt22  
**Description:** This documentary describes the courses and programs at the Arkansas School for the Deaf. In...

# *Future Data Modeling Ideas*

## Broadcast Series

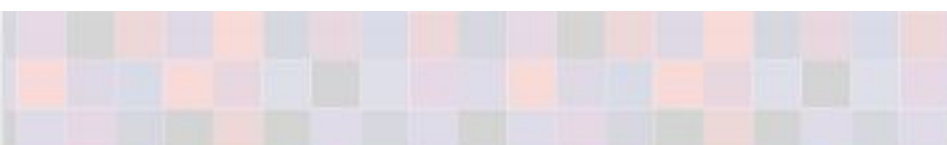
Currently series data is repeated on each Asset in the Series

Maybe use Custom Collection Type rather than Work Type

Assets can, but don't have to, belong to a Series

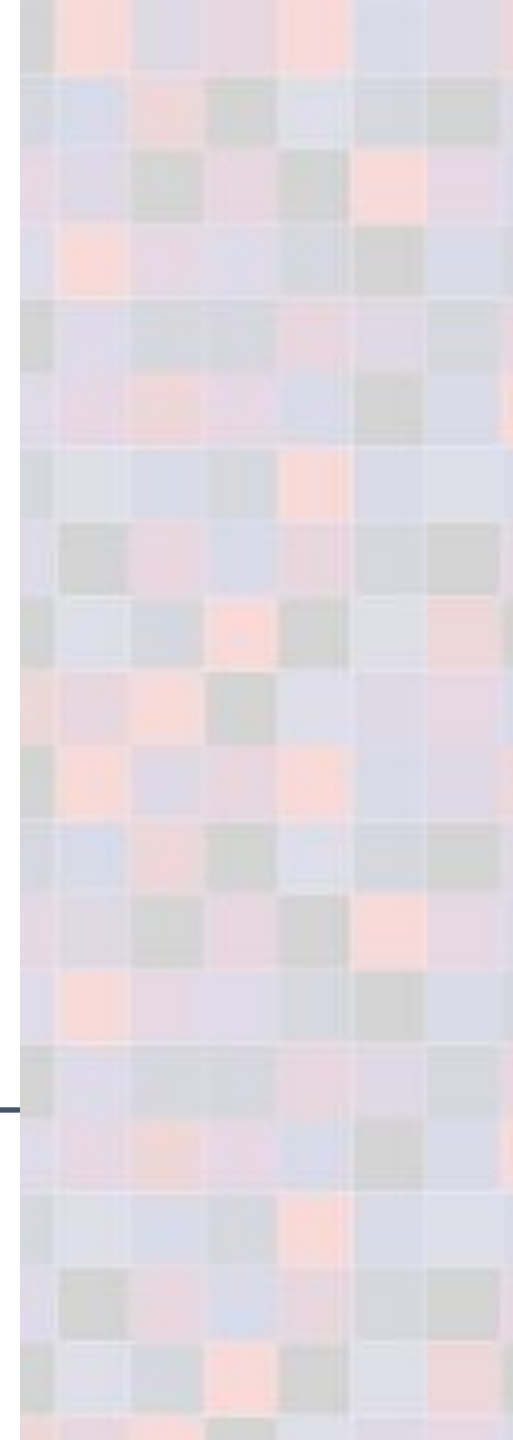
Assets can belong to multiple Series

When new Assets are imported an interface can help importers match them with Series they should belong to



# *ADVANTAGES OF HYRAX*

---

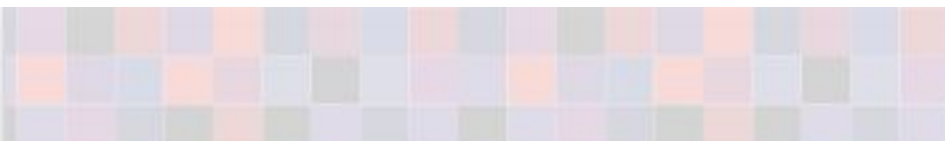


*What you get  
out of the box*

**Code Generation** – Simple to generate model, view, controller, actor stacks, tests

**Styling** – Looks nice and clean with minimal css additions required

**Flip flop functionality** – Easy to turn on/off features we may not need

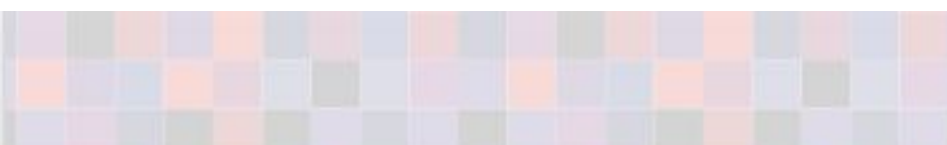


# *Flexibility*

**Configurable Work Types** – Easy for us to create multiple work types and define the possible relationships between them

**Custom Collection Types** – Looking into if we can use these to solve some data modeling challenges

**Changing Requirements** – Able to change requirement of file upload, checkbox on deposit agreement, and setting visibility levels

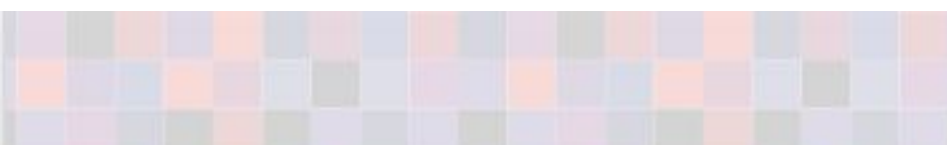


# *Samvera and Hyrax Communities*

**Real time support and feedback** – Regularly hear back quickly in Slack

**Access to original developers** – Easy to ping code authors with questions

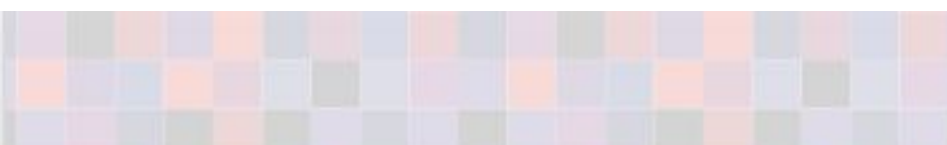
**Quick code resolution** – The PRs we've submitted back to Hyrax core have been merged within a day or so



# *Maturity*

More mature than previous  
Samvera applications

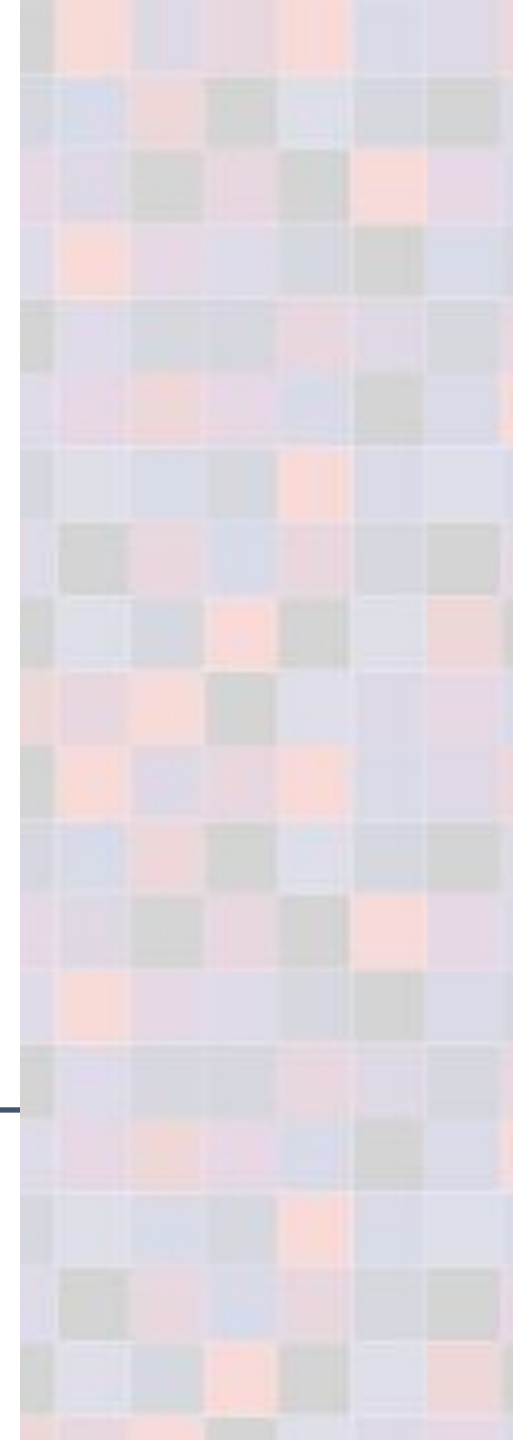
Started development on Hyrax  
2.0 and have been able to keep  
up with releases since then





# *CHALLENGES SO FAR*

---



# *Everything's under development*

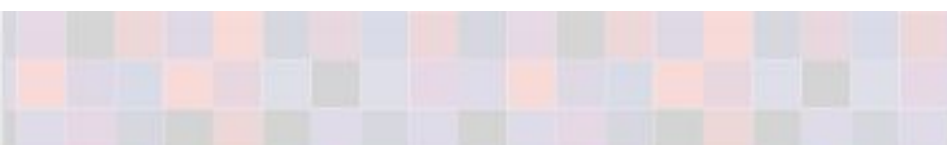
**Hyrax** – Building our application as our foundation changes

**Avalon Components** – In the process of componentizing features we want to use

**Do we wait?**

Sometimes yes – Avalon Media Player, Custom Collection Types

Sometimes no – generic batch import solution



*What should we  
contribute  
back?*

## **PBCore Data Model?**

Documentation, but not code

## **Batch Import component?**

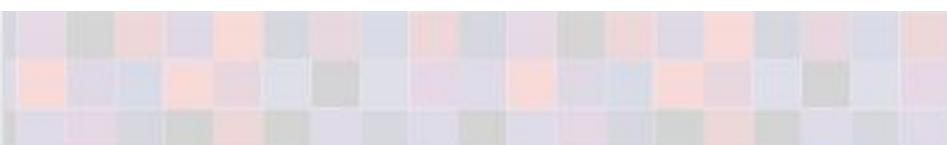
Building with the Avalon  
team, hopefully usable by others

## **Small Hyrax patches?**

Easier to submit, Quick turn  
around, helps us help ourselves

## **New minor features?**

As flip-flop features? Will  
anyone else use them?

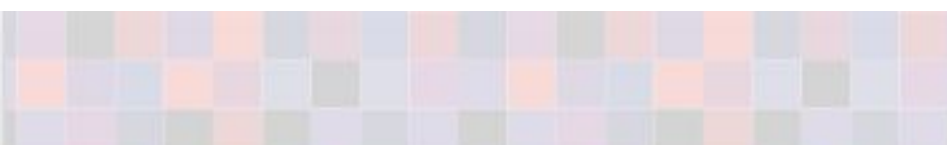


*Is there  
documentation  
for this?*

**Scattered Sources** – A lot of places to look, even for just one type of information or one section of the code

**Missing Information** – The documentation we have is great and helpful, but there isn't always a lot of it

**Known problem without enough resources to solve** – Join or support the Documentation Working Group if you can!



# *How much do we test?*

## **Feature specs**

- Initially relied on them a lot; then pulled back a bit
- What's the proper balance?

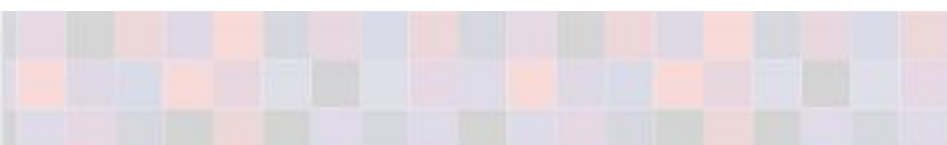
## **Setting up tests properly is cumbersome and mysterious**

- Factories that build/create dependent objects have helped
- Would be great if we could create mocks with more efficiency

## **Using Rspec and Capybara to DRY things up; improve consistency**

- We've created some custom matchers and helpers

These could be useful to the broader community



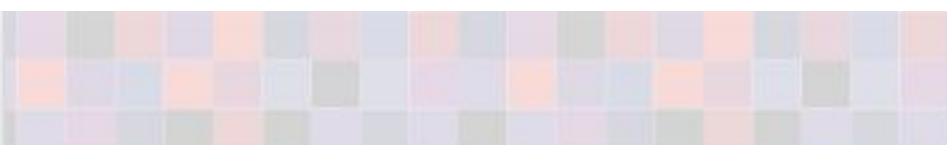
*Should we do  
our own thing?*

**Fighting the Framework** – When do we go around? And how?

**Is Hyrax a Framework or an Application?**

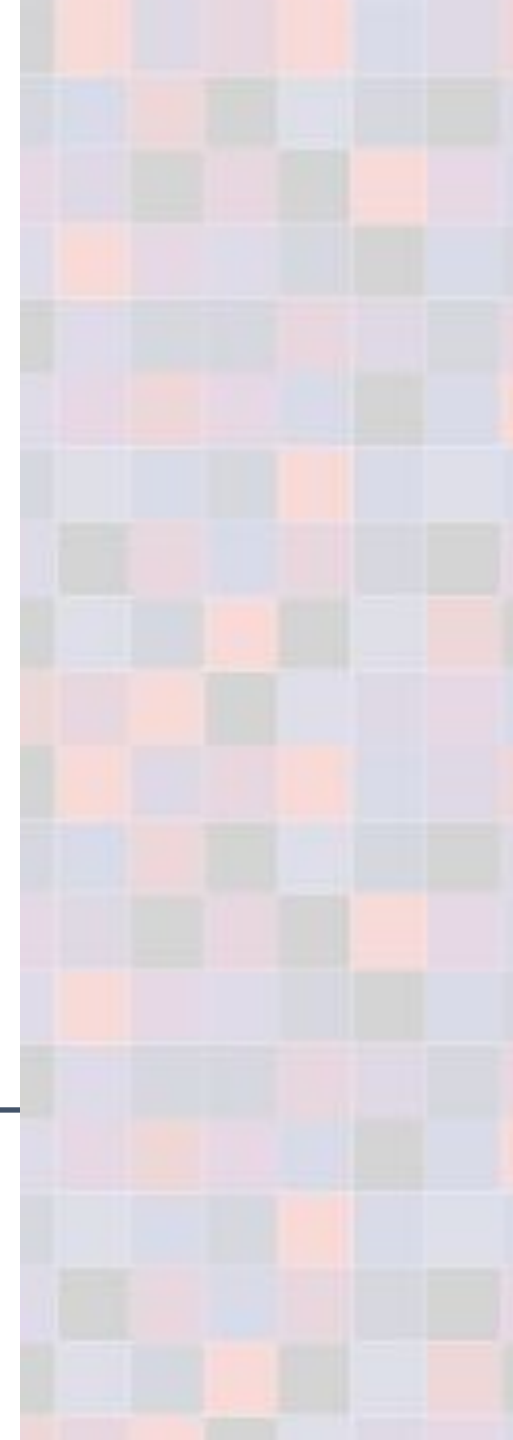
- Premise is a framework
- Behavior is an application
- Customizations seem expected

**A ticking time bomb** – Customizations might cause us more headache as we move to new versions of Hyrax.



# *The Future*

---



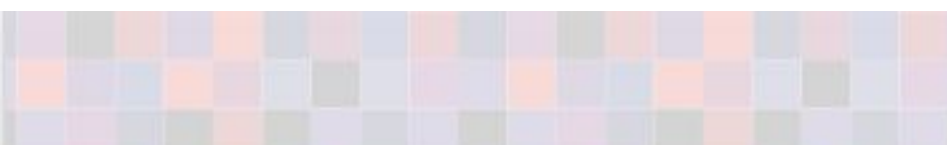
# *AAPB Future with AMS*

## **Continue Development**

- Hit the rest of our milestones
- Migrate in February 2019
- Continue to maintain and update after the grant ends to accommodate changing requirements

## **Roll Out to Contributing Organizations**

- Migrate their existing data for them
- Training on new cataloging and batch import workflows
- Promote as a viable option for small stations to manage their collections





# *Additional Future Development*



**WGBH Media Library and Archives received a 5 year challenge grant from NEH, in part to implement and further develop AMS for WGBH's needs.**

**Build additional features for WGBH's instance of AMS**

- Limiting access by IP
- Exploring using Valkyrie with Hyrax

**Integrate with WGBH's digital preservation metadata system Phydo**

- Share IDs
- Allow for navigation from one system to the other



[americanarchive.org](http://americanarchive.org)



[@amarchivepub](https://twitter.com/amarchivepub)



[facebook.com/amarchivepub](https://facebook.com/amarchivepub)



[@amarchivepub](https://instagram.com/amarchivepub)

# ***THANK YOU!***

**Casey Davis Kaufman**, *WGBH Media Library and Archives Associate Director / AAPB Project Manager*  
*casey\_davis-kaufman@wgbh.org*

**Sadie Roosa**, *WGBH Media Library and Archives Technical Project Manager / AMS 2.0 Product Owner*  
*sadie\_roosa@wgbh.org*

**Andrew Myers**, *WGBH Media Library and Archives Supervising Developer*  
*andrew\_myers@wgbh.org*

**Jason Corum**, *WGBH Media Library and Archives Web Developer*  
*jason\_corum@wgbh.org*

