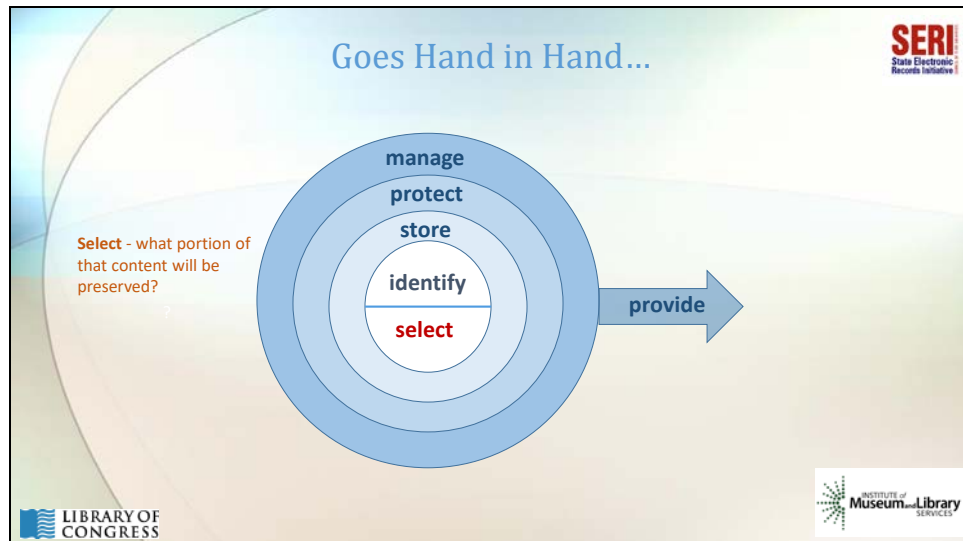




Welcome to Managing Digital Content Over Time. This training was produced by the State Electronic Records Initiative in coordination with the Council of State Archivists. It was developed under a grant from the Institute of Museum and Library Services and based primarily on training created by the Library of Congress. It is designed to help archivists and others who manage digital content understand the necessary steps of digital preservation. This is module 2, Select.

Slide 2



Note that the identify and select processes happen at the same time.



This is a picture of a log jam on the Saint Croix River in northwest Wisconsin in the 1880s.

After you've compiled your inventory, if you don't set some priorities, it can be easy to get overwhelmed. Meanwhile, you've still got more logs—more new digital content—coming in down the river. The goal in selecting specific types of content to preserve is to get your logs moving again—pick a few things to tackle so everything can start flowing more efficiently.


Not all of the content you're dealing with may in fact be appropriate or necessary for you to preserve, and you don't want to commit resources to preserving materials you don't have to.


**SERI**  
State Electronic  
Records Initiative


## Why Be Selective?

Storage may be cheap, management is not  
... especially over time

1 Tb Hard drive= \$100 + IT Department = \$100 hour



 LIBRARY OF CONGRESS

 INSTITUTE of  
Museum and Library  
SERVICES

Why be selective?


The hidden cost of a hard drive is in the people who will need to support it. With every terabyte you add, you add the need for hours of more expensive management.

You may hear people argue that storage is cheap so we should keep everything. Unfortunately, that perspective is rather short-sighted. Storage may be cheap, but preserving the **quality of content over the long-term** is not. And storage options are not equally reliable even in the short term.

There are **periodic migration costs**, moving the digital materials into systems where you will preserve it, plus **monitoring** files for **corruption and change**. Have you lost bits? Are the files degrading?

**Why be selective?**

Discovery and dissemination services  
... scale, scope, performance, sustainability



The screenshot shows a historical map on the left with numbered red dots (1-7) and a document snippet on the right. The map is titled 'Historical map portion of the map of the world' and the document is dated 'May 3, 1863'. The document text includes 'The 1863, the first year of the year of the world' and 'The 1863, the first year of the year of the world'.

**LIBRARY OF CONGRESS**

**SERI**  
State Electronic  
Records Initiative

**INSTITUTE of  
Museum and Library  
SERVICES**

Not to mention maintaining **access** to the files, which means updating your **discovery and dissemination services** every time hardware and software change, which is an ongoing, recurring cost.

The idea behind long-term preservation is that you will be making this **content available in the future**. It isn't enough just to save the content if you can't access it any more. Consider if you have the user interfaces to deliver the content you choose.

Why be selective?

Quality and significance of content

LIBRARY OF CONGRESS

Atlas of Digital Damages: <http://www.flickr.com/photos/69771128@N02/8187703241/in/pool-2121762@N23>

SERI  
State Electronic  
Records Initiative

INSTITUTE of  
Museum and Library  
SERVICES

The slide features a central collage of faces with various digital artifacts and distortions. The background is a light blue and white gradient with a large, faint, stylized 'C' shape. Logos for the Library of Congress, SERI, and the Institute of Museum and Library Services are present.

Even if we could keep everything forever, would we want to?

Is that manageable given the type of content that you hold?

**Not** all digital content may be **preservation quality** – if you have high resolution photos, do you also need to preserve lower quality versions of these images?

And not all will be **significant** enough to warrant preservation, including that string of emails about organizing the staff Christmas party.

Why be selective?

Match mission to content....

The mission of the Division of Archives and Records is to collect, preserve, manage, and provide access to information that protects citizen rights, documents North Carolina history, and culture, promotes transparency, and encourages stewardship of government records.

LIBRARY OF CONGRESS

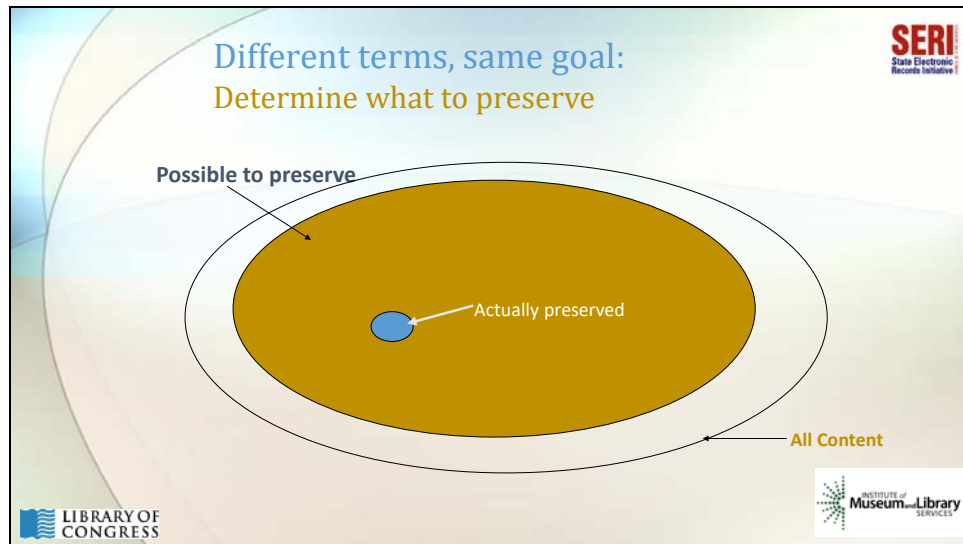
SERI  
State Electronic  
Records Initiative

INSTITUTE of  
Museum and Library  
SERVICES

Does the digital content we take in match our **mission** and **scope of collections**?

Quite often materials find their way to us that have little or nothing to do with our mission, yet we give them a home and expend our resources on maintaining them.

Maybe there is a better/more logical home for that content? Maybe you could partner with another organization that is better placed to hold and preserve that content.

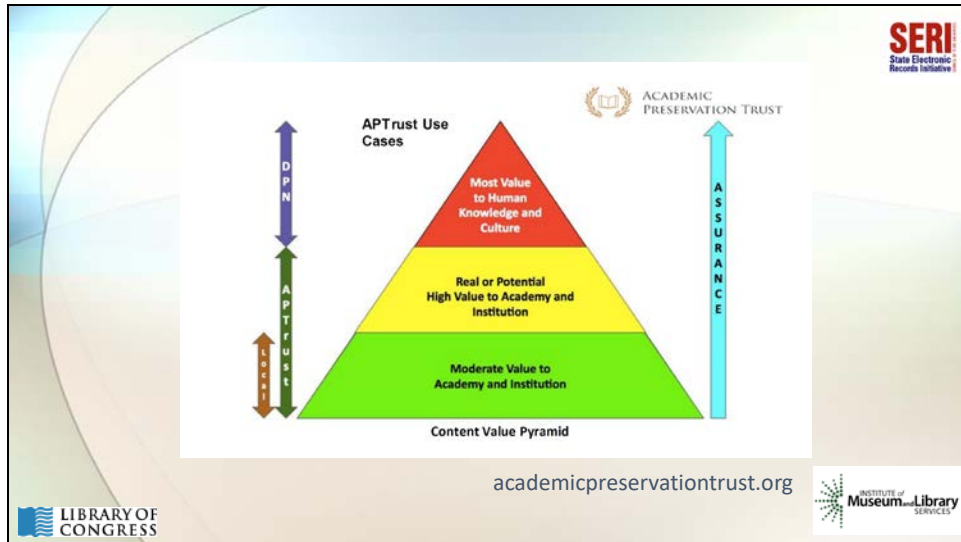


The selection process for digital content is very analogous to the selection process for non-digital materials – you don't collect materials for your archive that don't match your mission, and you should keep that same principle in mind when selecting digital content.



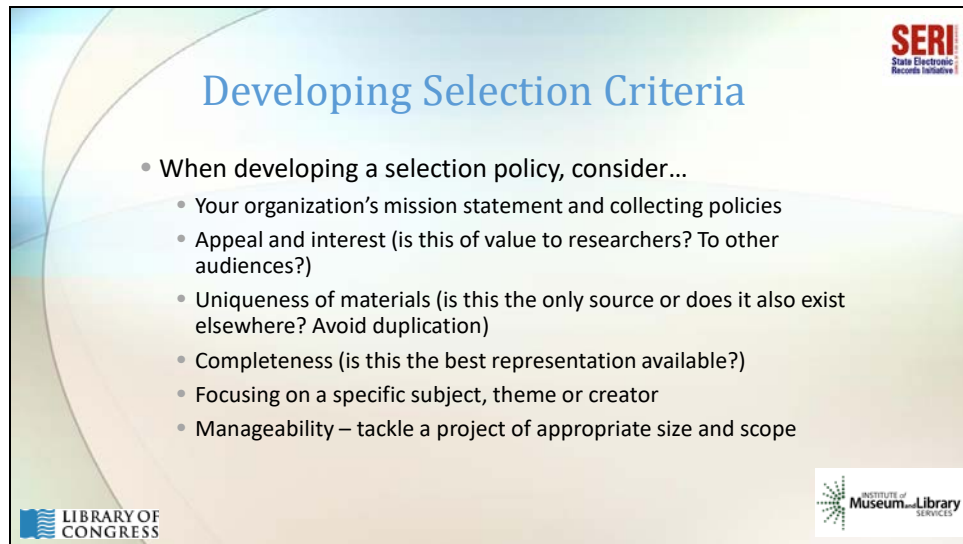
The **Basic steps for Selection** require you to:

- **Review** your potential digital content – start with the outcomes of your inventory; look over what you have and think you might have coming in. Understand the implications.
- **Define** and then **apply** criteria for what you will select to preserve. It's the best way to ensure consistency (across an organization, over time and staffing changes).
- **Document** (and **preserve**) selection decisions: Why are you keeping things? What is your rationale? You, your staff, and your successors, need to understand why you chose to keep that particular content. Don't assume it will be obvious to everyone.
- **Implement** your decisions – and stick to your criteria! Don't take in or keep content not in your defined scope of preservation. Review your selection criteria regularly to ensure they meet your needs. They are there to ensure consistency and can also be a helpful tool in controlling what content comes your way, which is an argument in your arsenal for those times when you need to say 'no' to someone.




This diagram comes from the Academic Preservation Trust which is a consortium of higher education institutions which provides both a preservation repository for digital content and collaboratively-developed services related to that content.


This provides a nice visual of things you might find as you move through your selection process. The largest piece may be those things of moderate value. Considerably fewer items will have high value and there will be a small list of treasures you feel you **MUST** have and preserve.



## Developing Selection Criteria

- When developing a selection policy, consider...
  - Your organization's mission statement and collecting policies
  - Appeal and interest (is this of value to researchers? To other audiences?)
  - Uniqueness of materials (is this the only source or does it also exist elsewhere? Avoid duplication)
  - Completeness (is this the best representation available?)
  - Focusing on a specific subject, theme or creator
  - Manageability – tackle a project of appropriate size and scope

 LIBRARY OF CONGRESS

 INSTITUTE of  
Museum and Library  
SERVICES

**SERI**  
State Electronic  
Records Initiative

Appraisal of electronic records is similar in many ways to that of other media. Each institution should have a policy which spells out records that are within their collection scope, including mission, customer interest, uniqueness, range of content, and specific subjects or themes which may be well-suited to your area.



**Consider your stakeholders**

- Content producers
- Researchers
- Community members
- Administrators
- Funders

The slide features a background with soft, overlapping circular gradients in shades of blue, green, and yellow. In the top right corner is the SERI logo (State Electronic Records Initiative). In the bottom left is the Library of Congress logo. In the bottom right is the Institute of Museum and Library Services logo. An illustration of seven stylized human figures in various colors (orange, teal, grey, pink, yellow, blue, and black) holding hands is positioned to the right of the list. Above each figure is a speech bubble in a matching color.

There's a lot to consider when creating a valuable preserve-able digital collection that can satisfy all the key stakeholders. It's clear that you need to develop a policy before coming to selection. You are making very important decisions that the organization will have to live with for the duration of the collection (hopefully a very long time).

**Practical Considerations**

Stop if or when the answer is **NO**

- **Content**
  - Does the content have value?
  - Does it fit your scope and mission?
- **Technical**
  - Is it feasible for you to preserve the content?
- **Access**
  - Is it possible to make the content available?
  - Are you the only holder of this content?

Logos: SERI State Electronic Records Initiative, LIBRARY OF CONGRESS, INSTITUTE of Museum Library SERVICES

Even if something fits your desired criteria, it still might **not** be reasonable for you to select it.

You've already considered the **content** in view of your selection criteria. And you should already have answered 'yes' to both of these Questions to continue considering the materials you hold.

—does the content have long term value?

—does it fit your scope and mission?

Next you need to consider **Technical** issues:

—is it **feasible** for you to **preserve** the content? [Is it a “**digital time bomb**”?]

—Some formats are a challenge to preserve, such as video/time-based media.

—Some may be too damaged to preserve. Do you have the skills and resources (either to undertake the preservation yourself or to buy the skills in)?

—Some types of material may require far more expertise and resources than you have available.

And **Access**.

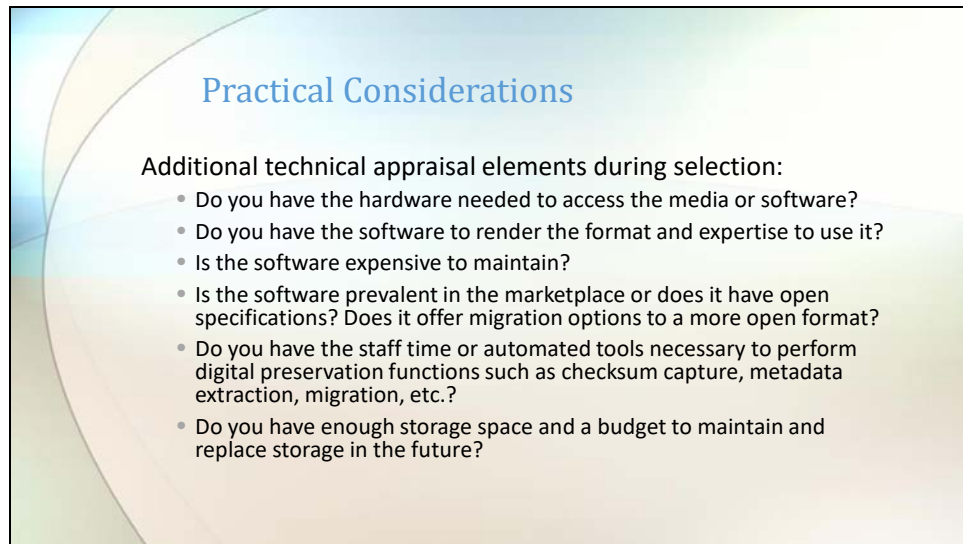
Even if we're not making it public, how useful is a server full of digital content that is safe, but that we can't access?

We need to ask:

—is it possible to make the content available over time?

—Are you the only holder of this content? [Is there duplication?]

If it is not feasible to preserve the content, and not possible to make it available and usable, then it probably shouldn't be included in your selection –especially if you know you are not the only holder of this digital content.



The slide features a light blue and green background with abstract curved lines. The title 'Practical Considerations' is centered at the top in a blue font. Below it, the text 'Additional technical appraisal elements during selection:' is followed by a bulleted list of seven questions.

## Practical Considerations

Additional technical appraisal elements during selection:

- Do you have the hardware needed to access the media or software?
- Do you have the software to render the format and expertise to use it?
- Is the software expensive to maintain?
- Is the software prevalent in the marketplace or does it have open specifications? Does it offer migration options to a more open format?
- Do you have the staff time or automated tools necessary to perform digital preservation functions such as checksum capture, metadata extraction, migration, etc.?
- Do you have enough storage space and a budget to maintain and replace storage in the future?

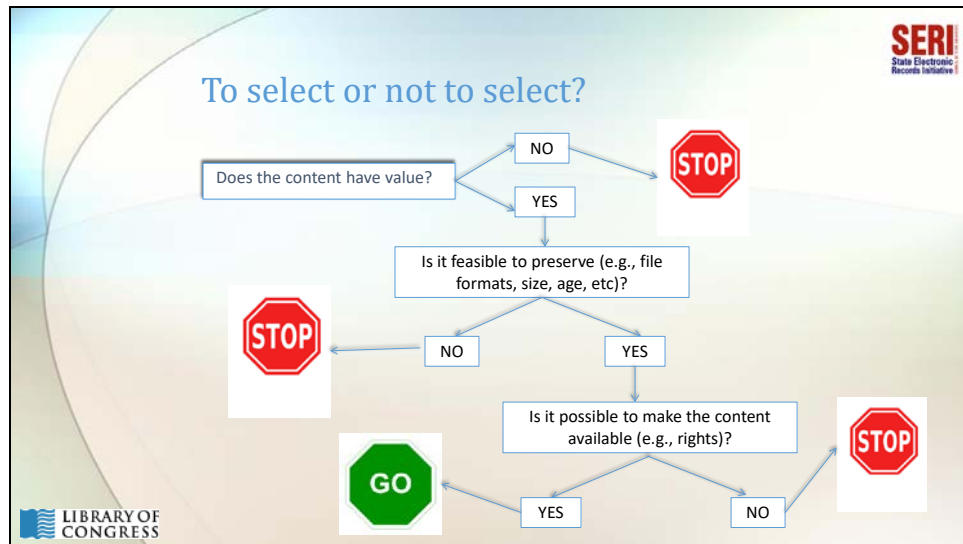
Here are some additional considerations for technical appraisal, which looks at things such as hardware and software availability, staff expertise, workload, automation, budgets, and the ability of the institution to physically care for these records, over and beyond whether or not the content is appropriate for your collection scope.

Slide 15

See full list at <https://archives.utah.gov/documents/appraisal-checklist.pdf>

Appraisal Checklist		Score (0-100)
Category		
<input type="checkbox"/> Cost to preserve Minimal		<input type="text"/>
<input type="checkbox"/> Cost to preserve Moderate and not within budget		<input type="text"/>
<input type="checkbox"/> Cost to preserve Moderate but within budget		<input type="text"/>
<input type="checkbox"/> Evidential Records provide evidence of creating entities' functions and actions		<input type="text"/>
<input type="checkbox"/> Fiscal Records identify fiscal actions and are subject to audit		<input type="text"/>
<input type="checkbox"/> Format Records are in a format that is hard to understand or use		<input type="text"/>
<input type="checkbox"/> Format Records are in a format that is most beneficial to designated community		<input type="text"/>
<input type="checkbox"/> Format Records are in a format that requires regular migration		<input type="text"/>
<input type="checkbox"/> Format Records are in a format that the designated community cannot use		<input type="text"/>

You may choose to create an appraisal checklist, combining content-based appraisal criteria with technical and access considerations. Adding a score to a line item can assist with hard decision-making. A score may have a positive or negative value, depending on the impact of the appraisal criteria.




Or you can use a decision tree to help you decide what's practical to preserve.

## Setting Priorities

To prioritize your review process, ask yourself which content is . . .

- most **significant** to your organization?
- most **extensive**?
- most **requested/used**?
- **easiest**?
- **oldest**?
- **newest**?
- **mandated**?
- **at risk**?



A woman helps four children organize a party  
Wisconsin Historical Society WHI-96879

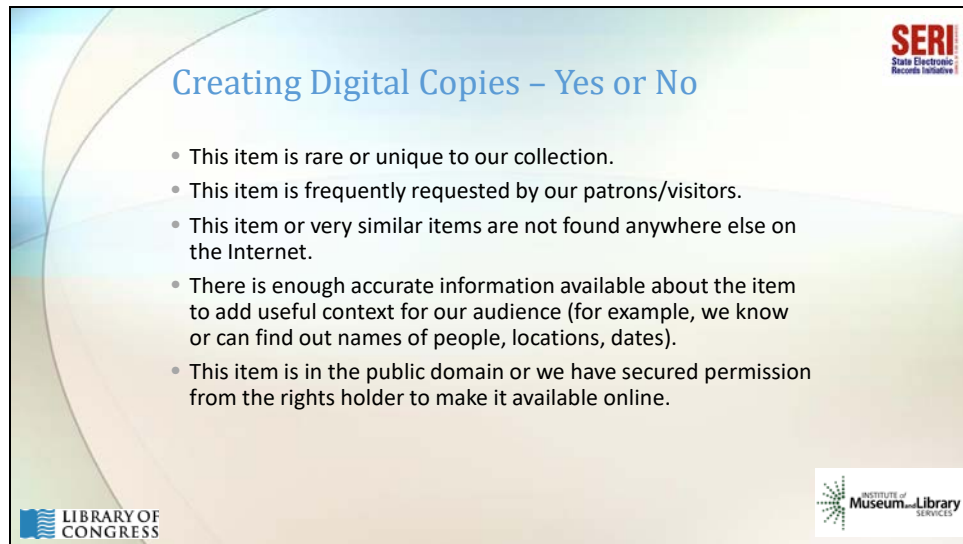
**SERI**  
State Electronic  
Records Initiative

LIBRARY OF CONGRESS

Once you have your selection criteria, it may not be possible to review/select everything at once, so how might you sequence the process? Again, the answer will be different for each organization.

Think about what's

- most **significant** to your organization?
- most **extensive**? (and therefore a more coherent body of material to manage)
- most **requested/used**?
- **Easiest** to tackle (e.g. most familiar, most ready for ingest – a quick win for your digital preservation process, which is very helpful when you are having to prove the value of your efforts to a reluctant administration)
- **Oldest** (possible historical importance)
- **Newest** (possible immediate interest)
- **Mandated** (via local policies, legislation, etc.)
- **At risk**? If it were no longer available, what digital files would be the hardest to replace? Some formats become obsolete a lot faster than other formats. PDFs are viable for a really long time – video files, however, get old very quickly.



**Creating Digital Copies – Yes or No**

- This item is rare or unique to our collection.
- This item is frequently requested by our patrons/visitors.
- This item or very similar items are not found anywhere else on the Internet.
- There is enough accurate information available about the item to add useful context for our audience (for example, we know or can find out names of people, locations, dates).
- This item is in the public domain or we have secured permission from the rights holder to make it available online.

**SERI**  
State Electronic  
Records Initiative

**LIBRARY OF CONGRESS**

**INSTITUTE of  
Museum and Library  
SERVICES**

You may choose to digitize your analog records.

Is this item rare or unique to our collection?

Is this item frequently requested by our patrons/visitors?

Is this item or very similar items not found anywhere else on the Internet?


Is there enough accurate information available about the item to add useful context for our audience?

Is this item in the public domain or have we secured permission from the rights holder to make it available online?

If you answer “no” to any of these questions, the item may not be a good candidate for digitization.

## Considering copyright

- Disclaimer: We are not lawyers.
- Owning a physical item does not necessarily mean you hold the copyright to that item.
- Public domain = no longer under copyright. In the US in 2016 that means the item was:
  - Published before 1926 –OR–
  - Unpublished; creator died before 1946 –OR–
  - Unpublished; unknown creator; made before 1896



Original image © 2008 University of Wisconsin-Milwaukee Libraries

LIBRARY OF CONGRESS

UW-Milwaukee Libraries

INSTITUTE of Museum and Library SERVICES

SERI  
State Electronic  
Records Initiative

If the records you collect are not government-produced and therefore in the public domain, then copyright is a consideration when determining if access rights will impede preservation. Be sure to document rights information in the record's metadata so that they are known, and perhaps the record can be available to the public in the future.


**SERI**  
State Electronic  
Records Initiative

## Considering copyright

- Works under copyright, copyright holder is known:
  - Contact copyright holder IN WRITING to request permission to make available online.
- Works presumed to be under copyright; copyright holder is unknown or cannot be located:
  - Due diligence has been made to identify and locate copyright holder.
  - Be prepared to remove item from digital collection if challenged.

*"In the Land of Woods, Water and Wildlife"*  
*Three Lakes, Wis.*  
**NORTHERNAIRE**  
*"Delightfully Different"*

- ON CHAIN OF 38 LAKES
- FISHING - perfect for record smuckers
- BERNIESE & HER WILDERNESS ORPHANS
- FLOOR SHOW & DANCING at our famous SHOWBOAT
- JOOL-F, FLYDAY & SPEED BOAT CRUISES
- INDOOR POOL & OUTDOOR BEACH
- MARKED NATURE TRAIL into the Wilderness



Three Lakes Historical Society

 LIBRARY OF CONGRESS INSTITUTE of  
Museum and Library  
SERVICES

Make every effort to contact copyright holders. Preserve any permissions given right along with the records they're about.

A presentation slide titled "Copyright Tools" with a light blue and yellow background. The slide lists two main categories of tools: "Public Domain Sherpa: Public Domain Calculator" and "Copyright Advisory Network". Each category includes a specific tool with a URL. Logos for "SERI State Electronic Records Initiative", "LIBRARY OF CONGRESS", and "INSTITUTE of Museum and Library SERVICES" are visible in the corners.

## Copyright Tools

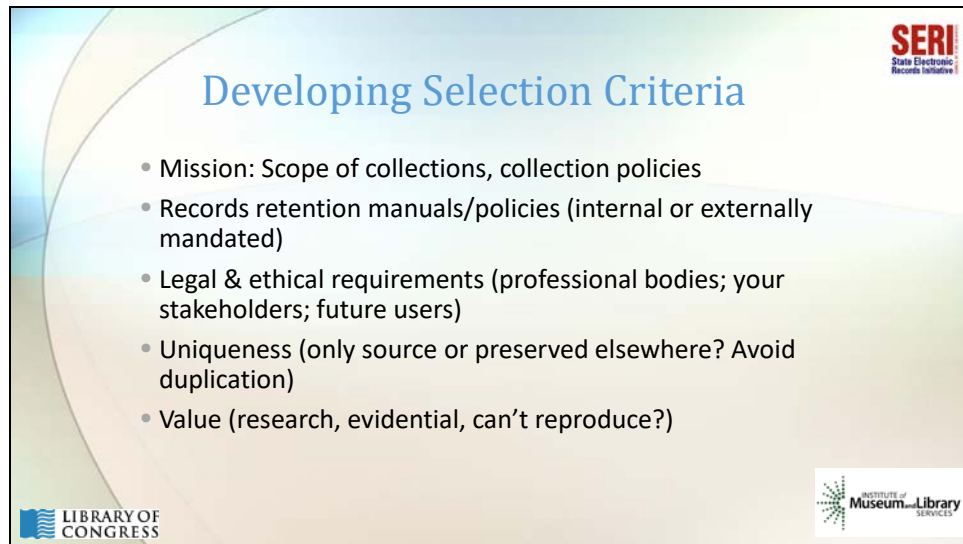
- Public Domain Sherpa: Public Domain Calculator
  - <http://www.publicdomainsherpa.com/calculator.html>
- Copyright Advisory Network
  - Copyright Slider: <http://librarycopyright.net/resources/digitalslider/>
  - Copyright Genie: <http://librarycopyright.net/resources/genie/>

LIBRARY OF CONGRESS

SERI  
State Electronic  
Records Initiative

INSTITUTE of  
Museum and Library  
SERVICES

Here are some links to resources that may be helpful when working with copyrighted material. The resources are from Public Domain Sherpa and the Copyright Advisory Network.



## Developing Selection Criteria

- Mission: Scope of collections, collection policies
- Records retention manuals/policies (internal or externally mandated)
- Legal & ethical requirements (professional bodies; your stakeholders; future users)
- Uniqueness (only source or preserved elsewhere? Avoid duplication)
- Value (research, evidential, can't reproduce?)

**SERI**  
State Electronic  
Records Initiative

LIBRARY OF  
CONGRESS

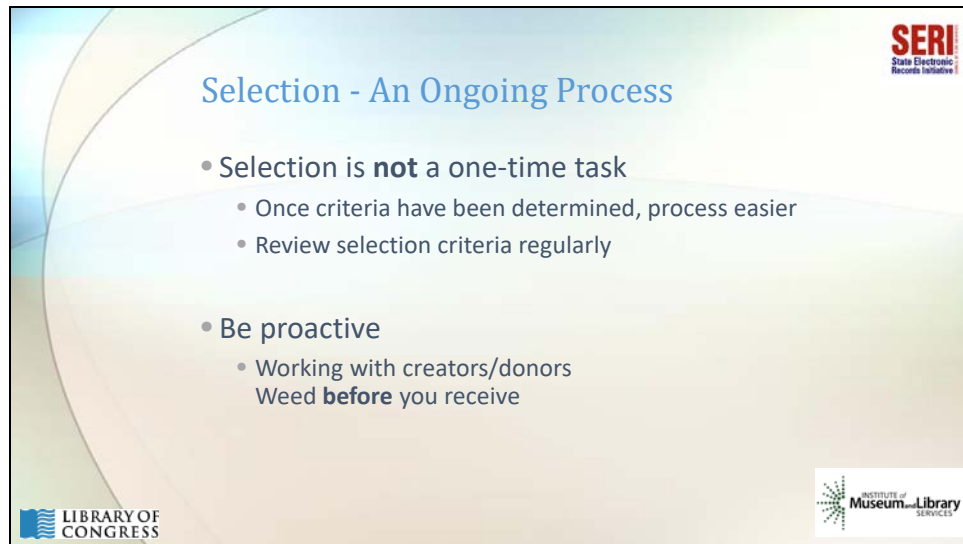
INSTITUTE of  
Museum and Library  
SERVICES

When you're first getting started, it's helpful to treat selection as a managed, **structured, project, in order** to plan and coordinate the process [and plan for the future]. The selection criteria you choose will be uniquely specific to your situation, your organization, and its mission.

So where can you go for guidance to begin this project of defining your selection criteria?

- Look **inside** your organization first: are there **mission-related documents** that might give you clues? Existing **manuals and policies**, such as records retention schedules? Or Collecting policies?
- Also look **outside** your organization: Are there **legal restrictions** and/or **ethical requirements** that will guide your choices?
- On the question of **uniqueness**, you may not want to include anything that is preserved elsewhere. You may want to focus only on what meets the needs of your primary audience.
- And the **value** of materials - determined by a variety of factors - must be assessed in light of your own situation, the materials themselves, and their place in their wider context, whatever that may be.

Taking this **wider view** will enable you to make intelligent choices regarding your selection. Once you have clarified the ideal of what you **WANT** to preserve, then you're ready to consider what you are actually **ABLE** to preserve.



The slide features a light blue and yellow background with abstract curved lines. The title 'Selection - An Ongoing Process' is in blue. The bullet points are in black. The SERI logo is in the top right, and the Library of Congress and Institute of Museum and Library Services logos are in the bottom left and right respectively.

## Selection - An Ongoing Process

- Selection is **not** a one-time task
  - Once criteria have been determined, process easier
  - Review selection criteria regularly
- Be proactive
  - Working with creators/donors
  - Weed **before** you receive

LIBRARY OF CONGRESS

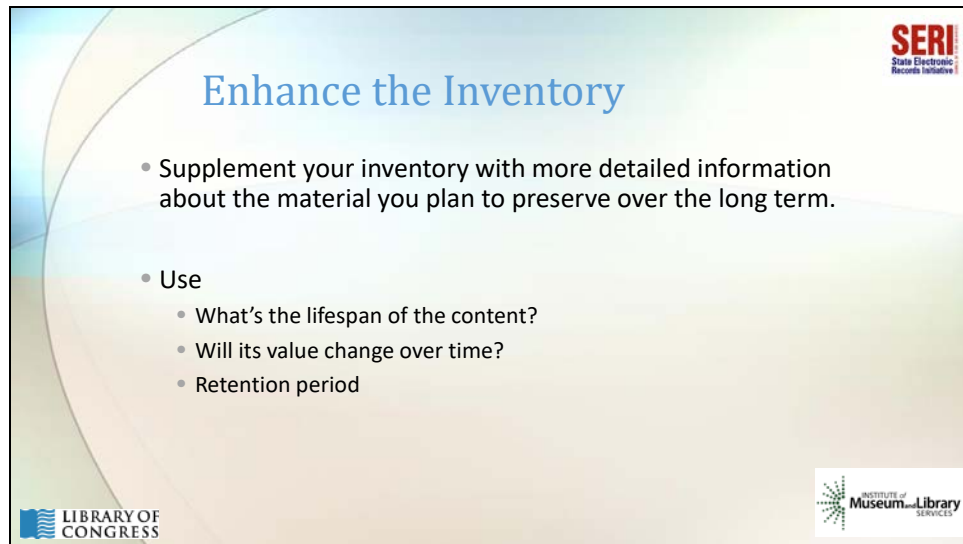
SERI  
State Electronic  
Records Initiative

INSTITUTE of  
Museum and Library  
SERVICES

Because digital preservation is a long-term commitment it's important to establish solid, ongoing relationships with the creators of your digital content. How many of you are managing digital content created by people outside of your library, archives, other departments, or maybe even other institutions?

**Communication is key** – particularly when the content is from external creators. You'll need to agree on terms for the transfer and retention of digital content to your library (and even where it's from others within your library).

Ideally, you'd want to review the content with the creators to determine which of their material is really important to be preserved, and ensure that what they're giving you meets your selection criteria. Be aware that most content creators don't have a clue as to what an archival format is, or how to create content that is likely to be manageable for long-term access. Education of content creators is very important. Working with them at the outset can save you many headaches later.



**Enhance the Inventory**

- Supplement your inventory with more detailed information about the material you plan to preserve over the long term.
- Use
  - What's the lifespan of the content?
  - Will its value change over time?
  - Retention period

**SERI**  
State Electronic  
Records Initiative

**LIBRARY OF CONGRESS**

**INSTITUTE of Museum and Library SERVICES**

Remember that you need to **document your selection process**. Start out by adding information to your inventory for material that you plan to preserve over the long term.

Supplement your inventory with usage statements.

What's the Lifespan of the content? Does its value/use change over time? When will content no longer be active? [retention period: how long will you retain it?]

**Access and Rights**

- Access
  - How will the public access the content?
  - Is access restricted? How? For how long?
- Rights
  - Who owns the rights to preserve and disseminate?

**SERI**  
State Electronic  
Records Initiative

**LIBRARY OF CONGRESS**

**INSTITUTE of  
Museum and Library  
SERVICES**

How will the public access the content? Who owns the rights to preserve and disseminate? Also include access and rights statements.

**Significance**

- Data criticality
  - Is it only in digital form?
  - Do we hold the only copy?
- Business/mission criticality
  - If we lose it, what's the damage to our reputation?
  - How would loss impact our services?

**SERI**  
State Electronic  
Records Initiative

**LIBRARY OF CONGRESS**

**INSTITUTE of Museum and Library SERVICES**

Don't forget the significance of the items. How critical are they to providing evidence of the activity they document? If the records are government-produced, they could document citizen's rights. If they are lost, what would be the impact to our user community as well as our institutional reputation?

## Data Criticality

**Data Criticality (1--> 5)**

- 1 - Digital and we hold the only copy - if we lose it, it's gone forever
- 2 - We have a digital copy but physical copies are at high risk (ex: Audio tapes)
- 3 - We have a digital copy but physical copies reside elsewhere
- 4 - We have a digital copy but digital copies reside elsewhere
- 5 - We Have a digital copy and still hold original physical item

**Business Criticality**

- 1 - Irrecoverable - Permanent loss of digital object would result in significant damage to our reputation and have a highly negative public relations impact.
- 2 - Major Impact - Temporary loss of digital object would result in Major damage to our reputation, or have a major negative public relations and/or financial impact (cost to replace or loss of revenue) or require major staff resources to correct.
- 3 - Minor Impact - Temporary loss of digital object would result in minimal damage to our reputation, or have a minor negative public relations and/or financial impact (cost to replace or loss of revenue) or require few staff resources to correct.
- 4 - No Impact - Temporary loss of digital object would result in no damage to our reputation, or have a negligible public relations and/or financial impact (cost to replace or loss of revenue) or require few staff resources to correct.

**LIBRARY OF CONGRESS**

**SERI**  
State Electronic  
Records Initiative

27

This is one way to determine which things you need to focus on after an event, which can be added to your inventory. It categorizes the degree of criticality.



Selection has several outcomes.

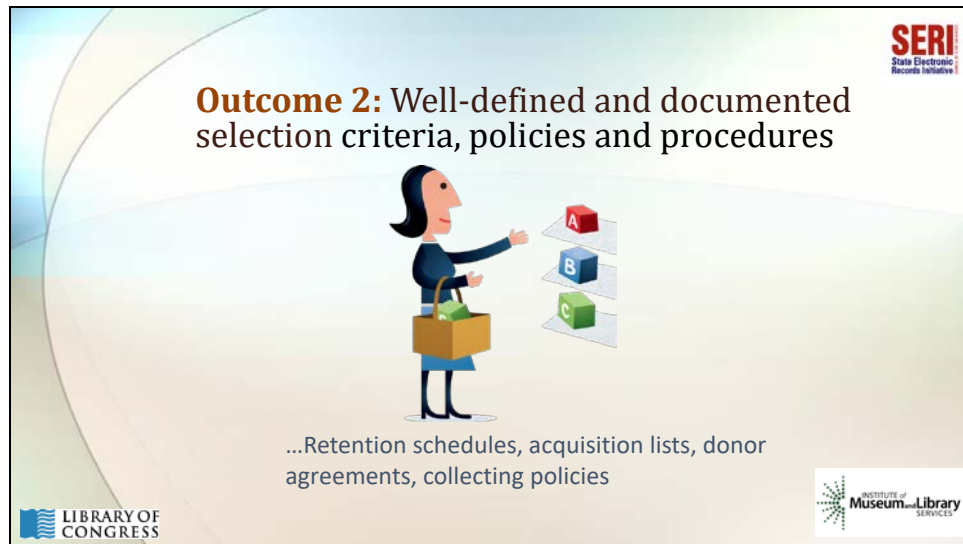
The first outcome should be an expanded inventory of content to preserve.

This allows you to focus your resources to maximum effect.

Know what you want/need to preserve.

Acquire accordingly.

Greater knowledge = greater control.



The second outcome should be well-defined and documented selection criteria, policies, and procedures.

So as you are going through this process, you need to document what your decisions are at each step. Why? So each time going forward, when you have a new collection and decision to make, you don't have to think too hard about how to proceed.

This is why we collect or don't collect something...

This way, you aren't the only one to know "the rules". Anyone in your institution should be able to pick them up and understand why a decision was made and you can consistently make the same decisions as an agency – it's an agreed-upon protocol.

It also gives you a leg to stand on when you need more resources. "This is our policy, but we can't right now because we don't have staff, enough storage, etc"

These also help manage expectations for others – "We won't take this amazing collection because it doesn't meet our institutional mission or collection policy," or "We should work together to transfer that collection because of that retention schedule."

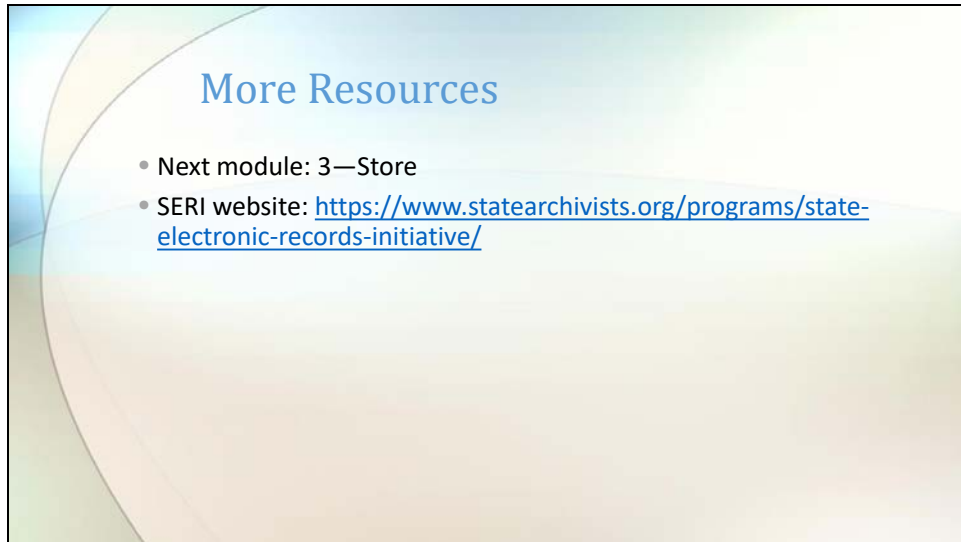


The third outcome should be a better understanding of content for future planning and growth.

Going through the work of selection is to gain a sense of control over what you have to deal with, what your scope is, and what your policies and priorities are for selection. This is critical to developing a sustainable program for support of long-term preservation and access.

By applying your selection criteria to your inventory, you will have more detailed information to work with in your planning. This documentation can also inform your work with creators of digital content. This might include the creation of submission agreements or other policies so that the content coming in to your organization fits your selection criteria for long-term support.

The selection process puts you on the path to a sustainable program. Selecting content is ultimately not a one-time project but a long-term, ongoing process, so formalizing it through policies, schedules and other documented criteria will help you avoid more log jams in the future.



This completes module 2, Select. If you are using these modules in order, the next one is module 3, Store. For additional resources on electronic records preservation and management, please visit the State Electronic Records Initiative webpage. This link is on your screen.