

# THE LIBRARY OF CONGRESS RECOMMENDED FORMAT SPECIFICATIONS

7/24/2014

Encouraging Preservation  
Without Discouraging Creation

# The Library of Congress has a mission

2

*“...to further the progress of knowledge  
and creativity for the benefit  
of the American people”*

Which means building a collection of content that  
can be preserved and made available  
for generations to come

# Identifying preservable content

3

The Library has relied traditionally on the Copyright Office's *Best Edition Statement* <http://www.copyright.gov/circs/circ07b.pdf> to determine the formats it wants

# The *Best Edition Statement* has two key drawbacks

4

- Much of its description of analog formats is dated
  - e.g. “cartridge rather than cassette”
- It says very little about digital formats at all
  - “Electronic Works Available Only Online” applies to serials only

We need a more current and comprehensive set of specifications to help us identify preservable content

# Recommended Format Specifications

5

The Library of Congress developed these to

- Serve as a set of hierarchies of the physical and technical characteristics of creative formats
- Apply to both analog and digital formats
- Seek to maximize the chances of survival and continued accessibility of that creative content well into the future

# Starting the process

6

The Library was able to take advantage of certain strengths

- Current documentation already in use at the Library
  - *Best Edition Statement*
  - Sustainability of Digital Formats guidelines
  - Collections Policy Statements
- Talented staff involved in the project bringing with them specialized knowledge in crucial areas
  - Technical aspects of preservation
  - Ongoing access needs of patrons
  - Developments in the marketplace and the publishing world

# Basic structure

7

We identified six basic categories of creative output

- Textual Works and Musical Compositions
- Still Image Works
- Audio Works
- Moving Image Works
- Software and Electronic Gaming and Learning
- Datasets/Databases

# Creating the hierarchies

- Most categories are broken down into subgroups (e.g. Photographs – digital)
- Within them are attributes which affect preservation and long-term access (e.g. formats)
- Each attribute has a list of the different options available (e.g. TIFF, JPEG2000, etc.)
- Two sets of these attributes and options are offered for most subsets: Preferred and Acceptable



# Recommended Format Specifications: caveats

- They complement the Copyright *Best Edition Statement*, they do not replace or supersede it
- They are not proscriptive, but acquiring and preserving content in unlisted formats will be challenging and costly
- They do not answer all questions raised in preservation and providing long-term access, e.g.
  - ▣ How to receive content into a repository
  - ▣ How to receive and use associated metadata
  - ▣ How to provide adequate inventory control
  - ▣ How to manage that content on an ongoing basis
  - ▣ How to provide access
- They merely help make it easier for those questions to be answered

# Recommended Format Specifications: internal focus

10

- The Library does not have the workflows or infrastructure in place as yet to acquire all listed formats – this will in some cases require custom design – but ...
- Staff can use the Specifications to inform potential acquisitions sources of formats the Library will prefer to collect, now or in the future
- Staff dealing with acquisitions sources can provide feedback to the Library on desired content currently available in the preferred formats
- And identify other formats in use of which the Library should be aware

# Recommended Format Specifications: external focus

11

- Use the Library's unique position and wealth of expertise to provide information useful for all stakeholders in the creative process
- Inform the creative, business, library and archival communities on best practices for ensuring the preservation of, and long-term access to, the creative output of the nation and the world
- Share the Specifications as widely as possible, so that all parties may take advantage of them

# Ongoing project

- The Recommended Format Specifications must keep up with ongoing changes in creative output
  - ▣ Input on them is requested from any and all stakeholders, e.g. publishers, producers, creative communities, vendors, libraries, etc.
  - ▣ Each spring, permanent technical teams at the Library will review and revise the six categories in light of all input, external and internal, they have received
  - ▣ As part of this, for 2015 the Library will focus on identifying preferred metadata schemas for inclusion
  - ▣ Each June 30<sup>th</sup>, the Library will issue the Recommended Format Specifications for the next twelve months

# Meet our common need

13

Encourage the creation, distribution and acquisition of content that will endure and remain accessible, which all involved in these processes need and from which all involved will benefit

# Questions?

14

[www.loc.gov/preservation/resources/rfs/](http://www.loc.gov/preservation/resources/rfs/)

General: Ted Westervelt, [thwe@loc.gov](mailto:thwe@loc.gov)

Textual Works, etc.: Ardie Bausenbach, [abau@loc.gov](mailto:abau@loc.gov)

Still Image Works: Phil Michel, [pmic@loc.gov](mailto:pmic@loc.gov)

Audio Works: Gene DeAnna, [edea@loc.gov](mailto:edea@loc.gov)

Moving Image Works: Mike Mashon, [mima@loc.gov](mailto:mima@loc.gov)

Software, etc.: Trevor Owens, [trow@loc.gov](mailto:trow@loc.gov)

Datasets/Databases: Donna Scanlon, [dscanlon@loc.gov](mailto:dscanlon@loc.gov)