

1 

Transfer Guidance Revision Project: Identifying “Fit-for-Purpose” Formats.

Donald Chalfant.

Kevin DeVorse.

Louise Guenther.

Kathryn Nielsen.

2 

Presentation Overview.

- A little background.
- Fun facts.
- What We’ve Learned: Ten years with the existing transfer guidance for federal agencies.
- What are our goals?
- Technical White Papers.
- The new look.
- Metadata for transfer.
- Q&A.

3 

Electronic Formats and Naara’s Transfer Guidance.

- What are file formats?
- Why are formats important to Naara?
- What does the Transfer Guidance apply to?
- How do we determine what things are?

4 

Fun Format Facts.

- Bytes 2-3 The second word of the file is the TIFF “version number.” This number, 42 (2A in hex), is not to be equated with the current Revision of the TIFF specification. In fact, the TIFF version number (42) has never changed, and probably never will. If it ever does, it means that TIFF has changed in some way so radical that a TIFF reader should give up immediately. *The number 42 was chosen for its deep philosophical significance.* It can and should be used as additional verification that this is indeed a TIFF file.

5 

The developers had a sense of humor...

6 

Current Guidance: limited scope that does not address all record/format types.

7 

Issues with current guidance products.

- Demonstrate a preference for standards based formats that agencies do not always use.
- Require that agencies transform to acceptable formats prior to transfer.
- Have proven an obstacle to Naara and agencies.

8 

Goals for revised guidance.

- Provide clear, concise, and consistent direction to agencies regarding formats that are acceptable for use when transferring records to Naara.
- Develop a flexible and extensible framework that can adapt to future needs.
- Balance preference for open formats with the business needs of agencies.
-

9 

Goals for revised guidance continued.

- Match record behavior and performance to the correct file formats.
- Adapt to the change from tapes to ERA.
- Support a life-cycle approach from creation through to researcher access.

10 

Goals for revised guidance continued.

- Expand the types of formats that Naara accepts.
- Acknowledge formats that are ubiquitous in the market place.
- Minimize the need for agencies to transform records prior to transfer.

11 

Where are we now?

- Revision project is almost ready for internal review.
- We’ve identified the categories and file formats that agencies are using today.

- We've conducted analysis of about 50 formats.
- We are compiling the results in the form of a revised guidance product.
- We are developing minimal metadata guidance.

12  **A Change in Approach.**


- We will identify formats that are:
 - Preferred .
 - Acceptable.
 - Acceptable for Imminent Transfer (to sunset previously acceptable formats).
 - Not Acceptable (this will be rare).

13  **Categories of E-records.**

- Digital Still Images.
- Digital Moving Images.
- Digital Audio.
- Text.
- Geospatial Records.
- CAD.
- Structured Data.
- E-mail.
- Web & Social Media.

14  **The importance of sustainability: Graphics Interchange Format (GIF).**

- Introduced by CompuServe in 1987.
- Unisys exerted patent claims against LZW compression in 1994.
- Alternate formats such as JPEG and PNG developed.
- Patent claims expired in 2003.
- In 2012, GIF was recognized as a verb.

15  **Measuring sustainability.**

- Disclosure: the degree to which complete specifications and technical integrity tools exist.
- Adoption: the degree to which the format is used by creators, disseminators, or users.
- Transparency: the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor.
- Self-documentation: formats that contain all the metadata needed to render the data as usable information.
- External dependencies: refers to the degree to which a format depends on particular hardware, operating system, or software for rendering or use.
- Impact of patents: Patents related to a digital format may inhibit the ability of archival institutions to sustain content in that format.
- Technical protection mechanisms: To preserve digital content and provide service to users and designated communities decades hence, Naara must be able to replicate the content on new media, migrate and normalize it in the face of changing technology, and disseminate it to researchers.

16  **Measuring Sustainability.**

- Analyzing format resources.
 - ISO/ANSI and other specifications.
 - Corporate sites.
 - LC Format Sustainability of Digital Formats.
 - TNA-UK PRONOM Format Registry.
 - CDL/LC UDFR.
 - Wikipedia.


17  **Determining “Fit for Purpose” Formats.**

- The conventional wisdom is not always correct.
- Open formats may not always be the best choice.
- Proprietary formats aren't always so bad.

18  **Metadata for Transfer**

- The types of electronic records in use have changed and so have our needs.

- There are more metadata rich formats.
- We currently ask for indexes and system documentation.
- More and more this exists as metadata that we can use to help automate processes.

19  **What are you using?**

- We would like to work with you to identify:
 - Existing standards that you are using.
 - Your capabilities to export metadata that you use to search and retrieve e-records.
 - A “simplistic” set of metadata appropriate for transfer.

20 

Kevin De Vorsey
Electronic Records Format Specialist
Policy Analysis and Enforcement
Office of the Chief Records Officer
Agency Services
National Archives and Records Administration

kevin.devorsey@Naara.gov
212-401-1631