

# Electronic Records Transfer Guidance at NARA

WHAT IS PAST  
IS PROLOGUE

Donald Chalfant & Kevin De Vorse

[donald.chalfant@nara.gov](mailto:donald.chalfant@nara.gov)

[kevin.devorse@nara.gov](mailto:kevin.devorse@nara.gov)

# Current Guidance (Developed Between 2002-2004)

THE U.S. NATIONAL ARCHIVES & RECORDS ADMINISTRATION

[www.archives.gov](http://www.archives.gov)

Printed on: Friday, October 22, 2010

[Code of Federal Regulations]

[Title 36, Volume 3]

[Revised as of July 1, 2003]

From the U.S. Government Printing Office via GPO Access

[CITE: 36CFR1228.270]

[Page 725-728]

TITLE 36--PARKS, FORESTS, AND PUBLIC PROPERTY

CHAPTER XII--NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

PART 1228--DISPOSITION OF FEDERAL RECORDS--Table of Contents

Subpart L--Transfer of Records to the National Archives of the United States

## Sec. 1228.270 Electronic records.

(a) Timing of transfers. Each agency is responsible for the integrity of the permanent records it transfers on physical media to the National Archives of the United States. For records transferred by a media-less method, NARA works with the agency to ensure integrity of the records during the transfer process. To ensure that permanent electronic records are preserved, each Federal agency must transfer electronic records to NARA promptly in accordance with the agency's records disposition schedule. Furthermore, if the agency cannot provide proper care and handling of the media (see part 1234 of this chapter), or if the media are becoming obsolete and the agency cannot migrate the records to newer media, the agency must contact NARA to arrange for timely transfer of permanent electronic records, even when sooner than provided in the records schedule.

## FAQs in Response to the What's & Why's of Electronic Formats NARA Accepts

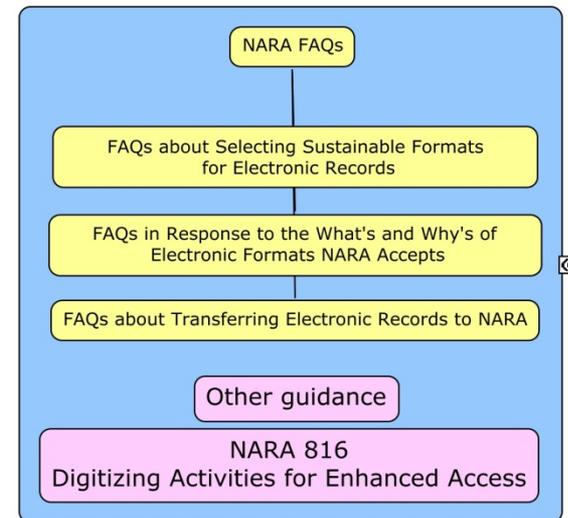
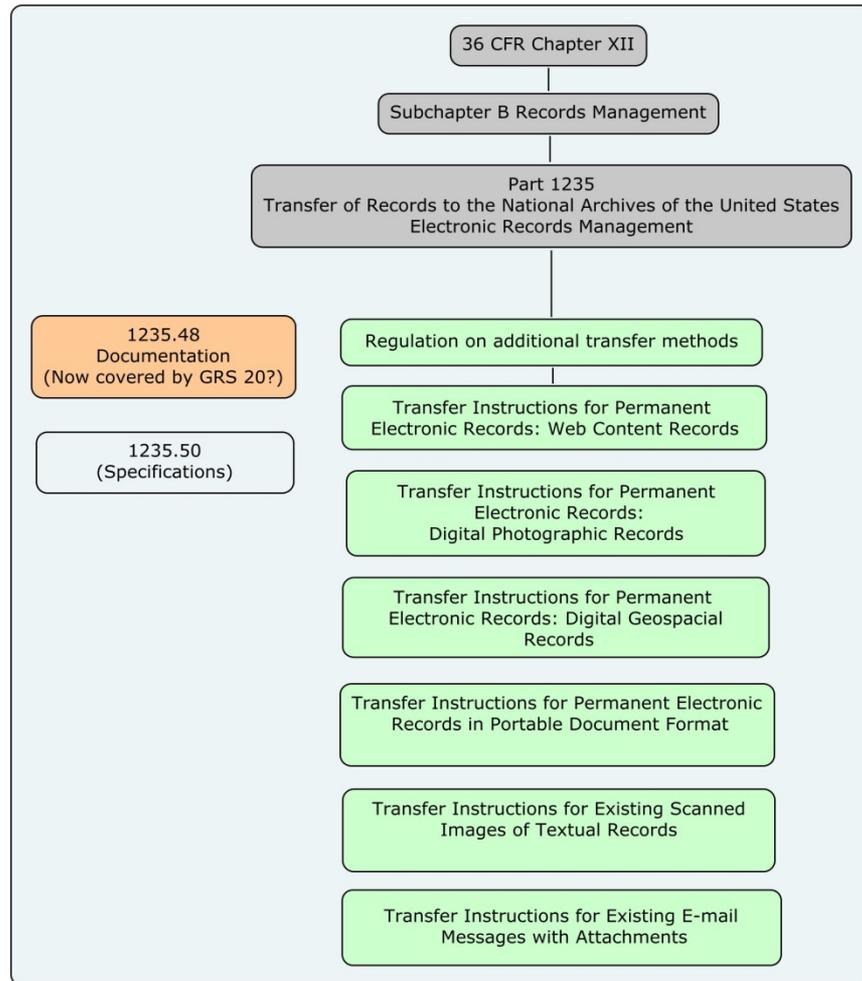
### *What formats of electronic records does NARA accept?*

According to its regulations (36 CFR 1228.270(d)) and transfer guidance developed during the Electronic Records Management [ERM] E-Government Initiative, NARA accepts over twenty formats (or versions thereof) of electronic records, ranging from digital photography to web pages to geographic information systems. In addition to the formats cited in the regulations, additional formats that NARA accepts can be found at <http://www.archives.gov/records-mgmt/initiatives/erm-products.html>.

NARA is currently developing an Electronic Records Archives [ERA] system that will allow it to accept even more formats of electronic records in the future.

## Reflect NARA's capabilities at the time.

# Current Guidance: limited scope that does not address all format types



# Current Guidance Products

- Demonstrate a preference for open, standards-based formats.
- Require that agencies transform or normalize data into acceptable formats prior to transfer\*.
- Have proven an obstacle to the steady transfer of records.
- Are referred to at many points of the lifecycle.

# Project Scope

- In scope:
  - This project seeks to support the work of federal agencies by providing **flexible and realistic electronic records file format guidance** on all electronic records types for use when transferring permanent records to NARA in accordance with the Federal Records Act.
  - This project **will identify and recommend changes** but the execution of any additional guidance including guidance for all types of metadata as well as the revision of business processes, or development of standard operating procedures is beyond the scope of this project.
- Out of scope:
  - Format guidance for other areas of the record lifecycle other than transfer to NARA
  - Guidance on physical media
  - Records of the Executive Office of the President and the records of the United States Congress. These branches are not covered by the Federal Records Act and are therefore excluded from consideration for this project

# Revised Guidance Content Categories

- Electronic Textual Records
- Digital Still Images
- Digital Audio Records
- Digital Moving Image Records
- *Structured Data*
- *Geospatial Records*
- *CAD and Vector Graphics*
- *Web Records*
- E-mail Records

# Relevant Content Categories Definitions

- **Structured Data** – includes the broad category of data that is stored in defined fields and includes:
  - **Databases** – Database formats are organized collections of associated data that conform to a logical structure. Database formats are determined by “data models” that describe specific data structures used to model an application and generally include navigational, relational, and hybrid models.
  - **Spreadsheets** – Spreadsheets are electronic simulations of paper accounting worksheets for financial plans, budgets, etc. Personal computer and server based spreadsheet programs [e.g. Microsoft Excel, Lotus 1-2-3, Open Office Calc.] can create both proprietary files as well as software independent files including text or XML. Cloud-based spreadsheets [e.g. Google Documents] include format export options such as .xls, .csv, .txt, .ods, PDF and HTML files as well as import and conversion options for common spreadsheet formats including .xls, .csv, and .ods.
  - **Statistical Data** – Statistical Data is the result of scientific quantitative research and analyses. Statistical data formats contain collections of data presented in both tabular and non-tabular form. Datasets are formatted as strings of characters contained within a markup language [e.g. XML] or as software dependent proprietary files by commercial statistical and qualitative data analysis software tools (e.g. SAS and SPSS).
  - **Scientific data** refers to research data collected by instrumentation tools during the scientific process. Scientific data formats are either domain specific such as those used within a single field of study [e.g. Flexible Image Transport System (FITS)] or are multi-domain formats useful for transfer of scientific data between domains [e.g. Common Data Format (CDF), HDF5].

# Relevant Content Categories Definitions

- **Geospatial** – Geospatial data includes files created by geographic information systems (GIS) or other software applications for spatial analysis using computer systems. The data may be contained within a database to enable analysis across the datasets (e.g. geo-database), united within a complex file format structure where one geospatial file is comprised of several distinct, but related, formats (e.g. shapefile), or contained within a single file (e.g. GeoPDF).
- **Computer Aided Design (CAD) and Vector Graphics**– Non-raster Vector graphics formats use mathematical expressions to create and manipulate computer graphics and animations. Computer Aided Design (CAD) are vector programs used in engineering and manufacturing design to create animations and represent three-dimensional surfaces of inanimate objects. CAD and Vector graphics programs can output binary and XML formats.

# Project Phases

- **Phase 1: Planning and Preparation**
  - July 1– December 30, 2011
- **Phase 2: Conduct Informational Meetings**
  - February 6– August 12, 2012
  - Internal NARA SMEs
  - *Future Perfect Conference*
  - Agency representatives
- **Phase 3: Develop and Publish Guidance Product**
  - May 29 – September 7, 2012
- **Phase 4: Evaluation and Completion**
  - December 12 - December 21, 2012

# Considerations\*

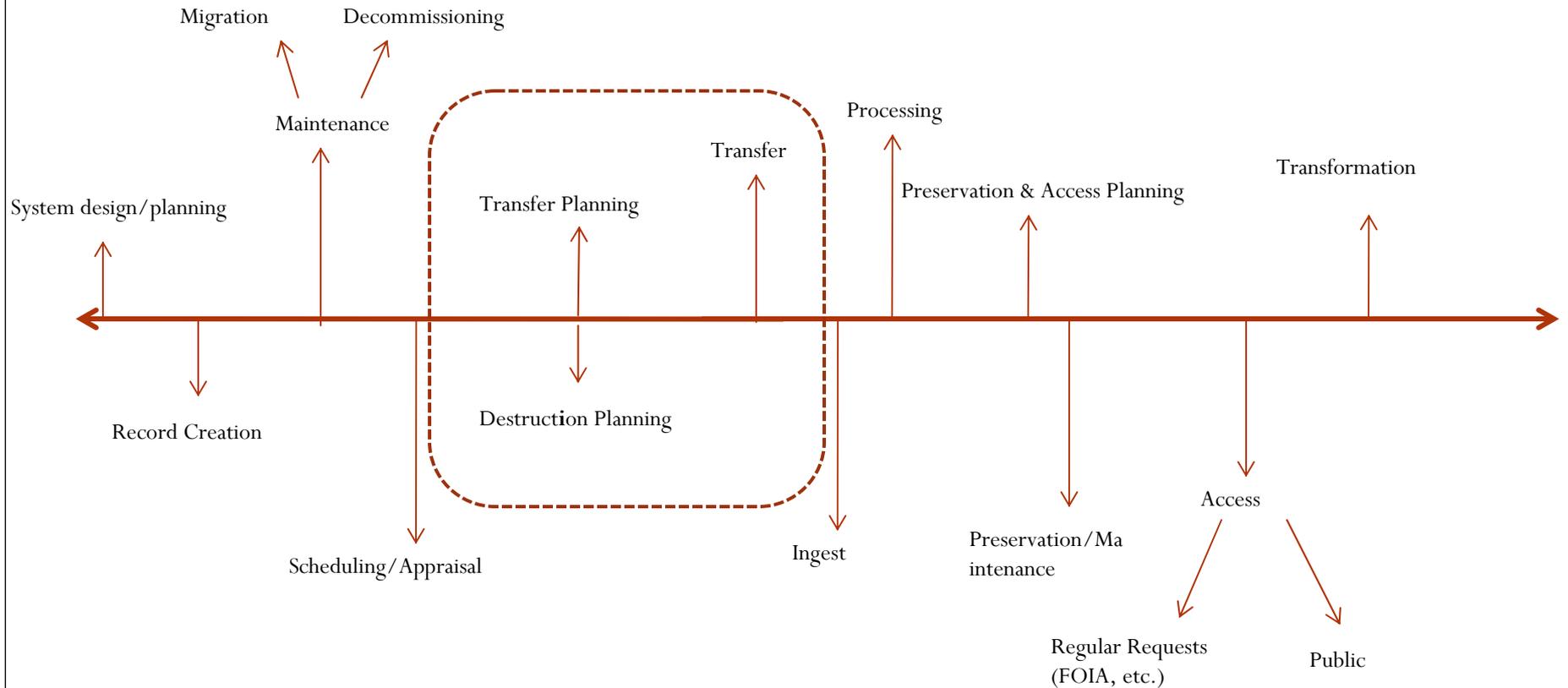
- What part(s) of the system represents the record?
  - Do we want to bring in the entire system?
  - If we only want a subset or can only accept an export then what is the “best” format for the electronic record type in question?
- Could ERA cope with the formats, file size, and/or volume or files?
- What additional information should accompany the data?
- How should we validate and verify this data?

\*These influence the transfer guidance but changes to existing work processes are out of the scope of this project.

# Goals

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

# Electronic Records Lifecycle



# Stress 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, NARA 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.

\*adapted from <http://www.digitalpreservation.gov/formats/>

# Concluding Thoughts

- The transfer guidance revision project will:
  - expand the types of formats that NARA accepts
  - balance the business requirements of agencies with NARA's preservation and access needs
  - minimize the need for agencies to transform records prior to transfer
  - Identify gaps that encourage development of guidance across the lifecycle of electronic records to support digital continuity

# Thank You!

Donald Chalfant & Kevin L. De Vorse

Electronic Records Format Specialists

Electronic Records Format Section

Policy Analysis and Enforcement Division

Office of the Chief Records Officer

Agency Services

National Archive and Records Administration

[donald.chalfant@nara.gov](mailto:donald.chalfant@nara.gov)

[kevin.devorse@nara.gov](mailto:kevin.devorse@nara.gov)