

General Considerations for Choosing File Formats

Andrea Goethals, Harvard University Library

Last modified: 07/31/09

Unfortunately it's not possible to provide a single list of file formats that are appropriate for all use cases. The best formats to use for each of an object's files will vary depending on how those files will be accessed and used, how they were captured or created, and their relationship to other files in the object.

When requested, OIS staff will provide guidance on the digital file formats to use for a project if the content will be deposited into the DRS. The factors that are taken into consideration include:

1. the intended use of the file for delivery and transformations
 - Is this intended to be a deliverable / use copy (either now or in the future)?
 - If so, who are the expected users of this content and what software is expected to deliver and render this content to these users? This software will need to support the format. Ideally the format should be also be supported by software known to be next-generation versions of the software users are currently using.
 - Is it possible that this file will be the source for generating future deliverables?
 - If so, the format will need to support all current and foreseeable uses of the content; the format should be well-supported by readily-accessible transformation software.
 - Is it possible that this file will be the source for generating future preservation copies?
 - If so, the format should support lossless transformations; the format needs to be capable of supporting all functionality needed by all uses of the content; the format should not be in any danger of becoming obsolete in the foreseeable future; the format specification should be readily accessible and well-understood; the format should be well-supported by readily-accessible transformation software.
2. the nature of the original capture or creation
 - Are you starting from already created/acquired digital files? For example, are you starting from digital files generated from a sound recorder?
 - If so, what are the formats and technical characteristics of these files? Any derivative formats will need to use these files as the transformation source. The metadata in the resulting files will be constrained by the metadata supported by the source files.
 - Is this the digital file closest to the original capture (camera, scanner, etc.) or creation (software application)?
 - If so, the format should have the same characteristics listed above under the heading "Is it possible that this file will be the source for generating future preservation copies?".
3. its relationship to other files
 - Are other files dependent on this file? For example, do other files link to this one? Does the delivery/rendering software expect the dependent files to be in particular formats?
 - Is this file the only deliverable file for the object?
 - If not, there can be more leniency in choosing a deliverable format reliant on uncommon software, as long as another deliverable file for the object is supported by common software.
 - Whether or not the file is intended to be delivered, is this file the only representation of some content (image, text, etc.) for the object, or does this content exist in other file formats deposited along with the object?

- If it is the only representation of the content it should have the same characteristics listed above under the heading "Is it possible that this file will be the source for generating future preservation copies?". If the format is not already supported by appropriate delivery/rendering applications consider generating additional representations of the content in deliverable formats.