Still Image File Format Comparison

Document III. Detailed Matrix

Version of July 30, 2013
For review by the FADGI Still Image Working Group

Background. The two FADGI Working Groups are exploring file formats for still images and video. The explorations are using similar, matrix-based tools to make comparisons relevant to preservation planning. The matrixes compare a limited number of formats in terms of roughly forty factors, grouped under the following general headings:

- Sustainability Factors
- Cost Factors
- System Implementation Factors (Full Lifecycle)
- Settings and Capabilities (Quality and Functionality Factors)

The still image effort is led by the Government Printing Office and it is comparing formats suitable for reformatting (digitization). The formats being compared include JPEG 2000, JPEG (DCT), TIFF, PNG, and PDF, and several subtypes. The findings from this project will be integrated into the Working Group's continuing refinement of its general guideline for raster imaging.

Other documents in the set

Document I - Narrative Introduction

Introduction: File Format Sub-Group
Guiding Principles and Selection of File Formats
Sub-Group Deliverables: Summary Table and Detailed Matrix
Findings and Next Steps

Document II. Summary Table

Category T: Textual and illustrated printed matter (books, journals, manuscripts, some maps).

ATTRIBUTES	Scoring Conventions	FORMAT: TIFF				FORMAT: JPEG 2000		FORMAT: JPEG	FORMAT: PNG	FORMAT: PDF		
		Common TIFF, Uncompressed	Common TIFF, Lossless Compressed	GeoTIFF/BigTIFF, Uncompressed	GeoTIFF/BigTIFF, Compressed	JPEG 2000: JP2	JPEG 2000: JPX	JPEG (JFIF with EXIF)	PNG	PDF (1.1-1.7)	PDF/A (1, 1a, 1b, 2)	GeoPDF*
Sustainability Factors												
Disclosure	Good Acceptable Poor	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
Adoption	Wide Adoption Moderate Adoption Limited Adoption	Wide Adoption	Wide Adoption	Wide Adoption (adoption tends to be limited to geospatial communities, but is widely adopted there)	Wide Adoption (adoption tends to be limited to geospatial communities, but is widely adopted there)	Wide Adoption (caveat: moderate adoption in cultural heritage community, but widely adopted in other communities. Not widely supported in browsers and cameras)	Moderate Adoption (some adopt JPX but it is not as adopted as core coding)	Wide Adoption (adoption is very high, ubiquitous)	Wide Adoption	Wide Adoption	Wide Adoption	Wide Adoption
Transparency	Good Acceptable Poor	Good	Acceptable (added layer of encoding due to compression)	Good	Acceptable (added layer of encoding due to compression)	Acceptable (encoded due to compression, compensated for by resiliency elements which are meant to overcome perceived transparency issues. Resiliency elements are listed on the digitalpreservation.gov website)	Acceptable (encoded due to compression, compensated for by resiliency elements which are meant to overcome perceived transparency issues. Resiliency elements are listed on the digitalpreservation.gov website)	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable
Self-Documentation	Good Acceptable Poor	Acceptable	Acceptable	Acceptable	Acceptable	Good (more capabilities for entering metadata)	Good (more capabilities for entering metadata)	Acceptable	Acceptable	Good	Good	Good
Native Embedded metadata capabilities	Good Acceptable Poor	Acceptable (limited to header tags)	Acceptable (limited to header tags)	Acceptable (limited to header tags)	Acceptable (limited to header tags)	Good (open and extensible)	Good (open and extensible)	Acceptable (limited to technical metadata, not descriptive metadata)	Minimal	Good	Good	Good
Embedded metadata capabilities through extension	Good Acceptable Poor	Good (XMP)	Good (XMP)	Good (Extended TIFF header elements are generally used rather than XMP)	Good (Extended TIFF header elements are generally used rather than XMP)	Good (open and extensible)	Good (open and extensible)	Good (XMP for descriptive and EXIF for technical information such as camera, shutter speed, etc.)	Not supported	Good	Good	Good
Level of Work necessary to embed native metadata	High Medium Low	Low (header tags)	Low (header tags)	Low (header tags)	Low (header tags)	Low (caveats: more time and effort may be required due to learning curve and available tools. Main obstacle is the format that metadata needs to adhere to, not inherent in the file format itself. There may be a need to establish your own specification for metadata)	Low (caveats: more time and effort may be required due to learning curve and available tools. Main obstacle is the format that metadata needs to adhere to, not inherent in the file format itself. There may be a need to establish your own specification for metadata)	Low	Low (tool may be required)	Low (tool may be required)	Low (tool may be required)	Low (tool may be required)
Level of Work necessary to embed metadata through extension	High Medium Low	Low (XMP)	Low (XMP)	Low (Extended TIFF header elements are generally used rather than XMP)	Low (Extended TIFF header elements are generally used rather than XMP)	No additional work required due to open and extensive native capability	No additional work required due to open and extensive native capability	Low (caveats: more time and effort may be required due to learning curve and available tools. Main obstacle is the format that metadata needs to adhere to, not inherent in the file format itself. There may be a need to establish your own specification for metadata)	Very Low	Low (XMP)	Low (XMP)	Low (XMP)
Geo-referencing Metadata	Good Acceptable Poor	Not supported	Not supported	Good	Good	Not supported (see JPX)	Good (OGC GMLJP2 specification available to handle this)	Limited grid coordinate data may be held in EXIF data. Richer GIS data provided by sidecar "world file" (jgw extension) supported by some applications.	Not supported	Not supported	Not supported	Good
Level of effort to embed geo-referencing metadata	High Medium Low	N/A (GIS data can be provided by sidecar 'world file' (tfw extension) supported by some applications.)	N/A (GIS data can be provided by sidecar 'world file' (tfw extension) supported by some applications.)	Low (open source tools)	Low (open source tools)	N/A	Low-medium (tools available to embed GML data)	Low (tools available in GIS software)	N/A	N/A	N/A	Low
Impact of Patents	Possible Impact No Impact	No Impact	No Impact (Patents on LZW compression have expired, alleviating a concern)	No Impact	Low Impact (Patents on LZW compression have expired, alleviating a concern)	Little or No Impact	Possible Impact (some patents may apply--impacts are not clear: see: http://www.digitalpreservation.gov/formats/fdd/fdd000141.shtml)	No Impact	No Impact	No Impact	No Impact	No Impact
Technical Protection Mechanisms	Possible Impact No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact (protection mechanisms are available but not required and not a deterent from choosing this format)	No Impact (protection mechanisms are available but not required and not a deterent from choosing this format)	No Impact (protection mechanisms are available but not required and not a deterent from choosing this format)
Cost Factors												
Implementation Cost	High Medium Low	Low	Low	Low	Low	Medium-High	Medium-High (may require added geo-referencing tool)	Low	Low	Medium-high (tools can be expensive)	Medium-high (tools can be expensive)	Medium-high (tools can be expensive)

* NOTE: this refers to either of two formats:

TerraGo GeoPDF
Adobe Geospatial PDF

Category T: Textual and illustrated printed matter (books, journals, manuscripts, some maps).

Cost of software tools	High Medium Low	Low	Low	Low	Low	Medium-High (best toolsets avialable currently are proprietary tools. Open source tools are not yet mature)	Medium-High (best toolsets avialable currently are proprietary tools. Open source tools are not yet mature)	Low	Low	Medium-High (best toolsets available currently for this use case are proprietary tools)	Medium-High (best toolsets available currently for this use case are proprietary tools)	Medium-High (best toolsets available currently for this use case are proprietary tools)
Cost of equipment needed to produce files	High Medium Low	Low	Low	Low	Low	Low-Medium	Low-Medium	Low	Low	Low-Medium	Low-Medium	Low-Medium
Storage Cost	High Medium Low	High (could do a sample test set of files to determine this?)	Medium for LZW on tonal images (NOTE: LZW on high-bit images will increase the size and therefore the storage footprint/cost) Low for bitonal with group 4	High (could do a sample test set of files to determine this)	Medium for LZW on tonal images Low for bitonal with group 4 (unlikely scenario)	**LOW**	**LOW**	Low-medium	Medium	Low (you would generally use PDF in cases where you could take advantage of compression)	Low (you would generally use PDF in cases where you could take advantage of compression)	Low (you would generally use PDF in cases where you could take advantage of compression)
Network Cost	High Medium Low	High (could do a sample test set of files to determine this?)	Medium for LZW on tonal images Low for bitonal with group 4	High (could do a sample test set of files to determine this)	Medium for LZW on tonal images Low for bitonal with group 4 (unlikely scenario)	LOW	LOW	Low-medium	Medium	Low (you would generally use PDF in cases where you could take advantage of compression)	Low (you would generally use PDF in cases where you could take advantage of compression)	Low (you would generally use PDF in cases where you could take advantage of compression)
Ongoing Cost of Production	High Medium Low	Medium-High	Medium	Medium-High	Medium	Low-Medium	Low-Medium	Low-medium	Medium	Medium (longer post process. could vary greatly dependent on original and number of pages, etc.)	Medium (longer post process. could vary greatly dependent on original and number of pages, etc.)	Medium (longer post process. could vary greatly dependent on original and number of pages, etc.)
Cost of Providing Access	High Medium Low	High	Medium-High (people working with CMS will be ok, but people with general image viewers may have issues)	High	Medium-High (people working with CMS will be ok, but people with general image viewers may have issues)	Medium-High (potential is there, but is not quite there yet)	Medium-High (potential is there, but is not quite there yet)	Low	Low	Low	Low	Low
Cost of Preservation Processing	High Medium Low	Medium (assumption is that raster easiliy available for migration processing)	Medium (assumption is that raster easiliy available for migration processing)	Medium (assumption is that raster easiliy available for migration processing)	Medium (assumption is that raster easiliy available for migration processing)	Medium	Medium (caveat: if your profile is known, it would be the same level as JP2, but if not, the cost may be higher)	Low	Low	Medium (could vary based on complexity)	Medium (could vary based on complexity)	Medium (could vary based on complexity)
System Implementation Factors (Full Lifecycle)												
Level of difficulty/complexity	High Medium Low	Low	Low	Low	Low	Medium-high	Medium-high	Low	Low	Medium (could vary	Medium (could vary)	Medium (could vary)
Technical Complexity	High Medium Low	Low	Low	Low	Low	Medium-high	Medium-high	Low	Low	Medium (could vary	Medium (could vary)	Medium (could vary)
Toolset Complexity	High Medium Low	Low	Low	Medium	Medium	Medium-high	Medium-high	Low	Low	Low	Low	Low
Availability of tools	Wide availability Moderate availability Limited availability	Wide Availability	Wide Availability	Moderate Availability	Moderate Availability	Limited to Moderate Availability	Limited to Moderate Availability	Wide Availability	Wide Availability	Wide Availability	Wide Availability	Wide Availability
Compatability in existing enterprise environment (e.g., OCR-ability, quality review)	Good Acceptable Poor	Good	Good	Acceptable	Acceptable	Acceptable to Poor	Acceptable to Poor	Good	????	Good	Good	Good
Ease and accuracy of File validation	Good Acceptable Poor	Good	Good	Good	Good	Good	Good (not clear about validating geo-referencing metadata)	Good	Acceptable	Good	Good	Good
Evaluating and Monitoring of Quality	Good Acceptable Poor	Good	Good	Good	Good	Good	Good (not clear about validating geo-referencing metadata)	Good	Acceptable	Good	Good	Good
Settings and Capabilities (Pass/Fail)												
Clarity	Pass Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass (DCT has lower level of clarity than DWT; and 8-bit has lower level of clarity than 16 bit)	Pass	Pass (for cetain categories of material, we would want a greater bit depth)	Pass (for cetain categories of material, we would want a greater bit depth)	Pass (for cetain categories of material, we would want a greater bit depth)
Support for Color Maintenance (could have a grading scale for some of these factors)	Good Acceptable Poor	Good (caveat: to insert an ICC profile or declare certain color spaces, you must use an "extended tag set")	Good (caveat: to insert an ICC profile or declare certain color spaces, you must use an "extended tag set")	Good (caveat: to insert an ICC profile or declare certain color spaces, you must use an "extended tag set")	Good (caveat: to insert an ICC profile or declare certain color spaces, you must use an "extended tag set")	Good (good but not perfect documentation of color space. Standards group working on these)	Good (better documentation of color space than JP2)	Good (Requires EXIF or other extension for embedding ICC profile. EXIF version is preferred for JPEG)	Good	Good	Good	Good
Searchable Text Embedding	Pass Fail	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Pass	Pass	Pass

Category T: Textual and illustrated printed matter (books, journals, manuscripts, some maps).

Multi-Page Capability	Pass Fail	Pass	Pass	Pass	Pass	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Fail (Not natively supported)	Pass	Pass	Pass
Notes on Maximum File Size	Actual data on maximum file sizes	Up to 4GB	Up to 4GB	GEO TIFF: up to 4GB BigTIFF: up to 18,000 petabytes Like TIFF format, GeoTIFF uses 32-bit offsets, thus limiting its extent to 4 gigabytes. The needs of GIS, large format scanners, medical imaging and other fields have prompted development of the variant BigTIFF format, which transcends the 4 GB TIFF limit using 64-bit offsets thereby supporting files up to 18,000 petabytes in size.	GEO TIFF: up to 4GB BigTIFF: up to 18,000 petabytes Like TIFF format, GeoTIFF uses 32-bit offsets, thus limiting its extent to 4 gigabytes. The needs of GIS, large format scanners, medical imaging and other fields have prompted development of the variant BigTIFF format, which transcends the 4 GB TIFF limit using 64-bit offsets thereby supporting files up to 18,000 petabytes in size.	Practical limits may arise depending on application and/or pixel count (may be limited to 537 megapixels)	Practical limits may arise depending on application and/or pixel count (may be limited to 537 megapixels)	Practical limits may arise depending on application and/or pixel count	Practical limits may arise depending on application and/or pixel count	Generally accepted practical limit is 2GB, based on reader applications	Generally accepted practical limit is 2GB, based on reader applications	Generally accepted practical limit is 2GB, based on reader applications