Federal Agencies
Digitization Guidelines Initiative

Raster Still Images for Digitization
A Comparison of File Formats

Part 1. Detailed Matrix (unified large table)

This table presents all of the information in a single table to facilitate comparisons. All pages after this cover are intended for printing on 11x17-inch paper. Part 2 provides the same information on multiple, easily printable pages.

Revised, September 2, 2014

The FADGI Still Image Working Group
http://www.digitizationguidelines.gov/still-image/

Raster Still Images for Digitization: A Comparison of File Formats

ATTRIBUTES	Scoring Conven-tions	Questions to Consider	FORMAT: TIFF				FORMAT: JPEG 2000		FORMAT: JPEG	FORMAT: PNG	FORMAT: PDF * NOTE: in this comparison, GeoPDF refers to both TerraGo GeoPDF and Adobe Geospatial 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	Does complete technical documentation exist for this format? Is the format a standard (e.g., ISO)? Are source code for associated rendering software, validation tools, and software development kits widely available for this format?	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
Adoption	Wide Adoption Moderate Adoption Limited Adoption	Is this format likely to become obsolete short, medium, or long-term? How widely adopted is the format in the digitization vendor community? Are there software tools available around this format? Are there user communitys/developer communities that are actively discussing the format and its further development?	Wide Adoption. Negligible support in browsers.	Wide Adoption. Negligible support in browsers.	Wide Adoption (Negligible support in browsers, adoption tends to be limited to geospatial communities, but is widely adopted there)	Wide Adoption (Negligible support in browsers, adoption tends to be limited to geospatial communities, but is widely adopted there)	Adoption varies: moderate for still images in cultural heritage community, wide for moving image content in production and archiving. Limited support in still image software, negligible support in browsers and still cameras.	Low to moderate adoption, less uptake than JP2 core coding.	Wide Adoption (adoption is very high, ubiquitous)	Wide Adoption	Wide Adoption	Wide Adoption	Wide Adoption
Transparency	Good Acceptable Poor	Is it a linear bitmap or is it more complex (e.g., compression). Need to consider ability of each format to compensate for lack of transparency. What is the impact of having many options and potentially complex implementations?	Good	Acceptable (added layer of encoding due to compression)	Good	Acceptable (added layer of encoding due to compression)	Acceptable. Compression is compensated for by resiliency elements, intended to mitigate low levels of transparency. However, the format offers many options (tiling, quality layers, progression order, more), and some users have found that "legal" variations may not interoperate from one application to another.	Acceptable. Compression is compensated for by resiliency elements, intended to mitigate low levels of transparency. However, the format offers many options (tiling, quality layers, progression order, more), and some users have found that "legal" variations may not interoperate from one application to another.	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable
Self-Documentation	Good Acceptable Poor	Does the technical metadata, typically in a header or equivalent, fully describe the characteristics of the file/file format? Does the format offer capabilities for descriptive metadata (aka "cataloging" or "about" metadata) that provide a reasonable level of information about the content within the file?	Acceptable	Acceptable	Acceptable	Acceptable	Good (includes Intellectual Property, XML, URL, and UUID metadata boxes)	Good (includes Image Creation, Content Description, History, Intellectual Property Rights, and Image Identifier metadata boxes)	Acceptable	Good	Good	Good	Good
Native Embedded metadata capabilities	Good Acceptable Poor	How well does this format support embedded metadata, including headers, as a part the format's own specifications (native metadata)?	Acceptable (limited to header tags)	Acceptable (limited to header tags)	Acceptable (limited to header tags)	Acceptable (limited to header tags)	Good (open and extensible, supports inclusion of user defined metadata and vendor specific metadata)	Good (open and extensible)	Acceptable (limited to technical metadata, not descriptive metadata)	Good	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., requires a compliant reader)	Good (XMP)	Good	Good	Good
Level of Work necessary to embed native metadata	High Medium Low	What level of effort is required to embed native metadata?	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	Low	Low	Low
Level of Work necessary to embed metadata through extension	High Medium Low	How well does this format support forms of embedded metadata that are not part of the format's own specification (metadata defined by extension)?	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)	Low to medium; (not all readers and writers support all metadata features)	Medium (not all readers and writers support all metadata features)	Low (XMP)	Low (XMP)	Low (XMP)	Low (XMP)	Low (XMP)
Geo-referencing Metadata	Good Acceptable Poor	How well does this format support embedded geo-referencing metadata?	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; TerraGo geo display functionality may be limited to Windows app
Level of effort to embed geo-referencing metadata	High Medium Low	What level of effort is required to embed extension metadata?	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	Are there patents related to this format that could have a direct impact on the long-term sustainability of files produced in this format?	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)	No Impact	No Impact	No Impact	No Impact	No Impact

Raster Still Images for Digitization: A Comparison of File Formats

ATTRIBUTES	Scoring Conventions	Questions to Consider	FORMAT: TIFF				FORMAT: JPEG 2000		FORMAT: JPEG	FORMAT: PNG	FORMAT: PDF * NOTE: in this comparison, GeoPDF refers to both TerraGo GeoPDF and Adobe Geospatial 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*
Technical Protection Mechanisms	Possible Impact No Impact	Are there technical protection measures inherenet to this format that would prohibit the creation of ample derviatives/other formats?	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact	No Impact (protection mechanisms available, not required; this factor would not deter users from selecting this format for scanned raster images in reformatting projects)	No Impact (protection mechanisms available, not required; this factor would not deter users from selecting this format for scanned raster images in reformatting projects)	No Impact (protection mechanisms available, not required; this factor would not deter users from selecting this format for scanned raster images in reformatting projects)
Cost Factors													
Implementation Cost	High Medium Low	Software/capture Software/deliver IT support [staff] Startup (training, support, expertise)	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)
Cost of software tools	High Medium Low		Low	Low	Low	Low	Medium-High (best toolsets avialable currently are proprietary; open source tools are not yet mature)	Medium-High (best toolsets avialable currently are proprietary; 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 (computationally intense compression)	Low-Medium (computationally intense compression)	Low	Low	Low-Medium	Low-Medium	Low-Medium
Storage Cost	High Medium Low	Are files created in this format usually large, medium, or small in size? (The values assigned in this category are especially rough-and-ready.	High	Medium for LZW on tonal images (NOTE: LZW on high-bit or pictorial images will increase the size and therefore the storage footprint/cost) Low for bitonal with group 4	High	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	Does the transfer of files in this format affect performance of internal networks to the point where it would cost more to implement this format? File transfer for ingest into archive, transfer to "working area" for processing and access derivative creation.	High	Medium for LZW on tonal images Low for bitonal with group 4	High	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	Scanner speed/file transformation and compression How many scans per hour can be accomplished? CPU usage calculations to produce derivatives?	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	Medium (derivatives needed) Low (copy of master serves access)	Are derivatives necessary in order to provide broad access?	Medium	Medium	Medium	Medium	Medium	Medium	Low	Low	Low	Low	Low
Cost of Preservation Processing	High Medium Low	Costs in relation to emulation, migration, etc. File integrity monitoring (bit level preservation, etc.) Tools that are needed to execute migration, emulation. Are there tools that are available that are cheap or free, or will there be custom development or large investment necessary?	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	What is the level of effort associated with the implementation of this format? Are there special requirements for this format that would change the nominal workflow for digitization/information life cycle? Cost of applications, software, etc.	Low	Low	Low	Low	Medium-high	Medium-high	Low	Low	Medium (could vary	Medium (could vary)	Medium (could vary)
Technical Complexity	High Medium Low	This is about the complexity of the implementation.	Low	Low	Low	Low	Medium-high	Medium-high	Low	Low	Medium (could vary	Medium (could vary)	Medium (could vary)
Toolset Complexity	High Medium Low	This factor relates to the level of difficulty/complexity of the toolsets avaialble to implement. Are there many or few applications that support the format?	Low	Low	Medium	Medium	Medium-high	Medium-high	Low	Low	Low	Low	Low

Raster Still Images for Digitization: A Comparison of File Formats

ATTRIBUTES	Scoring Conven-tions	Questions to Consider	FORMAT: TIFF				FORMAT: JPEG 2000		FORMAT: JPEG	FORMAT: PNG	FORMAT: PDF * NOTE: in this comparison, GeoPDF refers to both TerraGo GeoPDF and Adobe Geospatial 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*
Availability of tools	Wide availability Moderate availability Limited availability	Are there tools available for this format? Are the tools open source? Are tools reliable when creating files that precisely meet the format specification? If a future digital archeologist had the format specification, how easy would it be to write an application?	Wide Availability	Wide Availability	Moderate Availability	Moderate Availability	Limited to Moderate Availability (not all tools support all features)	Limited to Moderate Availability (not all tools support all features)	Wide Availability	Wide Availability	Wide Availability	Wide Availability	Wide Availability
Ease and accuracy for OCR	Good Acceptable Poor	Can the format be OCR'd at all? To what extent does the file format carry the optimal information necessary for clear and accurate OCR? Are there any distinguishing characteristics of this file related to OCR?	Good	Good	Acceptable	Acceptable	Acceptable	Acceptable	Good	[no information]	Good	Good	Good
Ease and accuracy of File validation	Good Acceptable Poor	Can the format be validated using DROID/PRONOM or JHOVE/JHOVE2, or other tools? Does the format specification include concepts and methods for conformance?	Good (JHOVE TIFF module; JHOVE2 module)	Good (JHOVE TIFF module; JHOVE2 module)	Good for GeoTIFF (JHOVE TIFF module; JHOVE2 module) Poor for BigTIFF (validation tool unknown)	Good for GeoTIFF (JHOVE TIFF module; JHOVE2 module) Poor for BigTIFF (validation tool unknown)	Good (JHOVE module)	Good (JHOVE module)	Good (JHOVE module)	Poor (validation tool unknown)	Good for versions 1.4, 1.5, and 1.6 (JHOVE module) Poor for version 1.7 (validation has ad hoc character)	Good for version 1 formats (JHOVE module) Poor for version 2 (validation has ad hoc character)	Poor (validation has ad hoc character)
Evaluating and Monitoring of Quality	Good Acceptable Poor	How easy is it to obtain or build a tool that would ensure that you are producing a well formed, high quality file that complies with a user specification profile for this format?	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	Does the format support elements that contribute to what is named by the deliberately imprecise term clarity? Two important characteristics are pixels per linear unit and bit depth ("bits per pixel"); clarity may also depend upon color accuracy and gamut, and will be adversely affected by lossy compression.	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	Good Acceptable Poor	How does the format support the documentation/metadata about the maintenance of color, e.g., tracking ICC profiles, or supporting the specification of sRGB, proRGB, eciRGB, Adobe RGB, or other color spaces?	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 (metadata possible for chromaticity, gamma, and ICC profile)	Good	Good	Good
Searchable Text Embedding	Pass Fail	Can searchable text be embedded? Note: Although this format comparison is focused on raster image data from scanning, some users who scan printed matter or manuscripts may be interested in identifying formats that can also carry searchable text.	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
Multi-Page (Multi-image) Capability	Pass Fail	Can the format carry multiple pages or images within the same file?	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