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.

April 17, 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 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*
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 their software tools available around this format? Are there user communitys/developer communities that are actively discussing the format and its further development?	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)	Moderate-to-Wide Adoption (moderate adoption in cultural heritage community, but widely adopted in communities such as moving images. Negligible support in browsers and still 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	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 format offer ample documentation (e.g., metadata) that makes the digital object a completely self-describing entity? Does the metadata fully describe the file/file format?	Acceptable	Acceptable	Acceptable	Acceptable	Good (more capabilities for entering metadata)	Good (more capabilities for entering metadata)	Acceptable	Good	Good	Good	Good
Native Embedded metadata capabilities	Good Acceptable Poor	What embedded metadata standards are available for this format? How mature are the schemas for each? What is the extent of use of the embedded metadata and who is using it?	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)	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.)	Good (XMP)	Good	Good	Good
Level of Work necessary to embed native metadata	High Medium Low	Is there a characteristic inherent to this format related to embedded metadata that distinguishes this format from others as far as level of effort required?	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 (XMP)	Low (XMP)	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; TerraGo geo display functionality may be limited to Windows app
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	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
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 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)

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*
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 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 (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
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
Compatability in existing enterprise environment (e.g., OCR-ability, quality review)	Good Acceptable Poor	What aspects should be looked at with the file format with respect to your enterprise environment 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 to Poor	Acceptable to Poor	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	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	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

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*
Settings and Capabilities (Pass/Fail)													
Clarity	Pass Fail	Does the format offer good support for extended bit depths and extended pixel counts, and such added features as multi-channel/multi-layer/multi-page files? (The values assigned in this category are especially rough-and-ready.)	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		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 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

AFTERSCHOOL TRAINING TOOLKIT

Tutoring to Enhance Science Skills

Tutoring Two: Learning to Make Data Tables

Sample Data for Data Tables

Use these data to create data tables following the Guidelines for Making a Data Table and Checklist for a Data Table.

Example 1: Pet Survey (GR 2–3)
Ms. Hubert's afterschool students took a survey of the 600 students at Morales Elementary School. Students were asked to select their favorite pet from a list of eight animals. Here are the results.

Lizard 25, Dog 250, Cat 115, Bird 50, Guinea pig 30, Hamster 45, Fish 75, Ferret 10

Example 2: Electromagnets—Increasing Coils (GR 3–5)
The following data were collected using an electromagnet with a 1.5 volt battery, a switch, a piece of #20 insulated wire, and a nail. Three trials were run. *Safety precautions in repeating this experiment include using safety goggles or safety spectacles and avoiding short circuits.*

Number of Coils	Number of Paperclips
5	3, 5, 4
10	7, 8, 6
15	11, 10, 12
20	15, 13, 14

Example 3: pH of Substances (GR 5–10)
The following are pH values of common household substances taken by three different teams using pH probes. *Safety precautions in repeating this experiment include hooded ventilation, chemical-splash safety goggles, gloves, and apron. Do not use bleach, ammonia, or strong acids with children.*

Lemon juice 2.4, 2.0, 2.2; Baking soda (1 Tbsp) in Water (1 cup) 8.4, 8.3, 8.7; Orange juice 3.5, 4.0, 3.4; Battery acid 1.0, 0.7, 0.5; Apples 3.0, 3.2, 3.5; Tomatoes 4.5, 4.2, 4.0; Bottled water 6.7, 7.0, 7.2; Milk of magnesia 10.5, 10.3, 10.6; Liquid hand soap 9.0, 10.0, 9.5; Vinegar 2.2, 2.9, 3.0; Household bleach 12.5, 12.5, 12.7; Milk 6.6, 6.5, 6.4; Household ammonia 11.5, 11.0, 11.5; Lye 13.0, 13.5, 13.4; and Sodium hydroxide 14.0, 14.0, 13.9; Anti-freeze 10.1, 10.9, 9.7; Windex 9.9. 10.2, 9.5; Liquid detergent 10.5, 10.0, 10.3; and Cola 3.0, 2.5, 3.2

Teaching tip: *The pH scale is from 0 to 14. Have students make two data tables, one with the data as given and one with the pH scale 0 to 14 with the substances' average pH in rank order on the scale (Battery acid at the lower end and Sodium hydroxide at the upper end) or create a pH graphic organizer.*

Example 4: Automobile Land Speed Records (GR 5-10)
In the first recorded automobile race in 1898, Count Gaston de Chasseloup-Laubat of Paris, France, drove 1 kilometer in 57 seconds for an average speed of 39.2 miles per hour (mph) or 63.1 kilometers per hour (kph). In 1904, Henry Ford drove his Ford Arrow across frozen Lake St. Clair, MI, at an average speed of 91.4 mph. Now, the North American Eagle is trying to break a land speed record of 800 mph. The Federation International de L'Automobile (FIA), the world's governing body for motor sport and land speed records, recorded the following land speed records. (Retrieved on February 5, 2006, from http://www.landspeed.com/lsrinfo.asp.)

Speed (mph)	Driver	Car	Engine	Date
407.447	Craig Breedlove	Spirit of America	GE J47	8/5/63
413.199	Tom Green	Wingfoot Express	WE J46	10/2/64
434.22	Art Arfons	Green Monster	GE J79	10/5/64
468.719	Craig Breedlove	Spirit of America	GE J79	10/13/64
526.277	Craig Breedlove	Spirit of America	GE J79	10/15/65
536.712	Art Arfons	Green Monster	GE J79	10/27/65
555.127	Craig Breedlove	Spirit of America, Sonic 1	GE J79	11/2/65
576.553	Art Arfons	Green Monster	GE J79	11/7/65
600.601	Craig Breedlove	Spirit of America, Sonic 1	GE J79	11/15/65
622.407	Gary Gabelich	Blue Flame	Rocket	10/23/70
633.468	Richard Noble	Thrust 2	RR RG 146	10/4/83
763.035	Andy Green	Thrust SSC	RR Spey	10/15/97

Example 5: Distance and Time (GR 8-10)
The following data were collected using a car with a water clock set to release a drop in a unit of time and a meter stick. The car rolled down an inclined plane. Three trials were run. Create a data table with an average distance column and an average velocity column, create an average distance-time graph, and draw the best-fit line or curve. Estimate the car's distance traveled and velocity at six drops of water. Describe the motion of the car. Is it going at a constant speed, accelerating, or decelerating? How do you know?

Time (drops of water)	Distance (cm)
1	10,11,9
2	29, 31, 30
3	59, 58, 61
4	102, 100, 98
5	122, 125, 127