16. IMAGE FILE FORMATS

Participants must make image versions of documents available when the evidentiary material is, in its essence, non-textual. The class of materials listed[16-1] in §2.1003 of the rule includes, among others, graphically oriented materials such as maps and photos, presentation materials such as overheads, computer printouts, and textual material that is handwritten. The NRC has also determined that forms may be considered non-textual (i.e., they may be imaged, rather than provided in full text).

Participants are not required by the LSN Rule to make image versions of **all** textual documents available online, although they may choose to do so as a means of fulfilling other obligations under the rule.[16-2]

The Licensing Support Network must ensure that delivery of image files does not become onerous to users because of large image files "downloading" via a slow telecommunications connection. Evidentiary materials that are essentially "graphic" (or authenticated images of textual materials that a participant may elect to make available online) are stored at -- and delivered to end users via -- each participants' document or document/image file server. As a result of this architectural approach, ensuring efficient and effective image delivery can only be accomplished by a cooperative design approach addressing the protocols used by the participants to store image files, or by some other strategy or software that a participant uses to deliver "workable-sized" image files.

When a user accesses a web page, they are actually downloading the page text and all the associated graphics from a web server. The user's web browser looks at the file extension (the letters following the "."). If it recognizes this type of file, it will display it. If the user's machine has the software application used to create the file, it will attempt to open the file using the application software package. If the browser does not recognize the file extension, it will ask if the user wants to configure a viewer (tell the browser which software program to use to view the file). Users also have an option to save the file to their hard drive or a disk.

What makes a file "workable-sized" is variable according to any given user's desktop and communications configuration, and is subjective according to that user's patience level. In general, a file 1.5 Mb in size will take approximately five minutes to download over a 56Kb (kilobyte) modem.

[16-1] Calibration procedures, logs, guidelines, data and discrepancies; gauge, meter, and computer settings; probe locations; logging intervals and rates; data logs in whatever form captured; text data sheets; equations and sampling rates; sensor data and procedures; data descriptions; field and laboratory notebooks; analog computer, meter or other device printouts; digital computer printouts; photographs; graphs, plots, strip charts, sketches; and, other descriptive material.

[16-2] § 2.1003(a)(1) requires that concurrent with the production of the text file will be an authentication statement that indicates where an authenticated image copy of a document can be obtained. Having an authenticated image version of a document available online would be one means of meeting § 2.1003.

16.1 Participant Files Stored in JPEG Format

JPEG (an acronym for Joint Photographic Experts Group, the committee that established the baseline algorithms) is a non-proprietary international standard (ISO/IEC 10918). For www use, JPEG/JFIF files are frequently smaller than GIF (Graphics Interchange Format) files, which results in faster transmission times. For full-color or grayscale images, JPEG is far superior to GIF, because it eliminates the banding common in GIF color transitions. JPEG/JFIF is not good for artwork with sharply defined lines, lettering, etc., as it tends to blur the borders between one color and another. It is generally not as good as GIF for palette-based images, since the palette is translated into truecolor and lossiness is introduced as soon as the image is JPEG compressed.

JPEG/JFIF and GIF are currently widely supported as inline image formats. That is, the browser's viewer will recognize those file types and open them without the assistance of a plug-in or without having to open the application package used to author the document. As a result, ***any participants' documents stored in a JPEG format do not need to be reduced to a single page-per-image format as a means to improve image transmission performance back to a user***.

16.2 Participant Files Stored in PNG Format

PNG (Portable Network Graphics, pronounced "ping") is a raster format which was adopted by the World Wide Web Consortium (W3C) in 1996 to provide improved www graphics unencumbered by legal restrictions. The PNG specification is currently a W3C recommendation. Unlike TIFF, PNG is designed for streamability rather than random access, which permits an image to start being displayed as it arrives. Unlike JPEG/JFIF, compression is 100% lossless (all image information is restored when the file is decompressed during viewing), so that images that are poorly suited to JPEG compression can still remain in a truecolor/grayscale colorspace as a PNG file.

Because W3C offers freely available libraries that support the full PNG specification, it seems likely that PNG will not have the problems TIFF has had with partial and incompatible implementations. It also seems likely that PNG and JPEG/JFIF (or JPEG/SPIFF) will coexist for the foreseeable future, as each format has advantages in different situations.

PNG is currently widely supported via browser plug-ins for early versions of currently supported Internet Explorer™ and Netscape™. Additionally, inline PNG, for example, is supported by Netscape Navigator 4.04 and higher and Microsoft Internet Explorer 4.x. That is, the browser's viewer will recognize those file types and open them without the assistance of a plug-in or without having to use another "plug-in" viewer. As a result, ***any participants' documents stored in a PNG format do not need to be reduced to a single page-per-image format as a means to improve image transmission performance back to a user***.

16.3 Participant Files Stored in PDF Format

PDF (Portable Document Format) is the format used by Adobe Acrobat™[16-3], which is a series of cross-platform software applications for the creation, sharing, and viewing of electronic documents. It is a free, cross-platform viewer with thumbnails and annotation and is derived from the Adobe Postscript™ printer language, although it also includes a number of additional features for active electronic documents (such as hypertext links). PDF 1.1 is the version used by the Acrobat 2.0 series. PDF is primarily intended for documents and is most widely used for electronic versions of print documents. The great strength of PDF is the functionality of the Acrobat software (including the free Acrobat Reader™). Acrobat enables the user to move from page to page, to zoom in and out, to follow hypertext links to other parts of the document, and more. This makes PDF a good target format for scanned documents.

PDF is widely used on the www, especially for documents that need their formatting tightly controlled. It has a multi-page format that makes document navigation easy for end users, however, the multi-page format can create performance issues when used in a WEB environment. Adobe and Web server vendors are trying to address these performance issues. (The new Acrobat 3.0 implements a variety of performance enhancements, including single page display, but will require special processing and HTTP byteserver software running on the server.) Additionally, it provides its own complete "search and view" tools. Inline support is beginning to appear, but it is unclear whether the inline PDFs will ever really supplant the Acrobat Reader product.

Note that there are three types of PDF files:

> "PDF Normal" - - PDF wrapped around text;
> "PDF Image" - - PDF wrapped around a raster/vector image; or
> "PDF Image & Hidden Text" - - PDF wrapped around a combination of text and image.

A simple scanned page image will generally be the middle option, usually in TIFF format; it will present an authentic representation of the original page, but it will not be searchable. If a page image subsequently goes through OCR processing, it can be either the first or the third option (with images for illustrations, signatures, etc.); the file will be significantly smaller and searchable, although it takes longer to create and will not appear as a "true" (clean) textual representation of the original page (for text searching purposes only) because the text file is "hidden" behind the image and it may be raw, unedited OCR output. As a result, ***any participants' documents stored in either a "PDF Image & Hidden Text" or a "PDF Image" format do not need to be reduced to a single page-per-image format as a means to improve image transmission performance back to a user***.

[16-3] Acrobat is the personal user version of the product family, suitable for an individual workstation or laptop machine. Adobe Capture™, is a more fully featured, production-scale version of the product.

16.4 Participant Documents Stored as TIFF Images

Participants may elect to use TIFF (Tag Image File Format) images for either non-textual materials, or to provide an authenticated image per §2.1003(a)(1). Each situation is slightly different and, therefore, guidance is addressed separately below.

TIFF images with CCITT 4 ("G4") compression are the de-facto standard in the document imaging market. There is substantial third-party and cross-platform support for this standard, it is "exportable" to different image formats, and TIFF headers are non-proprietary. TIFF is most useful for high-quality archive images and scanned documents, and is an effective intermediate format for feeding faxes and scans into OCRs and PDF. TIFF is preferable to PDF if further manipulation of the images will be needed (resizing, gamma corrections, etc.), but PDF is more effective for "end state" images intended for wide, cross-platform distribution.

As with PDF, TIFF files are of two types:

> TIFF Multi-page. This format enables easy document navigation because the entire document is downloaded to the client machine and the TIFF image viewer can control all navigation.
>
> TIFF Single-page. This format results in fast I/O (input/output) over Web environments because only one page is downloaded at a time.

16.4.1 TIFF Provided in Response to § 2.1003(a)(1)

The TIFF single-page format makes document navigation difficult because each page needs to be downloaded separately, externally from the TIFF image viewer. For text files in the LSN, where TIFF is being used by a participant to make the authenticated image available in addition to the text, losing navigation context within a document because images are downloaded one at a time is not a major consideration. However, TIFF multi-page format could cause I/O problems over the WEB because larger files need to be transferred; this is a problem to an unsuspecting user who requests that the TIFF image version of a 500-page document be downloaded. Therefore, ***any participant that chooses to provide an image version of a textual document, in addition to the text version required by the rule, should either:***

- ***Store the file in a TIFF single-page format as a means to improve image transmission performance back to a user***, or
- ***Implement a software (or package or application) solution which is able to deliver TIFF single-image pages.***

16.4.2 TIFF Provided in Lieu of Text for Graphic-Oriented Evidentiary Material

Where TIFF is used to image graphic-oriented materials such as maps or high resolution diagrams, **the participant should *store in a TIFF single-page format if a TIFF multi-page file for the entire document would exceed 1.5 Mb in size.*** Where TIFF single-page is used for large materials available in image only, the participant must provide a mechanism to link all associated image files with a single bibliographic header.[16-4]

16.5 Presentation Materials Stored in Other Formats

Participants may have authored materials in presentation software such as PowerPoint, Harvard Graphics, Corel Presentation or in any of a number of spreadsheet formats. In general, unless a user already has those packages installed on their desktop computer, a browser "plug-in" will have to be added to the user's machine in order to get a faithful image rendition of the document.[16-5] Therefore, participants have four alternatives available for making those files available on their server:

- ***Render and store presentation materials in a TIFF single-page format.*** Where a TIFF single-page is used for large materials available in image only, the participant must provide a mechanism to link all associated image files with a single bibliographic header; or

- ***Render presentation materials from the original authoring package format into a Windows Metafile (.wmf) format, which will create a single image per page.*** Participants must provide a mechanism to link all associated image files with a single bibliographic header; or

- ***Render presentation materials as a TIFF-multi page as long as it does not exceed 1.5 Mb;*** or

- ***Render the presentation materials as a PNG or PDF-image.***

16.6 Image Materials to Introduce into the Electronic Hearing Docket

The EHD is an official agency record repository and, therefore, is subject to NARA guidelines and regulations on acceptable electronic formats for any materials eventually retired to a Federal Records Center or the National Archives. This means that NRC will place strong emphasis on acquiring an electronic image version for storing of docketed materials.

[16-4] Document materials that are non-image in nature, such as Word, WordPerfect, Excel, etc., that may exceed 1.5mb in size should be expected by users because of the nature of scientific and engineering reports and studies. For overall system performance, it is advisable although not required that such large documents be unitized and packaged if the participant systems support packages.

[16-5] Users may save the file to their hard drive and import the file into a similar package they already have on their computer, but importing into another package will not guarantee faithful rendering of the original document.

Parties should anticipate that the documents they intend to introduce into the EHD will have to be in an image format, rather than text (although a text version would still be required to support text search and retrieval).

NRC's target standard for image storage is "PDF Image & Hidden Text." It is expected that the Presiding Officer(s) will require that all pages included in the offered exhibit (with exceptions for special handling items) must be in a single PDF file. Although NRC anticipates that the only acceptable image format for pre-filed materials offered by parties will be "PDF Image & Hidden Text," this does not preclude the introduction of non-digital materials, such as paper, in the courtroom that would subsequently be converted for entry into the docket by NRC.