

**Draft Meeting Minutes**

**PDF-Archive Meeting**

**October 3 – 4, 2002**

**Thurgood Marshall Building  
Washington, DC**

Dr. Stone and Mr. Levenson called the meeting to order at 10:15 a.m. on Thursday, October 3, 2002. They welcomed the attendees and explained the purpose of the meeting. Those participating in the meeting included:

Mrs. Mary Abbott	NPES
Mr. Stephen Abrams	Harvard University Library
Ms. Christina Bontempo	NPES
Mr. Robert Borochoff	Administrative Office of the U.S. Courts
Mr. John Brinkema	Administrative Office of the U.S. Courts
Mr. Jim Clements	Surety, Inc.
Ms. Jane Cohen	DTIC-OCQ
Mr. Roger Coombs	Quality Associates Inc.
Mr. Bryant Duhon	AIIM International
Ms. Betsy Fanning	AIIM International
Ms. Bette Fugitt	U.S. Dept. of Agriculture
Mr. Mark Gavin	Appligent, Inc.
Mrs. Pat Harris	NISO
Mr. Macduff Hughes	Adobe Systems
Mr. John Iobst	NAA
Mr. John Janick	Merck & Co., Inc.
Mr. Jose Juniel	EMC <sup>2</sup>
Mr. Jinsoo Kim	Image Solutions, Inc.
Mr. Tom Klaff	Surety, Inc.
Mr. William LeFurgy	Library of Congress
Mr. Steven Levenson	Administrative Office of the U.S. Courts
Ms. Linda Lorber	Millican & Associates
Mr. Jay Lucas	U.S. Patent & Trademark Office
Mr. Rich Lysakowski	CENSA
Mr. Joe McConnell	ProQuest
Mr. Basil Manns	Library of Congress
Ms. Pamela Mason	National Archives & Records Admin.
Mr. Charles Matheson	IBM
Mr. John Miller	U.S. General Accounting Office
Mr. Leonard Rosenthol	PDF Sages

Mr. Lou Sharpe  
Dr. Dana Stone  
Ms. Susan Sullivan  
Mr. Ricardo Thoroughgood  
Mrs. Melonie Warfel

Picture Elements, Inc.  
Merck & Co., Inc.  
National Archives & Record Admin.  
DTIC-BW  
Adobe Systems

Mr. Mel Bryson, Chief Information Officer for the Federal Judiciary welcomed everyone to the meeting and spoke of the excitement that the Federal Judiciary has for the project and the meeting.

Ms. Warfel, Mr. Levenson, and Mr. Hughes reviewed the initiation of the work that is resulting in the committee meeting on October 3 and 4. Meetings to begin discussing an archive format of PDF were initiated in May 2002. A number of U.S. government agencies had contacted Adobe Systems requesting assistance as they build their archives of PDF documents. It was felt that if some sort of standard were developed between industry and government that it would take some of the pressure off trying to determine how to archive PDF documents.

Mr. Hughes provided an overview of PDF including what it is, that Adobe only owns the copyright on the specification and that PDF is not trademarked. He also further defined the need for a PDF-Archive format.

The committee determined that the criteria for the success of the standard would be what elements or requirements were excluded rather than what is included in the standard. They feel it is important to be careful about what is excluded so that the work can be completed faster and the group can avoid having the scope expand unnecessarily. The goal of the committee will be to develop a specification that will provide guidance as to how to maintain all of the information that is needed to recreate the document within the document file.

The committee spent the remainder of the day and the second day identifying business needs for the standard, associating the needs to technical requirements and established working groups to further define the technical requirements. The fundamental goal for the committee is to develop a standard for the preparation of documents and associated metadata for archiving. This version of the standard will not address the physical storage media issues that will be addressed at a later date but will address emulating static paper. The users that will benefit from this standard include legal users, academics, archivists, records managers, historians, general technical and scientific users and office and business users.

The following are the general business needs identified for this standard:

- Self-contained documents
- Manageable file size
- Traceable history for the document beginning at its origin
- Scientific community want rotating graphics

- Text should be searchable
- Security
- Authenticate the document to ensure originality
- Embed meta information in the document
- Must not require additional files to accompany the PDF document file
- Extensibility
- Reference reader must exist
- Support wide range of page sizes
- Restrict extraneous/deleted information
- Require embedded fonts
- Preserve information present in original font
- Allow for limiting the choice of fonts
- Eliminate use of proprietary conversion methods
- External links and hyperlinks to other PDFs
- Common metadata

The committee also looked at the business needs as they apply to specific government agencies that were represented at the meeting. The following are those needs:

- Judiciary
  - Equivalent of microform that sets the quality benchmark
- Patent
  - Integrity
  - Authenticity
  - Word Search Capability
  - Color
  - Class/subclass data
  - Support for foreign languages and alternate character codes
  - Support mathematical and scientific formulas with search capability
- Library of Congress
  - Other image formats such as JPEG2000
- GAO
  - Security
  - Audit trail recording who has read the file
  - Metadata including the representation of copyright information
  - Embedding colors and fonts to ensure the document will render as in the original document (e.g., colors in pie charts must be retained; contrasting colors must contrast; matching colors must match)
  - TIFF and EPS must convert to PDF
  - Open access to documents without any digital rights constraints
  - Accessibility as it relates to ADA Section 508
- Defense
- Courts
  - Documents must be complete
  - No external links allowed

- FDA
  - External links including hyperlinks to other PDFs especially in the same submission

General industry business needs include:

- Compatibility so that new readers will read old files for at least 40+ years without having to migrate to a new format
- Need to archive content as well as rendered content
- Address XML and workflow
- Create final form rendered documents

Items for consideration of the committee as they draft the specification include:

- The standard must be ubiquitous and have freely available tools to enable adoption and use.
- The standard must be easy to implement by software developers in software and easy to use for end users.
- The standard should specify both file format and behavior of the readers.
- The standard should include a clear policy on undocumented keys to restrict inclusion in file or restrict use in implementation.
- The standard will not go beyond what already exists in the PDF specification.

The committee discussed the issue of compliance determining that some applications may be more restrictive than required by the standard. Therefore, compliance is not desired at this time and will be only if it is absolutely necessary. Further discussion on this issue will be required.

With regard to the reference reader, the committee acknowledged the need for one reference reader that will as a minimum read and print a document. For the government agencies to adopt the standard, it will need to show how the document can be read in 50 years with any associated guarantees that the standard can offer.

The committee discussed executables included in the file. No consensus was reached on this issue, which will require further discussion. They also discussed the issue of encryption learning that NARA will not accept encrypted data and Merck could accept the requirement to remove encryption prior to archiving the document. The committee agreed to prohibit encrypted files.

For software developers, the committee in cooperation with Adobe Systems will need to make sure all patent issues are covered for PDF 1.4. The committee believes a statement from Adobe covering patents that have not been issued may be required. Additionally, the committee discussed if the PDF-Archive standard would be based on PDF 1.4 or something earlier.

The following working groups were formed (\* denotes the chair of the working group):

- Rendering (M.Hughes\*, L.Sharpe, L.Rosenthol, and M.Gavin)

- Colorspaces
  - ExtGState, device dependent operators
  - PS Xobjects
  - Allowed image types
- Metadata (J.Brinkema\*, R.Lysakowski, J.Cohen, J.Iobst, J.Miller, J.McConnell, W.LeFurgy)
  - Copyright of embedded materials
- Fonts (S.Abrams\*, J.Miller, L.Rosenthal, M.Hughes, and L.Sharpe)
  - Allowable types
  - Encoding
  - Embedding rights preserved
- Multimedia (L.Sharpe)
  - Sound
  - Video
  - Animation
- Security (J.Brinkema\*, M.Gavin, R.Lysakowski, R.Thoroughgood, J.Janick, D.Stone)
  - Authentication
  - Encryption (Not to be included)
  - Signatures
- Intellectual Property/Referenced Specifications (L.Sharpe\*, M.Warfel, M.Abbott, J.Lucas, B.Fanning)
- Embedded Files (L.Rosenthal\*, J.Brinkema)
- Unrecognized Data/Extensions (J.Lucas\*, M.Hughes, M.Gavin, J.Brinkema, L.Sharpe)
- Hyperlinks, External References (B.Fugitt\*, J.Brinkema, M.Gavin, R.Lysakowski, L.Lorber, J.McConnell, W.LeFurgy, L.Sharpe)
  - File collections
- Annotations (L.Lorber\*, L.Rosenthal)
- XML (J.Iobst\*, R.Lysakowski, S.Abrams, L.Sharpe)
- General File Format (M.Gavin\*, M.Hughes, B.Manns, L.Sharpe)
  - ASCII/Binary
  - Compression Methods
  - Update Sections
  - PDF Version
- Logical Structure/Tagged PDF (J.Miller\*, S.Abrams)

Drafts from each of the working groups is due to be sent to Betsy Fanning, AIIM International at [bfanning@aiim.org](mailto:bfanning@aiim.org) by November 18, 2002. These drafts will be combined into one document that will be sent to the committee for review prior to discussion at the next meeting.

The next meeting of the committee will be held in Washington, DC on Thursday and Friday, December 12 and 13.

The meeting was adjourned at 2:50 p.m.