**PDF-Archive
Draft Meeting Minutes
February 27-28, 2003**

Mr. Stephen Levenson and Mr. Dana Stone called the meeting to order at 9:00 a.m.  The meeting participants included:

| | |
|---|---|
| Mary Abbott | NPES |
| Stephen Abrams | Harvard University Library |
| Jane Cohen | DTIC |
| Marion Ellis | Merck & Company |
| Betsy Fanning | AIIM International |
| Scott Foshee | Adobe Systems |
| Bette Fugitt | Dept. of Agriculture |
| Kenneth Haller | INS/ORM (Tessada & Assoc.) |
| Pat Harris | NISO |
| Macduff Hughes | Adobe Systems |
| Jinsoo Kim | Image Solutions, Inc. |
| Bill LeFurgy | Library of Congress |
| Steve Levenson | Administrative Office, U.S. Courts |
| Linda Lorber | Millican & Associates |
| Rich Lysakowski | CENSA and GERA |
| Joe McConnell | Proquest Co. |
| David McDowell | NPES |
| Basil Manns | Library of Congress |
| Pamela Mason | NARA |
| Jeffrey Reed | NARA |
| Lou Sharpe | Picture Elements Inc. |
| Paul Showalter | Internal Revenue Service |
| Dana Stone | Merck & Company |
| Susan Sullivan | NARA |
| Doug Vander Wilt | Honeywell |
| Melonie Warfel | Adobe Systems |

The February 27-28, 2003 meeting agenda (PDF-A 2003-002) was approved with the addition of developing an advanced meeting schedule for the working group. (Cohen/Foshee)

The draft meeting minutes for the December 12-13, 2002 (PDF-A 2003-001) meeting were approved as amended with the removal of the statement regarding NARA on page 3.  The sentence regarding the title was amended to " Mr. McDowell suggested a title for the document other than PDF/A." (Cohen/Mason)

During the discussion on the meeting schedule it was also identified that an ISO project leader was needed.  Mr. Stephen Abrams and Mr. Stephen Levenson were nominated as co-leaders for the project at the ISO level.  The committee unanimously approved the appointments.

The tentative meeting schedule moving forward will be:
May 1 and 2
July 10 and 11
September 4 and 5
October 20 – 24 (New Orleans for ISO Joint Working Group)
November 20 and 21

Meetings will take place approximately every two months.  The committee unanimously agreed to this schedule.

The Annotations working group presented their proposed work which resulted from discussion.  They have concerns on the format for sound annotations.  Drop down forms need to be easy for users so that all they are doing is click and point.  This working group is deferring the handling of widgets, describe actions on forms for the printing process, to the forms group.  The group discussed picking and choosing annotations that should be included in PDF-A.  Since Annotations covers a wide variety of items, the committee needs to determine what should be included and excluded from this area.  The working group thinks it includes visual markups but not multimedia.  The group discussed the need to include markups in PDF-Archive.  The Courts would not want it while others do not see it as important.  Any forms issues the working group identifies will be sent to the forms working group.  The PDF-A specification will include requirements for the reader.  The question then is does the reader need to inform that an annotation exists and allow the annotations to be turned off.  The committee identified a need to have some metadata that would identify that something was included in the document and what it was.  This would be described in terms of what the output is.  The committee agreed to disallow multimedia and agreed to the proposal from the Annotations working group.

The Metadata working group presented their report.  This working group recommends the PDF-A spec should address metadata for identification and description and to document the technical and administrative aspects of the document.  The spec must support additional requirements an organization may have – so it must be flexible.  It should ensure consistency among different applications and provide for a tight bond with file documents.  The working group proposed using Adobe XMP for PDF-A metadata.  They recommend the specification should require a minimal set of mandatory descriptive and/or technical properties, have the properties embedded in each file or document and preserve schemas in close conjunction with the files.  The working group reviewed the Extensible Metadata Platform (XMP) that defines a basic set of metadata properties.  For more information,
http://partners.adobe.com/asn/developer/xmp/pdf/MetadataFramework.pdf.  XMP is a constrained subset of RDF (Resource Description Framework).  The committee discussed why some metadata fields should be included and aliases.  They also discussed how metadata is entered in a Word document at the time of creation and how it is handled by Acrobat that currently handles most of it.  Other concerns discussed included the impact of XMP on the file size and the author and create date.  PDF is a conversion tool and the metadata may convey the conversion date rather than the creation date.  If there is to be mandatory metadata, the values will have to be defined in the specification.  It was acknowledged that the metadata couldn't be verified.  The committee agreed to use XMP as the container for the metadata approach; however, it may want to revisit this at a later time if Adobe does not decide to take on XML schemas

in the Acrobat product. The committee asked the working group to go to the next level of specificity as to the fields that should be required.

The Logical Structure working group reported on its work. Tagging is not necessary for visual reproduction but is necessary for U.S. ADA Section 508 compliance so therefore, the specification must allow for tagged documents. We need a reader that reads tagged and non-tagged PDF. The datastream will be defined to have both. A profile will need to be defined that will work with the accessible and non-accessible PDF based on specified criteria. The committee discussed compliance levels. If the specification only has one compliance level, an informative annex will need to be included that will explain how to create accessible formats. The structure must be able to be translated into Unicode to be accessible compliant. PDF-X includes multiple conformance levels. Mr. LeFurgy proposed to allow tagged PDF in the standard and not split it out into two compliance levels but to say if tagged do this and if not move on to X. A profile in the document would specify if you want to accessible what needs to be done. Section 15 of the document will be revised to be clause 1, Tagged PDF and clause 2, Non-tagged PDF. It needs to be noted that tagging, as an issue is not the same as accessibility. The tags should initiate a pathway to accessibility. The specification will allow the tags agreed upon. If there are not two conformance levels specified then a named profile will be needed. The committee moved to include two specified named profiles or conformance levels in the normative text of the standard. Affirmative 12 Negative 2 Abstain 0 Motion carried. Compression and encryption are considered more file format than logical structure and will be covered by that working group.

The Forms working group conducted a discussion session to determine what the committee's views on forms was. Ms. Fugitt discussed how the government uses forms and the importance of maintaining the options from the drop down lists on electronic forms. It was noted that electronic forms incorporate Java scripts and calculations that can be performed in the background. The committee identified the need for self-revealing forms/documents that contain all the options offered. This includes not only the form but also all the associated data and programs that accompany the form. The committee suggested the working group limit their work to the frozen form as it is represented today. Ms. Fugitt and Mr. Showalter are to be added to this working group. Ms. Warfel and/or Mr. Hughes are to identify an Adobe technical person to participate on this working group.

The Compression working group reviewed various compression algorithms. The committee agreed to allow Flate compression and disallow ASCII85 Decode and ASCIIHEX Decode as ASCII has been obsolete for several Acrobat releases. The need to make compression easy for the archivists was noted. The following compression algorithms will be allowed in the specification JBIG2, JPEG, and CCITT Group 4. Adobe will find out about the Intellectual Property issues associated with including JBIG2 and provide a recommendation to the committee at the next meeting. Adobe recommended JPEG2000 not be included in this version of the standard. The committee agreed to this with 11 votes to not allow JPEG2000, 1 vote to allow it and 2 abstaining. The committee also discussed compression should not be included in the standard other than in an informative annex. The PDF-A standard will allow lossy compression. The committee proposed that once a file is received by NARA it would need to comply to the compression NARA has selected and not be lossless. As long as the standard maintains the ability to include an uncompressed image it would be acceptable to NARA.

The standard needs to include a bullet for uncompressed compression modes.  The current PDF specification allows for lossy, lossless or uncompressed compression.

The Rendering working group reviewed the changes they made to the latest version.  The committee agreed that thumbnails should not be made mandatory but be allowed.  The section will be amended to say that thumbnails will be allowed but not mandatory.  While this standard can reference sections in PDF-X, the committee would prefer to include the appropriate sections in this standard.  We will need to ensure that there is compatibility between PDF-X and PDF-A.  PDF-X is thought to be able to feed into PDF-A.  Mr. McDowell will research X2 and provide feedback to be added to section 4.1 where a point or two of ICC's data registry will be included.  The committee needs to review how colorspace is defined as there is a difference in color recorded in the file and color rendered in the file.  In section 4.7, the working group may want to explicitly state that all transparency is disallowed.  Reflex objects were left out as a result of not allowing transparency.  The committee requested more explanatory text be added to this section.

The Fonts working group reported on the revisions to the text that they made.  The approach they took is to recover semantic level information about the document so that there is no reliance on any external files.  The committee discussed restricting Unicode.  The PDF-A validation software will flag that a font is not allowed to be embedded if the permission to embed the font does not exist.  Section 6.8 should be modified to state "in a file complying with…"

As part of the Intellectual Properties working group report, Adobe reported on a draft statement it has obtained from its legal department with the same releases as Adobe 1.3.  The IP statement will not be included in the standard but will remain on file with the committee and with ISO.  PDF-X did not require any statement other than for use of the name.  It was noted that usually an IP statement is completed at the end of the process because it is not known what is needed until that point.   The committee also noted that ISO will not allow any licensing terms that are discriminatory in that it would charge one company but not another one.  The committee agreed that it must make sure that royalty free technology is used in this standard.  Adobe stated that they have no intent on charging anything for use of the technology.  The Secretariat will check with ISO on royalty free patents and/or the IP issue.  Since PDF is not owned by Adobe, the committee may use PDF-A.  The committee will need to have a statement from Adobe that if Adobe chooses in the future to trademark PDF, the committee may still be able to use the name.  Adobe will work with its legal department to finalize the wording on the statement.

The Hyperlinks working group reported on their work stating that links in the document are allowed but they need to be visible and stagnant.  With relative and absolute links we can get into trouble.  The committee decided it would allow GoToR, but the specification would be for relative file specifications.  If there is a remote GoTo dictionary, the relative path would be X.  The standard will restrict the file format rather than the viewer.  If links were removed, it would not necessarily be the same document even though it would read correctly.  Many of the documents that will be archived will be converted to another media for long-term storage.  These documents must work well in hard copy as well as in digital format.  Mr. Hughes rejected this idea stating that if you have an Excel spreadsheet with formulas, when you print to paper or to PDF, you lose all the formulas.  Therefore, the primary goal should be to convert the documents as with paper.  When a

document is submitted to archives, expectations may be set as to what can be reproduced.  With URI, it may be assumed that the functionality and security is preserved.  The committee discussed the issue of whether the standard can regulate the records management and the file format without losing functionality.  In addition to the paper analogy, archivists maintain books with bibliographies but do not necessarily maintain all the books in the bibliography.  Links do not need to be a URI, however, the standard should specify how to preserve the URI that should be fully documented but not actionable.  The committee wants the ability within the metadata to confirm that there was a link but that the content of the link is not correct.  The committee discussed forbidding actionable URIs. Strict forbidding of URIs is not attainable – 5 members are against this, 9 agree that strict forbidding is attainable and 1 voted to abstain.  The standard may need to include a few paragraphs discussing both sides of the issue.  All committee members are requested to send to Ms. Bette Fugitt, proposed wording as to how you would like to see URIs handled.  The committee discussed having URIs be fully documented but not actionable.  The committee recommended the working group provide a full-text write up of the issue and mark it as undecided.

The following URI outcomes were discussed:
- URI – Forbidden (1)
- URI – Actionable URI forbidden (text is there but cannot do anything)
- URI – Allowed, (must be fully articulated) but compliant readers are not required to utilize it; thus it is advised that the complete text of the URI also is made visible as text (2)
- URI – Fully documented so that action could be taken (allowed)
- URI – Converted to static display
- URI – Undecided (3)

The committee identified the three most likely conditions (numbered above as 1, 2, 3) and then voted on the most likely of those as follows: 1 received 2 votes; 2 received 10 votes and 3 received 3 votes.  The committee raised the question of where the URI text would be visible as if it is fully articulated the URI text would change the visual appearance of the page.  If we cannot or do not want to change the visual appearance of the page the URI text should be included in the metadata and in the reproduction of the page.  The committee recommended URI outcome 2 and requests the working group to flesh out the details within the next 3 weeks (by March 24, 2003) and post them to the full group for discussion.

The Multimedia requirements will be included in the Annotations working group work and section of the standard.

The Unrecognized Data working group is to be merged with the File Formats working group.  Mr. Rich Lysakowski requests to be a member of this working group.

Mr. McDowell proposed the title of the standard to be "Document management – Long-term electronic preservation – Use of PDF (PDF/A)."  He also recommended the following as the modified scope statement for the document:

"This International Standard specifies the use of Portable Document Format (PDF) for long-term preservation of black and white and color compound documents as electronic data.  Compound documents may contain combinations of character, raster, vector, and

other data.  This International Standard also specifies methods for creation from these data of an exact visual reproduction of the document as it appeared at the time it was submitted for preservation.  It also enables the preservation and retrieval of appropriate metadata."

In a committee straw vote to accept the modified scope and title, the committee unanimously agreed with the title and scope.  They will be incorporated in the next revision of the document.

Mr. McDowell proposed the following as the Introduction to the standard:

**Introduction**

In the traditional documentation world, long term document storage and preservation has been accomplished by a combination of careful storage of paper records under controlled conditions and the use of optical reproduction of these materials in a variety of reduced size photographic formats such as microfilm, microfiche, etc.  However, these methods do not address the growing number of documents that are created and used as electronic records in a wide variety of data formats.

There was an urgent need to archive electronic documents in a way that would ensure preservation of their contents over an extended period of time, and to further ensure that those documents would be able to be retrieved and rendered with a consistent and predictable result in the future.

Even more important was the need to define a both a file format and the behavior of retrieval devices (readers) that is compatible with both electronic documents and scanned images of traditional documents. This need has existed, and continues to exist, in a growing number of international government and industry segments, including legal systems, libraries, newspapers, regulated industries, and others.

The initial activity to define the business and technical requirements, and study possible solutions, was sponsored by a joint committee formed under AIIM International (the Association for Information and Image Management, International) and NPES (NPES The Association for Suppliers of Printing, Publishing and Converting Technologies).
The use of a restricted subset of the Adobe Portable Document Format (PDF), a publicly available published specification, similar to the work done in the printing and publish industry and known as PDF/X (ISO 15930), was identified as a solution path and the project became known as PDF/A.

That work has led to this International Standard, which addresses the use of PDF for  the long term storage of multi-page documents that may contain a mixture of text, raster images and vector graphics. It addresses the features and requirements that must be supported by reading devices that will be used to retrieve and render the archived documents. A goal of this initial version is to emulate static paper with the added need to include electronic annotations, electronic signatures, marginalia, approvals, etc.

Because there is a significant need to index, inter-relate, and search such archived records, considerable effort was made to insure that the file format itself includes an appropriate level of information about the document as well as enabling the association of appropriate metadata with each file. However, it must be noted that such information requirements vary widely among the various anticipated user communities.  Therefore, emphasis was placed on minimizing the required information but ensuring that the metadata capability was versatile enough to accommodate a wide variety of user needs.

This standard does not address the media used to record the electronic data or the associated requirements for its storage and/or maintenance.  Such requirements are addressed by other ISO technical committees and International Standards.

It is anticipated that a variety of products will be developed based on PDF/A, such as readers (including viewers) and writers of PDF/A files, and products that offer combinations of these features. Different products will incorporate various capabilities to prepare, interpret and process conforming files based on the application needs as perceived by the suppliers of the products. However, it is important to note that a conforming reader must be able to read and appropriately process all files conforming to a specified conformance level.

An ongoing series of Application Notes [1] is maintained for the guidance of developers and users of the ISO PDF/A family of International Standards. They are available from TBD at http://TBD.

The committee moved to include the introduction as proposed by Mr. McDowell in the ISO New Work Item (NWI) proposal. (Fugitt/Vander Wilt) 14 Affirmative, 0 Negative, 1 Abstain.

The committee recommended the order of the standard should parallel the PDF Reference Manual.   The next revision will incorporate this in it and will move the file format up in the outline.

It was noted that the working groups might want to put questions in the front of the sections for review purposes while we are in the draft stages so that issues are not lost and can be addressed appropriately.

The committee discussed forming a separate U.S. TAG for the Joint Working Group to handle the project as it moves into the ISO arena.  It may be more cost effective to have the group be a sub-committee off of an existing TAG.

The committee unanimously appointed Mr. Stephen Levenson and Mr. Stephen Abrams as the ISO project leaders from the U.S.

The meeting was adjourned at 1:55 p.m. (Fugitt/Cohen)

**Draft Meeting Agenda**
**PDF-Archive**

**Hilton Silver Spring**
**8727 Colesville Road**
**Silver Spring, MD 20910**
**Conference Room: TBD**

**February 27, 2003   9:00 a.m. – 5:00 p.m.**
**February 28, 2003   9:00 a.m. – 3:00 p.m.**

1. Introductions                                                  Levenson/Stone

2. Approval of Draft Agenda (PDF-A 2003-002)

3. Approval of Draft Meeting Minutes, December 12 & 13, 2002   (PDF-A 2003-001)

4. Reports and Review of Drafts of the Working Groups

   Each working group chair will present the work the group accomplished since the October meeting and lead the discussion of the draft and technical issues.  Each group will have approximately a half hour.

   | | |
   |---|---|
   | Annotations | L. Lorber |
   | XM/Metadata | W. LeFurgy/R. Lysakowski |
   | General File Format | M. Gavin |
   | Logical Structure | J. Miller |
   | Forms | M. Warfel |
   | Compression | L. Sharpe |
   | Rendering | M. Hughes |
   | Fonts | S. Abrams |
   | Multimedia | L. Sharpe |
   | Security | J. Brinkema |
   | Intellectual Property | L. Sharpe |
   | Embedded Files | L. Rosenthol |
   | Unrecognized Data | J. Lucas |
   | Hyperlinks | B. Fugitt |

5. New Business

6. Wrap-up and Next Steps                                         Levenson/Stone

7. Adjournment