



Self-contained

***“Everything” required for reproducing document inside document – extreme position
No: we don’t have to do this***

File Size

***Documents typically in large groups with common elements
Needs resolution – example fonts***

How detailed must spec be?

***Must PDF spec be copied or referenced
Is spec clear enough for this purpose
Must verify that referenced portions are adequately detailed and unambiguous***

Do not address [physical storage] media issues

Should multimedia in PDF be addressed?

No consensus yet. Note sound can be embedded in PDF today but video cannot.

Origin of document, history traceable

Further discussion necessary

Searchable text (and retrievable)

***Yes, this is important.
Should have searchable metadata too.
Details TBD***

Ubiquitous – freely available tools

***No legal obstacles to making free tools
ISO already requires that technology is either free or available on terms “reasonable and non-discriminatory”
Should we insist on free? Needs discussion***

Security – don’t embed executables

***Agree this is an important goal
Review what has been done for EDGAR (SEC) and others
Make sure content can’t change***

Goal for v1: emulate static paper

***Agree as high level goal
Need to resolve marginalia and private tags issues***

Classes of users – self-evident benefit to:

Legal users
Academic
Archivists and records managers
Historians
General technical and scientific users
Office and business users...

Must have means of authenticating – verify document has not changed over time, is the “original”.

Agreed as desirable goal.
Not necessarily contained in file.

Spec must be easy to implement by software developers
Easy to use for end users

Average individuals or skilled experts?

Summary of above two lines: Spec must make it easy to develop software that is easy to use.

Unambiguous

Standard must be unambiguous enough to allow for true and accurate validation of whether a file is PDF/A compliant

Able to embed metadata in document

And also require specific kinds?
Need further discussion

Do not require additional files to accompany the PDF

Agree to allow PDF/A as atomic unit, but may also allow workflows where additional files are involved
Rich L. notes there is an Australian solution for multi-file using XML (VERS).

Collections of PDFs submitted as a package

Probably defer to v2?

Extensibility

Able to add features in the future
Set clear expectation for how future versions will be handled
One reader should support all past formats
Or should we prohibit private metadata?
This could prevent archiving actual working documents

**Perhaps allow private metadata with some compartmentalization
Needs more discussion**

May support different compliance levels targeted at different industries

**Not desired. Only if absolutely necessary.
Some applications may be more restrictive than required by standard
Needs more discussion**

Standard should specify both file format and behavior of readers

agreed

“Standards for preparation of documents and associated metadata for archive”

Far in future, must be able to write program from scratch using the PDF/A standard, PDF Reference Manual and associated documents; resulting document will render same (for evidentiary legal purposes) as rendered today.

**What degree of exactness of reproduction is required?
Comparable to paper, photocopy, fax, microfiche?
“Recognizably same”**

Support wide range of page sizes

May need to clarify implementation details

Must spec be self-contained (included PDF Ref., JPEG, font formats, ...?)

ISO allows references to “publicly available” documents, providing date or version optional.

So many specs referenced in PDF that making it self-contained would be impractical

Should we have higher standard than ISO?

Require that every referenced document have a Library of Congress number?

Every referenced spec should be as well available as the PDF/A spec itself

Clear policy on undocumented keys

Restrict inclusion in file or restrict use in implementation?

Must be able to have process-specific metadata

Restrict extraneous/deleted information?

Needs discussion

Judiciary requirement: equivalent of microfiche

Sets quality benchmark

Not necessarily good enough for all applications

Needs more discussion

Patent Office requires:

integrity, authenticity

word search to extent possible.

Color "reasonable".

Class/subclass data.

Support foreign languages

Lib of Congress: Would like varied image formats (JPEG2000, others?)

Defense: security issues tbd (discuss again tomorrow)

Would like Audit of who has read file

Probably not part of PDF/A standard but allowed as part of metadata

Need general discussion of metadata needs

Encryption: should it be forbidden?

NARA currently will not accept encrypted data

Merck could accept requirement to remove encryption before archive

GAO: separates classified documents in separate storage.

Anything in public domain has no encryption.

Agreed: prohibit encrypted files

Consider representation of copyright information in metadata. Discuss with metadata issue.

Should a PDF/A reader obey DRM restrictions, e.g. no print?

No issue if no encryption in PDF

GAO: Colors and fonts render as in original document.

Fonts must be embedded.

e.g. Colors in pie chart must be retained. Contrasting colors must contrast, matching colors must match.

GAO: Images: TIFF and EPS can be converted into PDF.

GAO: open access (no DRM restrictions)

PTO: support mathematical and scientific formulas with ***searchability. (e.g. search for Benzene Ring symbol)***
Censa: note standards are being developed for these things, e.g. MathML.
Can do in PDF in non-standard ways today with AltText or metadata.

Don't go beyond what's already in PDF.

GAO: accessibility. Must be able to include accessibility information e.g. description of chart.

Do not do anything to prevent inclusion of enough accessibility information for 508 standards
Must be able to make scanned document images accessible – current PDF solution ok.

Image Solutions speaking of FDA: need external links (hyperlinks to other files especially in same submission)

Could restrict to relative links
Needs more discussion

Want common metadata.

What kinds?
Identify related documents, history
Part of general metadata discussion

Courts require documents complete: no external links.

But maybe not PDF/A requirement
Part of links discussion

Industry requirements

Keep compatibility – would like new readers to read old files for at least forty years without having to migrate files to new format

Library (Harvard)

Keep in mind even longer periods, hundreds of years

Print publishing (Newspapers)

Moving to XML workflow

Need to archive content that's rendered as well as rendered content

Lightweight structure

Need separate discussion on XML

Software developers:

Need to make sure all patent issues are covered for PDF 1.4.

Want statement from Adobe covering unissued patents.

But will PDF/A be based on 1.4 or something earlier?

Issues from May document "PDF Archiving":

Require embedded fonts

Make sure font permissions are preserved/honored?

Preserve any information present in original font

Note not all fonts allow embedding

Limit the choice of acceptable fonts?

No. Definitely not acceptable internationally.

Some agencies may want to limit fonts

No proprietary compression methods (LZW, JBIG2?)

External file refs

Executables

Could allow embedding but require reader to ignore

Simpler to just forbid all embedded file attachments

Needs more discussion

3/b/4/g – ascii PDF? No. May have been driven by a past NARA requirement, no longer needed.