The principle feature of this proposed new standard is to allow the use of the PDF 1.7 (ISO 32000-1) file-embedding feature (the embedding of one or more instances of any arbitrary file format) in an instance of an archival PDF file. Per the Introduction (page *vi*):

> This part of this International Standard adds a new goal which is to enable PDF documents to serve as containers for other file formats, so that a single physical file can contain not only the visual representation but also other representations including the original authored version, richer semantic formats, and others.

We believe this proposal entails a significant redefinition of the key term "electronic document" and "archival format", and constitutes a significant departure from the warranties implicit in terming PDF/A an archival format.

The intent to include files in any arbitrary format is in contravention to the principle that informed the choices made in the specification of PDF/A-1 and PDF/A-2: that the document instance contain within itself everything necessary (given a conforming reader) to extract the complete semantic value of the document. Hence the restrictions on fonts not completely contained within the document; on links to destinations outside the document; on the use of Reference or PostScript XObjects; and on the use of 3D, Sound, Screen, and Movie Annotation types. Hence also the requirement (See 6.6.2.3.2) that all "extension schemas referenced from any metadata stream in a conforming file shall have their descriptions embedded within the referencing metadata stream or the stream that is the value of the Metadata key in the Catalog."

The proposal comprises an extension to the change made in PDF/A-2 (which we did not review or comment upon), which extended PDF/A-1 to allow the use of embedded files, provided that those embedded files themselves complied with PDF/A-2 (all fonts embedded, no links to external URLs or external files, etc.). There is currently no restriction of any sort on the type (document or otherwise), purpose, or format of files that may be embedded – an editor's note to paragraph 6.8 simply states:

> This is the primary section that differs between A-2 and A-3. In this initial NWI, I have simply removed the requirements that embedded files need to also comply with PDF/A. I expect that we will want more requirements to appear here, but these changes set the stage.

There are arguably good reasons to implement this proposal, all of which obtain principally for producers of PDF content. This change seems to have been motivated by the desire to enable the packaging of the source material from which the PDF was produced, or the data used to produce it, or other related collateral materials, along with the document representation ("PDF classic"), making use of PDF's digital signature capabilities if desired. The weight of an international standard that defines a format as "archival" would presumably increase the incentive for toolmakers to extend their tools to facilitate embedding of source and related content, and the creation of the file provenance information for the container file itself (though not necessarily for embedded files, and not necessarily including any information about the relation of these files to the "document" itself) in both this proposal and in PDF/A-2, particularly for those users with statutory and other archival obligations. The effect could be to encourage the practices of creating at least minimal provenance and event metadata for at least the document itself, and of capturing in a single container associated source artifacts (early versions, including versions in formats other than PDF/A, equivalent documents in various natural languages, as well as other arbitrarily related artifacts) at a critical point in a document's evolution as an artifact.

Leaving aside the practical issue of potentially extremely large file sizes, the principal downside of this proposed archival format, ironically, is for archives themselves. While the proposal does make statements about what a conforming PDF/A-3 application must, should, and may do, the ISO standard process does not mandate the creation of a reference application, much less a free and/or open source reference application. Those requirements, with respect to the embedded files, are minimal:

- A conforming reader must have an annotation handler for a FileReference annotation (ISO 32000, 12.5.1), but what that handler does for any specific (embedded or otherwise) file format is unspecified. A conforming reader must merely provide a mechanism to display the Contents key of the annotation dictionary, which will provide a text description of the file inserted by the creating application.
- Per the specification, an action which entails navigation to embedded files (GoToE) must display the F (the root document of the target relative to the root document of the source, which, for an embedded file, is the current document) and D (the destination in the target to jump to) keys of the GoToE dictionary – i.e. merely displaying the location information of the embedded file is the only required action. The proposal continues:

> In addition, since the actual invocation of these three actions by a conforming interactive readers involves the locating of and interacting with other files that may or may not be conforming, the reader may choose to not allow the actual invocation of these actions.
>
> NOTE For purposes of archival disclosure of the complete information content of conforming files, it is important for interactive readers to provide some mechanism to expose the destination of such actions. However, this part of ISO 19005 does not prescribe any specific behaviour.

- With respect to the file specification dictionary for the embedded file itself, a conforming reader is required only to display name strings pertaining to the file. Per the proposal:

> A conforming reader shall provide a mechanism to display the name strings from the value of the EmbeddedFiles key in the names dictionary of a conforming file. In addition, a conforming reader may also choose to display information from the associated embedded file stream dictionaries or their Params dictionary.

There is no requirement for a conforming reader to provide a "display" capability for the files which are embedded. More critically, from an archival perspective, there is no requirement for a conforming reader to provide a delayer capability – one that can serialize, or otherwise reconstitute, the sequence of bytes that comprise the original embedded file.

There are other less purely "archival" use-cases envisioned for this "container" feature of PDF – including delivery use-cases where the source format of the information from which the PDF is constructed is, ironically, a more desirable format from an archival point of view (e.g. XML). While it is certainly desirable, from an archival point of view, that multiple manifestations of the information be stored in a single container, it is not at all clear that PDF/A-3 is an optimal choice as an archival container format. Nor, since the metadata ("file provenance") requirements in the proposal are minimal, and are restricted to the "wrapper" PDF/A-3 file itself, is it not clear that the requirements for a package manifest that a true archival container format should contain are being met. Further, nothing in the proposed standard in fact constrains the embedded content to “other representations” of the document image constituted by the PDF/A document instance.

We do not wish the best to be the enemy of the good. However, we feel that expanding archival PDF into a container for any arbitrary format instance elides the necessary hard work of specifying and meeting the requirements of an archival container format. One of the many reasons for the ubiquity of PDF as a document format was the hard work that was done in the original specification in anatomizing the significant properties of a page image format and in delineating and constraining the operators available for use in implementing the PDF image model. Over the history of the format, that core architectural metaphor of a page image along with the definition of “document” was elaborated to the point where a PDF application has begun to resemble a total environment - -much like a web browser—for all sorts of electronic behaviors and interactions. That such elaboration was problematic in any PDF instance intended to be accessible for the very long term was in fact the motivation for the PDF/A standard. Absent any reference implementation for “unpacking” a PDF container, and unless and until a sufficient anatomy of the archival requirements for a container format and its metadata is performed and the implementation requirements specified, we cannot support this proposal.