Long live ISO 32000-1

The PDF standard

by James C. King, Adobe Systems Incorporated

The portable document format (PDF) was invented by Adobe Systems Incorporated in 1993. Since that time, it has become the predominant file format for general electronic documents.

The success of this format is due to two facts:

- Any computer application that can print to paper can also create an electronic version as a PDF file
- Free PDF viewing software for Windows, Mac and Linux computers is readily available.

Today, billions of PDF files exist, as do thousands of applications written by hundreds of organizations. PDF became a *de facto* standard because it was so widely used.

Changing the status quo

Even though Adobe did not put any restrictions on other organizations with respect to creating software that views, creates or modifies PDF files, it kept ownership of the PDF Reference Manual[1] that defined the standard. Adobe "owned" PDF.

In January 2007, Adobe decided to change that, and asked the Association for Information and Image Management (AIIM), also known as the enterprise content management (ECM) association, to work together with the American National Standards Institute (ANSI) to submit PDF to ISO, with a view to its becoming a publicly-available International Standard. By January 2008, the project had been approved by the ISO membership, and the completed specification was published a few months' later as ISO 32000-1:2008, *Document management – Portable document format* – Part 1: *PDF 1.7*.

Trustworthiness and stability

So what difference has it made? The answer has to do with trustworthiness, stability, change and responsibility.

If an organization is going to invest in a given technology, it wants to be assured that its investment will be valuable in the future and not subject to the proprietary interests of some other organization.

Today, the PDF technology is captured in the ISO 32000-1 standard that is owned by ISO. Adobe is just one organization that builds PDF products based upon the ISO standard. This removes any risk that Adobe might move PDF in a direction that could be harmful to other organizations. It is not that Adobe is not trustworthy, but that ISO, being a neutral public organization, can be regarded as more trustworthy.

"The answer has to do with trustworthiness, stability, change and responsibility."

Once published by ISO, the standard is stable. Small mistakes and errors may be corrected but ISO 32000-1 as a whole cannot be changed. This brings a strong level of stability to PDF that frees people and organizations to invest in PDF documents and PDF software. It also enables exchange and interchange since everyone is working from the same common specification. A file either is, or is not, a

1) **www.adobe.com/devnet/pdf/pdf_reference.html**

PDF file according to ISO 32000-1:2008. Software is compatible with the standard if it can process PDF files correctly.

There are established and well-known procedures for introducing revisions and new editions of ISO standards. Open contributions from the member countries are solicited. Draft standards are created and circulated for review. Votes are taken. It is an open and transparent process. Anyone with a vested interest should be able to become a member of their national organization and take an active role in establishing a new direction for PDF.

"Once published by ISO, the standard is stable."

The other side of the coin of "ownership" is responsibility. From June 1993 until January 2008, Adobe was responsible for PDF. Today, PDF is ISO's responsibility. That means that if a revision is needed to change or add some new function to PDF, it is up to the relevant ISO technical committee – ISO/TC 171, *Document management applications*, subcommittee SC 2, *Application issues* – to propose and develop those changes, get them thoroughly reviewed, approved and published.

About the author

Dr. James C. King is a Senior Principal Scientist at Adobe Systems Incorporated and works in the San Jose, California, USA, Headquarters of Adobe as the PDF Architect. He has worked at Adobe for over 20 years starting the Advanced Technology Group in 1988. Before that, he was at IBM Research, both in Yorktown Heights, New York, and San Jose California. He is a member of ISO/TC 171/SC 2. Dr. King received a Ph.D. in Computer Science from Carnegie Mellon University.

Given the size and complexity of the standard, 756 pages of slow reading, this is no small undertaking. On the plus side, the "ISO" that now owns PDF, is not a faceless organization, but one composed of bright, technical people who are willing to spend time to understand, discuss, look at alternatives and help decide where PDF goes next.

A standard among standards

The ISO 32000-1 standard makes use of many other standards. Its normative reference section lists 79 standards documents upon which PDF is built, including eleven ISO standards such as JPEG, JPEG2000 and JBIG2 image formats (see **page 15**), ICC Profile formats and ASN.1 notation. Other notable standards used directly by PDF include the OpenType font format, XML (see **page 27**) and Unicode. When PDF was being developed, every effort was made to use existing standards.

Some ISO standards have been developed to define constrained PDF subsets for special applications. These include PDF/A for document archiving (ISO 19005-1:2005), PDF/X for professional publishing (ISO 15930 series) and PDF/E for engineering documents (ISO 24517-1:2008). All of these standards actually pre-date the publication of ISO 32000-1.

"ISO is not a faceless organization, but one composed of bright, technical people."

PDFs initial strength comes from its ability to capture digitally an exact replica of the document as it would have printed onto paper. This means that material that would invariably end up as some kind of paper document can instead be represented as a computer file and printed later or at some distant place. Or better yet, it can be viewed on a screen and never require paper.

This is one of the best examples of moving bytes instead of atoms as promoted by Nicholas Negroponte[2]. As useful and well-loved as paper and paper documents are, it is essential to have an electronic analogue. PDF serves that purpose.

Advanced capabilities

But once a document is available in an electronic form, there is an immediate desire to do more with it than could be done with the paper version. For example, electronically searching the document for particular words is more effective than scanning pages with our eyes. The introduction of hyperlinks that can immediately carry us from one page/view of a document to another, and back again, gives us a function that is unavailable or very clumsy with paper.

Conversely, since people make notes on documents published on paper, they similarly need to be able to annotate electronic documents. Functionality to support these examples – word/phrase search, annotations, hyperlinks – was added to PDF from an early date.

Advanced capabilities, well beyond these more obvious examples, are also part of ISO 32000-1. They include:

- Forms for interactive completion and submission
- Multimedia content, including video, Flash and sound
- Portfolios of files of any kind stored, indexed and accessible from within one PDF file
- Content layers
- 3D interactive rendering
- Digital signatures and digital rights management.

Documents plus

So not only does PDF allow us to have the exact representation of a document in electronic form, it also allows us to exploit the power of the computer to extend what we can do with those documents.

Long live ISO 32000-1 ! ■

2) Author of the book *Being digital* (1995); known for his forecasts on how the interactive world, the entertainment world and the information world would eventually merge.