

From PDF to PDF/X



Finding the right balance



ISO/TC 130, Graphic technology, develops standards for all aspects of graphic technology. With the strong involvement of computer technology in all sides of the preparatory part of the graphic technology industry, the current motivation for graphic technology standards is driven largely by two industry needs. These are standard file formats that enable the unambiguous exchange of print-ready material as electronic data and supporting standards that allow the

between
reliability,
flexibility and
application
complexity

By David Q. McDowell, Chair, ISO/TC 130, *Graphic technology*, WG 2, *Prepress Data Exchange*

The first question most people have is: "Why do we need more file formats? Aren't the formats out there already good enough?" The answer is both "yes" and "no". The most reliable file exchange format for graphic arts applications has been TIFF/IT (ISO 12639:1998, *Graphic technology – Prepress digital data exchange*

– *Tag image file format for image technology (TIFF/IT)*). It has served, and will continue to serve the industry well, but it is restricted to raster data. That means that all graphic objects and text must be converted to raster data before data can be exchanged using TIFF/IT. A standard file format is urgently needed that will support

meaning of that data to be fully defined.

This article focuses on a family of standards being developed within ISO/TC 130 that is revolutionizing the exchange of material between providers and publishers/printers. It is also unique in that it demonstrates a new style of cooperation between the owner of an industry specification and the activities of an international standards body to formalize the use of certain aspects of that specification.

object-based elements in addition to raster data.

As we look at the file formats available in the industry, we really have too many to choose from – most of them unfortunately unique to specific application programmes. PDF is a more general format, but has so many options that it is very

Help! How to find your way in and around the graphic technology world

Blind exchange: An exchange of printing information (digital data files) in which the sender and receiver do not require advance communication to properly process the information and produce the expected results.

CGATS: (Committee for Graphic Arts Technologies Standards): The standards committee responsible for graphic technology standards within the USA. NPES The Association for Suppliers of Printing, Publishing and Converting Technologies serves as the secretariat of this committee.

CMYK: The printing industry generally uses Cyan, Magenta, Yellow, and Black ink in combination to create the visual impression of full colour images.

Colour managed workflow: A method of exchanging image information where input and/or output device data spaces are related to a colorimetrically defined reference space through transforms carried in "profiles" (see ICC).

DDAP (Digital Distribution of Advertizing for Publications): An international industry group formed to support and publicize the use of accredited standards for the distribution of advertising in the publication segment of the printing and publishing industry.

FAQ (Frequently Asked Questions): Listings of frequently asked questions and their answers often found on Web sites.

ICC (International Colour Consortium): An industry group developing an architecture and data formats to enable open exchange of colour information.

ICC Device Profile: The collection of transforms necessary to relate device colour data of input, display and output devices to the colorimetrically defined profile connection space.

Object data: Imaging data that is created by a series of vector line segments and/or fill commands to create graphical elements, text, etc.

OPI (Open Prepress Interface Specification): A method by which the creator of a document will use a low resolution image as a place holder which will later be replaced by a image with sufficient resolution for quality printing. Image files with sufficient resolution for printing generally are generally about 60 kilobytes per square centimetre.

PDF (Portable Document Format): A format developed by Adobe Systems Incorporated that builds on the PostScript page description language and layers a document structure and interactive navigation features on its underlying imaging model.

PDF/X: A family of graphic technology standards that standardize the used of the Adobe PDF for eXchange of print-ready material.

PDF/X-1 (ISO 15930-1): Complete exchange of CMYK data.

PDF/X-2 (ISO 15930-2): Exchange of CMYK and/or colour managed data where some of the data required for final output (e.g. fonts, high resolution picture data, etc.) is uniquely identified but exchanged separately from the main body of the data exchange.

PDF/X-3 (ISO 15930-1): Complete exchange of CMYK and/or colour managed data.

Raster data: Imaging data that is organized into sequential lies of information that are displayed one following the other as in a television display.

RGB: Red, green and blue are the colours obtained from an image scanner and/or used to drive a computer monitor. They are also the colours used in many computer imaging and drawing packages.

TIFF and TIFF/IT: A file format that uses "tags" to carry information about the data content and structure of the file.

Transforms: See ICC Device Profile.

***Further information:** There are a number of useful Web sites available that provide additional information about PDF/X and related issues. ISO/TC 130/WG 2 and CGATS information can be found at www.npes.org/standards/workroom.html. The Digital Distribution of Advertising for Publications (DDAP) Association site, www.ddap.org, is also a good general source for information about TIFF/IT and PDF/X and their use in the United States publication marketplace. An excellent FAQ on PDF/X, by Martin Bailey of Global Graphics, can also be found on the DDAP Web site at www.ddap.org/resources/pdf_x_faqs.html. (Martin is the chair of both the CGATS and ISO task forces working on PDF/X.)*

difficult to be sure that the sender and receiver of a file have the same understanding of the options chosen and how to handle them. In addition, a basic premise of PDF seems to be that, if there are questions or problems, it is better to do something than nothing. Therefore, PDF by itself does not provide the level of reliability and predictability required for the graphic arts industry – particularly for catalogue and advertising material.

This is where the "/X" in PDF/X comes in. Although never formally defined, the "/X" to most of the people involved in the standards activity represents reliable eXchange. This is accomplished by identifying a limited set of PDF objects that may be used, and adding restrictions to the use, or form of use, of those objects and/or keys within those objects. The goal is that as long as both the sender and receiver of a data file use applications that meet the same PDF/X restrictions then the received file, when printed, will be exactly what the sender intended – no surprises and no detailed conversations to iron out the details!



The term "blind exchange" conveys some of the intent behind the development work. The belief is that, particularly for the placement of advertising, the prepress or preparatory shop that created the final advertisement file should be able to send the same file to anyone the agency has contracted with to print the advertisement with confidence that the printed results will be consistent. PDF/X is expected to become an electronic

substitute for the distribution of film separations that worked so well for so long. The same philosophy applies to the distribution of any print-ready work – not just advertising material.

A single format does not meet multiple workflow needs

When the several task groups in ISO/TC 130 were working on their individual parts of the PDF/X family of standards, we realized that we needed a common set of guidelines and ground rules. In addition, we believed that there would be occasions when national bodies and industry groups will want to prepare their own specifications for restricted subsets of PDF for very specific situations – too narrow in scope to justify an International Standard. We wanted to be sure that it would be as easy as possible for all such specifications to adhere to the same guidelines and principles and were as compatible as possible.

Therefore, the PDF/X family of standards includes a standard on how to write PDF/X standards. It is ISO 15929:2002, *Graphic technology – Prepress digital data exchange – Guidelines and principles for development of PDF/X standards*.

The rest of the current ISO PDF/X family is contained in the multi-part standard ISO 15930, *Graphic technology – Prepress digital data exchange – Use of PDF*. There are currently three parts in preparation. These are:

- **Part 1:** Complete exchange using CMYK data (PDF/X-1 and PDF/X-1a),
- **Part 2:** Partial exchange (PDF/X-2), and
- **Part 3:** Complete exchange suitable for colour managed workflows (PDF/X-3).

A common question is “why so many different parts or options?” All graphic arts electronic data exchange is a compromise between three issues: reliability, flexibility, and application complexity (and cost). If we had only a single format specification for the most reliable and simplest exchange, it would not meet many real workflow needs. Conversely, a very flexible all-in-one approach would put a burden

“PDF by itself does not provide the level of reliability and predictability required for the graphic arts industry – particularly for catalogue and advertising material.”

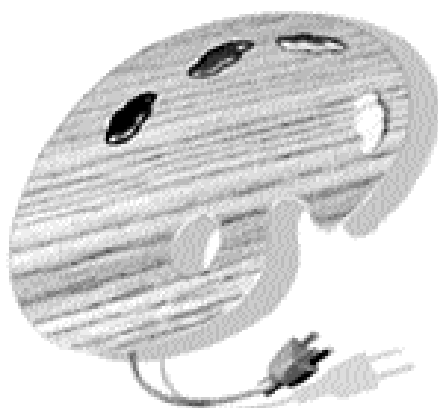
on many application areas that do not want to pay for features that they will not use. A good example of this is the support for OPI (Open Prepress Interface Specification – a method used to reference external files for substitution at time of final imaging). Some markets absolutely insist that it is a required feature, others say that requiring the inclusion of the OPI capability will make the receiving applica-

tions too costly for widespread use. The nub of the problem is this: if PDF/X is to be successful, every receiving software application must be able to read and appropriately process all conforming PDF/X files as defined for that compliance level.

Avoiding conflicts of the basic functions

A key requirement of ISO 15929 (or Base PDF/X) is that: “The PDF file format, as defined in the Portable Document Format Reference Manual, shall not be violated.” It further states that: “PDF/X standards should be developed as far as is possible without adding extensions to the PDF file format upon which they are based.” When it is found absolutely necessary to extend beyond the relevant version of PDF, the following guidelines shall be taken into account:

Open exchange of colour information



The International Color Consortium (ICC) was established in 1993 for the purpose of creating, promoting and encouraging the standardization and evolution of an open, vendor-neutral, cross-platform colour management system architecture and components. The outcome of this cooperation was the development of the ICC profile specification and the architecture upon which it is based.

The basic concepts of this architecture involve the use of transforms to relate colour data of input, display, and output

devices to a colorimetrically defined reference profile connection space (PCS). These transforms can be designed to preserve the image colorimetry between the PCS and either input or output device space, or they can be designed to perceptually relate device space to the PCS in order to preserve the “appearance” of the image (referred to by the ICC as colorimetric intent or perceptual intent). All of the various transforms necessary to relate colour data in a particular device space to the PCS are carried together in a device “profile”.

It is important to note that the PCS is not used to store data but rather to “connect” the profiles of input and output devices. This allows device profiles prepared in accordance with this specification to be used in pairs to translate colour data created on one device into another device's native colour space.

The current version of the ICC File Format for Color Profiles may be found at www.color.org along with a number of papers and articles that more fully describe colour management and the work of the ICC. It is expected that the specifications being developed by the ICC will eventually be brought into ISO when the technology is more fully developed.

PDF vs Acrobat and Distiller

Many people confuse the Adobe Systems Incorporated products, such as Adobe Acrobat and Adobe Distiller, that use the Adobe Portable Document Format (PDF), with the PDF file format specification itself. They are two different things. Both Acrobat and Distiller are based on the PDF file format specification and read or create many different types of PDF files whose features depend on the particular settings used. The Adobe Portable Document Format itself is separately documented, requires no license or fees, and may be used by anyone. The PDF specification may be found at

<http://partners.adobe.com/asn/developer/acrosdk/DOCS/PDFRef.pdf>

- Care should be taken to avoid the mechanisms and constructs used in a part of ISO 15930 coming into conflict with the mechanisms and constructs that might be expected to be added to a future version of PDF to provide the same functionality.
- Extensions to PDF shall be constructed in such a way that the resulting file may still be read without error by a PDF compliant reading application that is not also compliant with the part of ISO 15930 standard being developed, and by reading applications compliant with other parts of ISO 15930.
- Wherever possible the extensions should be specified in such a way as to maximize the probability that they may be implemented using plugins or other minor extensions to existing mainstream PDF applications.
- Extensions that will cause a significant difference in output when processed by a PDF/X compliant application as opposed to a baseline PDF application should only be added after conclusive discussion and shall be clearly identified in the defining part of ISO 15930."

The technique used to enable these requirements has involved cooperation between Adobe Systems Incorporated and

the standards activities to identify (and justify) additions to the PDF file format needed by the printing and publishing industry. These are both documented in Adobe Technical Notes to allow immediate use and added to future editions of the Adobe PDF specification. For both the standards community and Adobe Systems Incorporated, predictability of results in a printing and publishing environment is critical.

As a mechanism to allow identification of add-ins or extensions to the PDF file format, Adobe Systems Incorporated has created a series of name registries. "GTS_" is a second class name prefix assigned by Adobe Systems Incorporated to the United States ANSI Committee CGATS (and its secretariat, NPES The Association for Suppliers of Printing, Publishing and Converting Technologies). The GTS_ PDFX name prefix is an extension of this prefix adopted by the ISO/TC 130 committee and defined in ISO 15929 to allow unique identification of any PDF/X file format specification.

As part of ISO 15929, NPES The Association for Suppliers of Printing,



With this new suite of standards the opportunity exists for a variety of applications based on standard exchange formats to meet virtually every publication and commercial printing workflow need in the printing and publishing industry.

Publishing and Converting Technologies has agreed to make the use of the "GTS_ PDFX" name prefix available to all industry and standards groups within the worldwide graphic technology industry and to maintain an open registry of all such uses.

About the author, David McDowell



David Q. McDowell is retired from the Professional Imaging Division of Eastman Kodak Company where he worked for 42 years. As a Senior Technical Associate,

he represented Kodak in a variety of standards and industry activities, and within Kodak provided broad technical support to product development activities that involved or impacted the graphic arts market.

In retirement, he is continuing his involvement with Kodak in the area of colour and imaging standards, on a part-time basis. He is also working as a volunteer with NPES The Association for Suppliers of Printing, Publishing and Converting Technologies, the graphic arts trade association that provides coordination of all United States graphic arts accredited standards activities, to further the development of graphic arts standards.

David McDowell is the Chairman of the ISO Steering Committee for Image Technology (SCIT), Chairman of the US Technical Advisory Group (USTAG) to ISO/TC 130, *Graphic technology*, Chair of TC 130/WG 2, *Prepress Data Exchange*, Chair of ISO/TC 41/JWG 21, *Revision of ISO 5 Densitometry Standards*, Chair of CGATS/SC 8, *Color Data Definition*, and Secretary of CIE Division 8, *Image Technology*. He was a key participant in both the creation of the ANSI accredited standards effort within the United States graphic arts industry and in the reactivation of ISO/TC 130, *Graphic Technology*.

An earlier version of this article appeared in *The Prepress Bulletin*, the journal of the International Prepress Association, and their assistance is much appreciated.

ISO 15929:2002 is due to be published shortly.

When the sending and receiving organizations use "standard" applications, blind exchange works!

ISO 15930, Part 1 (PDF/X-1 and PDF/X-1a) is largely intended to meet the needs of the traditional publication marketplace – in essence a digital equivalent of film separations. To quote directly from its introduction:

"Part 1 of this International Standard defines a data format and its usage to permit the predictable dissemination of a compound entity to one or more locations as CMYK data, in a form ready for final print reproduction, by transfer of a single file. This file must contain all the content information necessary to process and render the document, as intended by the sender. This exchange requires no prior knowledge of the sending and receiving environments and is sometimes referred to as "blind" exchange. It is platform and transport independent.

"These goals are accomplished by defining a specific use of the publicly available Adobe Portable Document Format as specified in Version 1.3. In order to achieve a level of exchange that avoids any ambiguity in interpretation of the file, it identifies a limited set of PDF objects that may be used and adds restrictions to the use, or form of use, of those objects, and/or keys within those objects. It includes two compliance levels, PDF/X_1 and PDF/X_1a, that differ only in their allowed use of OPI references and encryption, which are allowed in PDF/X_1 but not in PDF/X_1a."

The term "compound entity" is defined as a "unit of work with all text, graphics and image elements prepared for final print reproduction and may represent a single page for printing, a portion of a page or a combination of pages".

PDF/X-1a is essentially a simplified version of PDF/X-1 and, therefore, a PDF/X-1 receiving application will be able to easily receive both PDF/X-1 and -1a files. For smaller and/or simpler installations PDF/X-1a may suffice. In all cases the writing applications only need to support those features in the standard that the user requires.

ISO 15930-1 is derived from ANSI CGATS.12/1 - 1999, *Graphic technology – Prepress digital data exchange – Use of PDF for composite data Part 1: Complete exchange (PDF/X-1)*, which was based on Version 1.2 of the Adobe PDF specification and which included OPI as a required capability. However, as the ISO version was being prepared, it was clear that it needed both to be compatible with Version 1.3 of the PDF specification and to include a compliance level that did not require support for OPI.

ISO 15930-1:2001 has been published.

Exchanging "colour-managed" information

Let's jump to PDF/X-3.

PDF/X-3 has many things in common with PDF/X-1a. They are really sisters. The one big area in which they differ is that PDF/X-3 allows the use of colour-managed three-component data in addition to CMYK data.

What does "colour-managed" data mean? One of the ways in which the PDF/X family differs from plain old PDF is that all data in a set of files being exchanged must be colour-corrected and adjusted for a single characterized printing condition prior to exchange. That means that all process

colour data must be based on the same inks, densities and dot gain. In PDF/X-1, allowing only CMYK data was one step taken to make this more predictable. In addition an ICC (International Color Consortium) input profile (device-to-PCS transform) had to be included, or referenced, to define the particular type of CMYK printing.

Why standards will be one step behind

I am sure many are saying: "Why can't the standards committees keep more current with the Adobe PDF file format specification releases?"

Therein lies a problem. Because the Adobe PDF file format specification is a private specification (albeit documented and publicly available), Adobe has chosen not to make the various versions of the specification publicly available until Adobe products implementing the specification are available in the marketplace. Although many developers (even those participating in the standards process) have signed non-disclosure agreements and have advance insights, the whole concept of non-disclosure runs counter to the open standards process. As long as companies maintain such policies, the standards will be one step behind. For example, although Acrobat 5 is based on Version 1.4 of the Adobe PDF file format and is currently being sold, an Adobe Technical Note describing the changes between Version 1.3 and Version 1.4 was only recently made available. To wait for final documentation to allow incorporation of Version 1.4 would have delayed the standards approval process significantly.

"The prepress or preparatory shop that created the final advertisement file should be able to send the same file to anyone the agency has contracted with to print the advertisement."

In PDF/X-3 the standards committee got braver and agreed to allow the inclusion of an ICC output profile (PCS-to-device and device-to-PCS transforms) to define the relationship between three-component data and the data to be used for final rendering. (See box p. 8 on the International Color Consortium.) While this is typically CMYK data for printing and publishing applications, this mechanism allows any output colour space to be used; for example web RGB, some special RGB or CMYK



The choices available to users will depend, in part, on the vendors and the applications they choose to develop. As experience and confidence develop, the use of the various versions of PDF/X is probably going to become very mixed and job-dependent. Ideally they will be like flavours of ice cream – mix and match as taste (or need) dictates.

Photo Courtesy
International Prepress
Association

printing device, photographic output, etc. The requirement still exists that all data has to be prepared for the same "printing" condition so that all data in the files being exchanged will work together. The recent Adobe Technical Note (No. 5413) Recording Output Intentions for Color Critical Workflows describes the use of the OutputIntents array in the Catalog object as the way that this information is carried in a PDF file (another example of the way in which Adobe has worked with ISO/TC 130 and the standards community to document new features in PDF which can then be used in PDF/X).

ISO 15930-3 (PDF/X-3) has passed its first ISO ballot (Committee Draft) and is in ballot as a Draft International Standard (DIS). ISO 15930-3 (PDF/X-3) will be published in 2002.

The problem of font identification

PDF/X-2 is both the most difficult and potentially the most useful of the current PDF/X family. It is based on PDF/X-1a and PDF/X-3 and will be compatible with both. Both CMYK data as well as properly characterized three-component data are sup-

ported. However, as its title says, it also adds something called "partial exchange".

That word "partial" can be misleading. What it really means is that in certain workflows some or all of the referenced elements may be more logically present at the receiving site, or may be exchanged at a different time. These include fonts, high resolution contone image files, or line art files. What makes this difficult is that the scope currently includes the statement "where all elements necessary for final print reproduction are either included or uniquely identified". It is that "uniquely identified" that has been causing problems.

The committee believes that predictable results is the real requirement for reliable data exchange. Finding the best compromise between the three issues of reliability, flexibility, and application complexity is the issue. Because a PDF/X-2 exchange is defined to require communication between sender and receiver (it is not a "blind exchange") some additional flexibility is available – but not much.

A good, and the most difficult, example is fonts. While PDF/X-1 and PDF/X-3 require that all fonts used are embedded, there are some fonts whose license does not permit

embedding (or, for some Asian fonts, are too large to embed practically). PDF/X-2 would not require embedding, just unique identification. However, how does one ensure that the sender and receiver have exactly the same version of a font, including any edits that might have been made to font metrics? Most of the tools to accomplish such checking are based on proprietary technology and require licenses or fees for use. To require use of such proprietary technology in a standard is a last resort.

ISO 15930-2 (PDF/X-2) is being developed jointly by ANSI CGATS (Committee for Graphic Arts Technologies Standards) and ISO/TC 130. Subcommittee 6 Task Force 1 of CGATS is taking the lead and developing the details that are being reviewed by a Task Force in TC/130/WG 2. As soon as a satisfactory draft is available it will be placed in ballot as an ISO Committee Draft (CD).

Where will different versions of PDF/X be used?

The choices available to users will depend, in part, on the vendors and the applications they choose to develop. With this new suite of standards the opportunity exists for a variety of applications based on standard exchange formats to meet virtually every publication and commercial printing workflow need in the printing and publishing industry. However, user commitment and vendor implementation is required to make it happen. TIFF/IT will also continue to be useful in situations where the simplicity and reliability of raster data is the preferred format.

PDF/X-1a will probably be the preferred choice for most of the publication markets. Initially, the principal users of PDF/X-3 will probably be found in the newspaper area and possibly used for individualized printing using digital presses. The catalogue and commercial printing areas will benefit most from PDF/X-2. However, as experience and confidence develop (and as various application programmes become available), the use of the various versions of PDF/X is probably going to become very mixed and job-dependent. Ideally they will be like flavours of ice cream – mix and match as taste (or need) dictates. □