TECHNICAL GUIDE

PREPRODUCTION DATA MANAGEMENT

Second Edition

FESPA
profit for purpose

Introduction to preproduction data management

Printing is all about delivering data for output, so preproduction data management is as important for graphics companies as all other management activities.

Data manipulation in the graphic arts industry enables new production models, including variable data output across different delivery platforms and media. Image and document data is initiated, edited, managed and delivered digitally to electronic screens and to physical substrates, so critically, effective data management influences profits.

The industry's focus on data is part of a tradition that started with the advent of electronic scanning to capture colour originals. Introduced in 1969, John Crosfield's Magnascan, with tools for image retouching, signaled how digital data processing would change print media production workflows.

In 1984 the introduction of the Apple Macintosh desktop computer, the Apple LaserWriter printer and the Adobe PostScript page description language brought composition to the desktop. The desktop publishing revolution brought device independent integrated text and graphics datastreams to printers and publishers, and to a new generation of developers. For the first time lay users had control over composition

and output choices. They could design and lay out their own pages and an army of entrepreneurial developers rose up to serve them. This surge in innovation continues to reinvent how graphics data is processed.

The PostScript page description language is still used in the market however Adobe's Portable Document Format (PDF), introduced in 1993 and an evolution of the PostScript language, now dominates. Digital print production workflows are increasingly automated, sophisticated, fast and reliable. Since the introduction of digital colour printing in 1993, the range and capabilities of such systems has grown. This technology can print everything from labels to books, from newspapers to building wraps. Managing data effectively for all these print applications depends on data management processes that work in all print applications, from sign and display work through to commercial print and publishing.

A Disconnect Overcome

Understanding how data works and understanding how to fully exploit a printing press are not the same thing. Yet successful print media production companies combine precisely these skills. They mix information technology and computing knowledge, with manipulations of inks, substrates and press behaviours using colour data management expertise. Such companies exploit the power of data to control complex workflows using process automation, enhancing job margins and company profits.

Digital data is a series of electronic signals with only two states: on or off, like a light switch. A series of binary integers, or bits, is configured into sets of eight bits, or bytes, which historically provided enough on/off combinations to describe a single character. It's got more complicated over the years, but in essence computer processors represent information and instruction sets as multiples of bits and bytes: 8-bit, 16-bit, 32-bit or 64-bit. For processing speed and efficiency, the more bits the better because this determines how quickly computers process data. Network bandwidth, is defined in bits per second, so the bigger the channel, the faster data is delivered.

The Central Processing Unit (CPU) is a computer's brain, and having sufficient processing capacity is especially important for printing

applications. Printing and publishing processes are data intensive, so they require lots of processing power. Read Only Memory (ROM) is where hard-wired instructions, such as those relating to the operating system, are processed. Processing instructions for different software applications (apps), for example for image processing or email, happen in Random Access Memory (RAM). The CPU accesses RAM directly and much faster than is possible from a hard drive, so the more RAM available the better. The alternative is grindingly slow instruction processing, so it's common to have 16 Gigabytes (GB) of RAM. A GB is 1000 million bytes, which is 8,000 million bits. If your prepress workflow is struggling to keep up, you may need to go no further than upgrading the RAM in your computers.

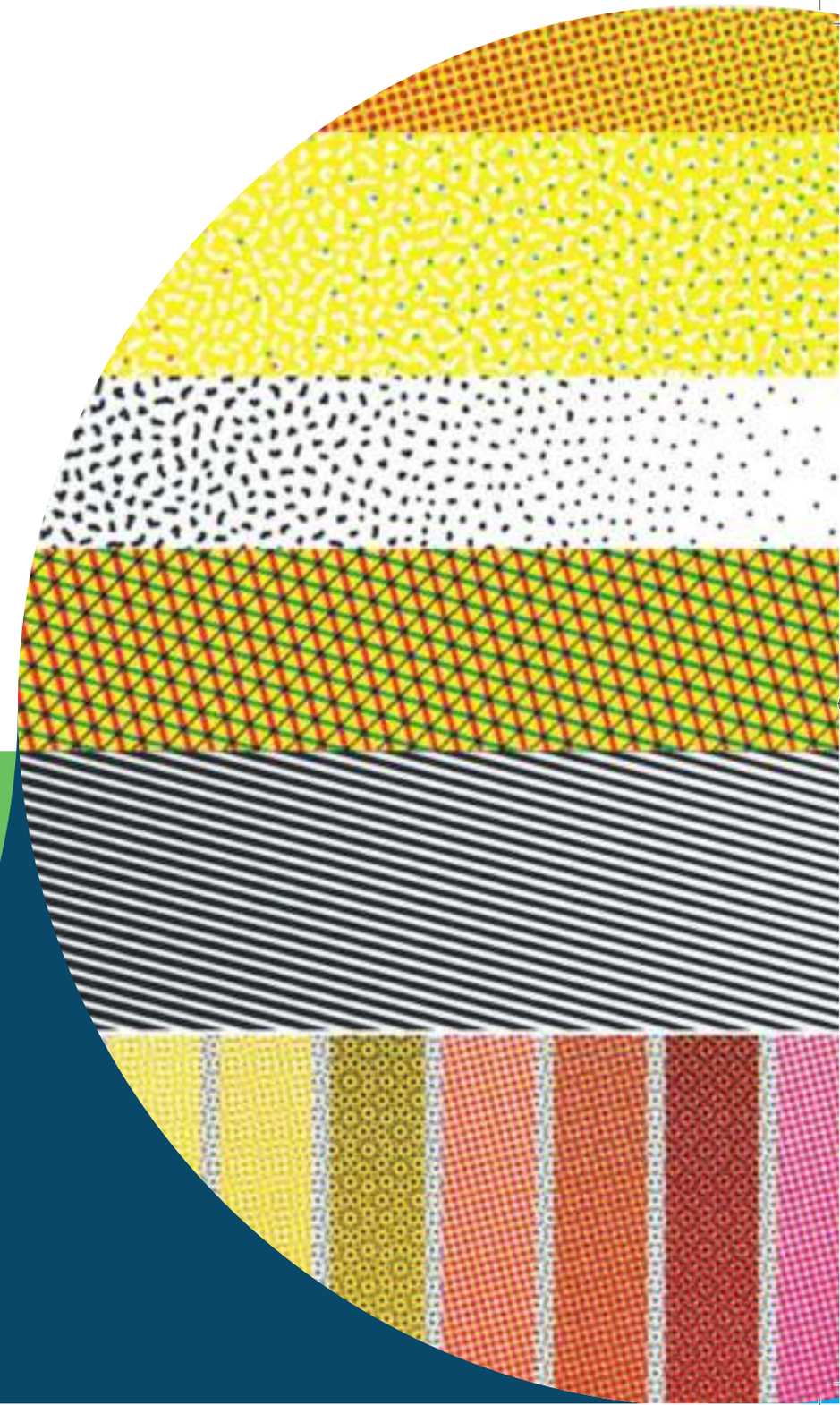
Processing speed is only useful if it receives and executes instructions efficiently. The Operating System (OS), for instance the Mac OS or Windows sits in ROM between the CPU and everything else, including you. It provides common services to software and hardware, including the screen.The OS is an intermediary, matching software and hardware requirements to the computer's processing resources. Having enough RAM for the business need and a robust operating system influences productivity. Operator efficiency and the performance of apps depends on the OS, its stability and consistency and the intuitiveness of the user interface.

Mac Attack

The Apple Mac was for many years the preferred choice for graphic arts professionals and, following a few years in the Windows wilderness, the Mac is increasingly popular for both design and production. Apple has pioneered colour management and its technology is known for reliability and processing consistency, both of which are profoundly important for graphic arts applications. The technological foundation used for production tasks and workflow management influences preproduction data management policies, so they must have the scope to support business and operations expectations. This includes such things as links to accounting and related Management Information Systems (MIS). Investment and upgrade decisions should take into account how the digital infrastructure is expected to perform and evolve, and how it will support the printing systems fleet.

In addition to the OS, apps, which appear everywhere from your desktop to your mobile 'phone, must adequately support the workflow. They increasingly reside in the cloud and include business tools such as spreadsheets, presentation tools and word processing. For the graphic arts community Adobe Creative Cloud is the preferred apps suite. This subscription based set includes PhotoShop, Illustrator, InDesign, and Acrobat. Other common softwares include QuarkXPress and Corel Draw, plus preflight tools such as Enfocus Pitstop, and production management technologies from manufacturers such as Agfa, Fujifilm and HP.

Workflow systems can be extremely complex with multiple components such as RIPs and colour servers, plus associated apps. They operate in diverse environments supporting PCs, Macs, media servers and mobile devices. Online and local servers provide dedicated functions such as colour management or RIP-ing, and are accessible to a network of desktop devices. They can drive multiple engines from a single system and may reside in the cloud where they link to other remote systems.



Raster Image Processors (RIP)

The heart of any production system is the RIP which converts data into bitmaps. RIPs turn electronic descriptions of output pages into a series of precisely defined dots. With device profiles, RIP software takes into account the output device's particular characteristics for further precision. Suitably configured, a single RIP system can drive individual or multiple output devices, and manage load balancing across output devices.

At Your Service

The traditional client/server architecture used in the graphics industry is giving way to cloud services accessed from the desktop and via mobile devices. Client/server architectures comprise co-located networked devices that provide or consume dedicated services, such as managing file servers, printers or other networks. They provide shared computing resources and make available common resources or additional processing power. They can be used for dedicated applications such as file delivery as part of an online workflow system or for load balancing, supplementing processing resources for data intense apps such as colour management.

Cloud computing takes this model one step further, with apps accessing remote servers managed offsite via the internet or an encrypted Virtual Private Network (VPN), supporting mobile and desktop devices. These servers can be owned or managed by third parties which commit to provide software and hardware support under license. A cloud can be public or private, or a mixture of the two and they can be very large, such as the Google cloud, or very small such as proprietary web-to-print systems use to provide print on demand client services.

Naming Conventions

A computer at the desktop, on a 'phone or tablet, or online in the cloud, is a digital filing system. It's made up of datafiles stored in sub-folders and folders. Just as conventional filing systems can comprise many drawers and cabinets full of folders, a digital system can be very extensive with complex file hierarchies and access privileges. Just as a filing system in your office needs organization to be useful, digital systems require logic to identify and source the things they store. Software developers use suffixes to identify file types and to instruct OSs for opening a file. Mac OS is very good at knowing how to open different file types, however Windows needs to be told what something is, in order to open the file. If a file doesn't end with .docx or .xlsx or .pdf for instance,

Windows may struggle to open it. In automated workflows, the system relies on naming conventions to know how to process the file. The file's name is a unique identifier, and can be used to route the file automatically through the workflow to final production.

The filename can also prevent the file from being lost or overwritten in error and makes file retrieval and processing easier. Naming conventions can be set up so that the file name instructs the processing software as to where to send it and what to do with it. File naming conventions for prepress production workflows can be assigned according to the range of work produced, in line with degrees of automation and customer expectations. The naming system should include supported formats and output resolution requirements. A directory structure helps manage different file types for different customers, storing files as digital assets for subsequent use. Customer numbers or workflow type can be assigned within the file name, so that files are automatically routed for the correct output path via hot folders.

Data Formats & PDF

File names often include a suffix to identify their data format. Data formats standardise data encoding, so that different software applications know what to do with the data. Both standard and proprietary data formats exist, and the printing industry uses many of both types.

No data format is as important for the graphic arts as the Page Description Format (PDF). PDF has its origins in the Adobe PostScript language which describes the content of a page, so that the data from which it is made can be interpreted and used to render bitmaps for CMYK separation sets. But PostScript code describes the appearance of an integrated text and graphics file in a single entity, so PostScript files can be huge, difficult to edit and slow to process. Contained in a PDF, the content is a list of page components and attributes, a database, so output files are smaller and the data is easier to access and manage.

PDF 2.0 (ISO standard 32000-2) published in 2017, includes important features for graphic arts professionals. It supports Open Web Platform (OWP) technology which simplifies turning PDFs into HTML for the web. Security features provide data encryption and support for digital signatures. Black Point Compensation for graphic components in a document can be specified in PDF 2.0. Output intents, already a strong feature of PDF, support spectral data and the Colour Exchange Format (ISO 17972).

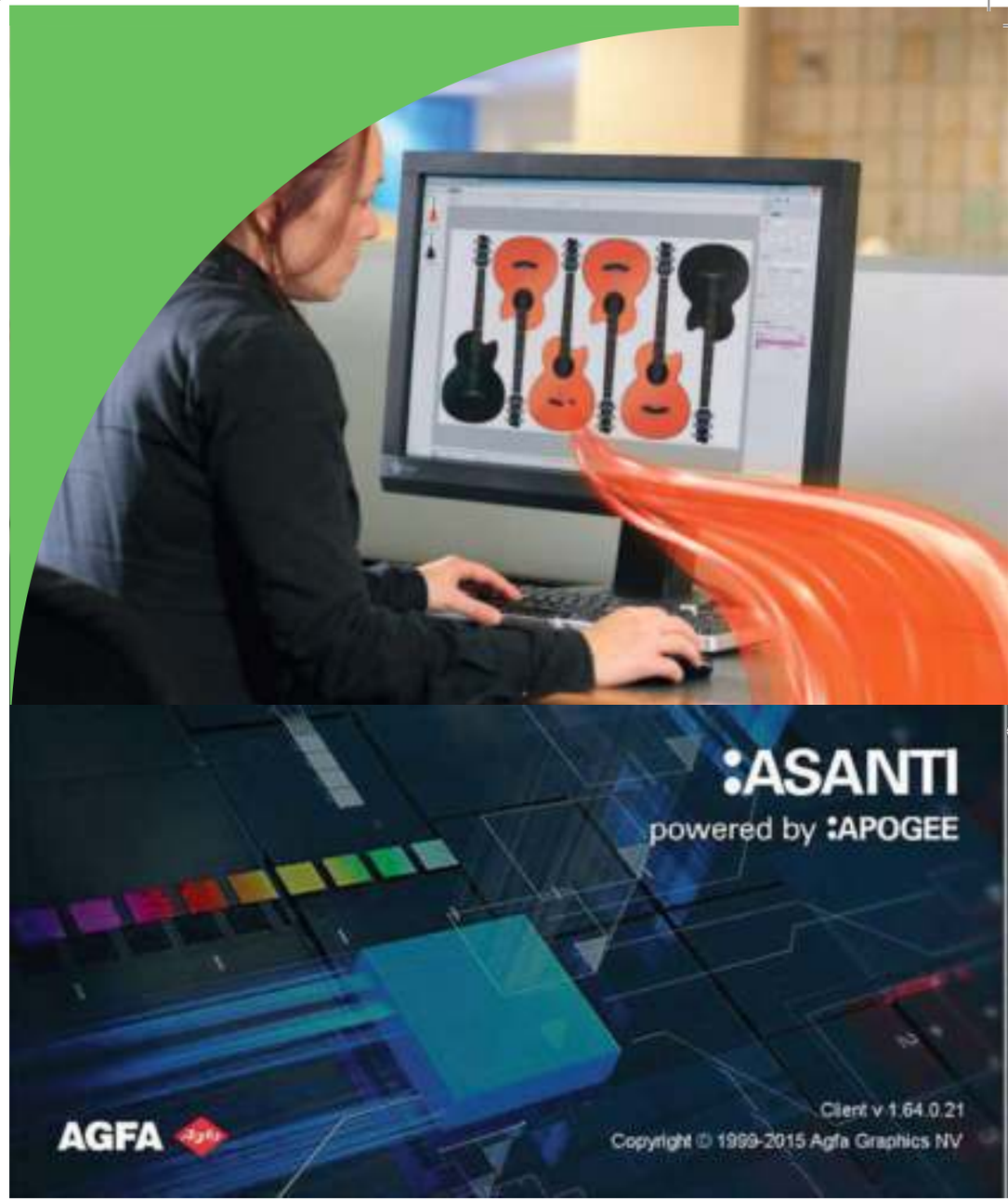
PDF benefits from a lively development community, which produces a huge range of tools for PDF processing. They range from automated file delivery and preflight checking, through to conversions to HTML.

Adobe PDF Print Engine (APPE)

PDF is page independent, lean and flexible in ways that are impossible with a full blown page description language. Consequently graphic arts workflow and data processing systems are generally based on the Adobe PDF Print Engine (APPE), rather than the PostScript language. APPE offers flexibility, processing speed and an architecture that is very developer friendly.

APPE is a native PDF RIP that Adobe provides as a Software Development Kit (SDK), so that a single platform can render input and output PDFs and the job content they contain. PDF library instructions handle data rendering to output bitmaps for greater rendering speed and control. This is especially important for processing colour data and variable data, both of which can be complex. But APPE is also designed for distributed data processing, particularly for complex applications combining multiple data streams for a single output path.

POWERED BY
Adobe PDF Print Engine



Agfa was the first company to build a workflow system based on PDF. Apogee, widely used and robust, is the foundation for Agfa Asanté a popular workflow system for wide format digital printing.

About PDF/X

The PDF/X series is a collection of ISO standards based on PDF. It was created to constrain PDF file content for file quality control and process automation in the graphics business. Collectively known as ISO 15930, the various parts of the series define different expectations for PDFs entering print media production workflows.

Work on improving PDF file exchange began soon after the format gained acceptance within the graphic arts. PDF has huge functionality, but much of it is irrelevant for printing applications, so controls on how PDFs are created for print were needed. ISO 15930-1 is based on PDF version 1.3 and came out in 2001. PDF/X-1a allows for blind PDF file exchange with specification of the output intent. All fonts are embedded and colours are described as CMYK or spot colours. Components must be identified as untrapped or trapped. In 2003 ISO published ISO 15930-2. PDF/X-2 was based on PDF 1.4 but the market largely ignored PDF/X-2 and it was eventually withdrawn.

PDF/X-3 allows CIE La*b*, calibrated RGB, spot colours and CMYK, and embedded ICC profiles, so it was relatively popular. It was replaced in 2008 with PDF/X-4 based on PDF 1.6 with support for layers and transparency, which soon became the preferred format for file submissions. Extensions to PDF/X-4 include PDF/X-5g for including external graphics, and PDF/X-5p for including an ICC profile with an external file, and PDF/X-5n that allows that ICC profile to use alternative colour spaces. PDX/5 has not been widely adopted. ISO is currently working on a replacement version of PDF/X-4 based on PDF 2.0, PDF/X-6 expected to be published in 2020.

Controlling PDF settings and robust preflight checks reduce the risk of error. Outputting to a wide format digital device such as this HP Scitex 9000 has to be right first time, to avoid costly remakes.

PDF/X-6, ISO 15930-9, defines parameters for complete and partial data exchange taking advantage of the page level intents (printing different pages on different devices) included in PDF 2.0. There are two sub-versions. PDF/X-6p applies for documents that include an external output intent ICC profile and PDF/X-6n is used for n-colorant profiles, so that complex colour combinations can be referenced.

PDF/X-6 requires that Black Point Compensation is enabled, if there is no setting for it in the PDF. CxF encoded spectral data can be included in PDF/X-6 compliant files, as long as it doesn't conflict with the output intent ICC profile. It is also possible to include annotations, as long as they conform to PDF/X requirements.

PDF/X-6 is also the basis for the latest variable data iterations of PDF/X, PDF/VCR-1 and PDF/VT-3. PDF/VCR-1 (ISO 16631-1) for variable content replacement, is for high volume applications such as personalized direct mail. It is essentially a template with placeholders, into which variable data can be merged. PDF/VT3 (ISO16612-3) has rules for how to deal with transparency and repeated contents for more streamlined data processing.

Variable data printing

Variable data processing technology makes possible completely different content on each page in a digitally printed run. It is used to produce customised documents such as travel itineraries including luggage labels and excursions vouchers, or bespoke car and property brochures. Variable data printing's greatest advantage is that it can generate accurate high quality variable data colour output for high value print. Wedding photo albums, commemorative print, localised posters and rapid turnover instore signage can all be produced using PDF/VT. This technology can also be used in other ways, such as for logistics management. Using the variable data capabilities in a PDF/VT workflow it is possible to organise output so that print is produced to suit complex installation logistics for signage work. Using PDF/VT a production run of multiple prints localised for different installations, can be set up to cut turnaround times so that work is dispatched starting with the most remote installation site.

AFP/IPDS

An alternative means of producing variable data documents and one widely used in the transactional market is AFP/IPDS, the Advanced Function Presentation-Intelligent Printer Data Stream format. Like PostScript, AFP/IPDS is a programming language, but there any similarities end. AFP/IPDS is used to instruct high speed, high volume data stream creation. It creates page instances, the content of which disappears as each page is printed, keeping the data totally secure. The content exists only on the printed page. The format dominates the transactional market and is being developed for improved colour handling as colour becomes more important for transactional applications. AFP/IPDS is not widely used in commercial printing, where variable data printing only accounts for a small percentage of printed pages and where PDF is the preferred format. But it is now possible to embed containers for single or multi-page PDFs in an AFP data stream. ISO 22550 (AFP interchange for PDF) is in its final stages of development and is expected to be published by 2020.

Widely used for transactional applications, this printer can process both PDF and AFP/IPDS datastreams. ©Ricoh Infosystems

Common data file formats

PDF	Compact and flexible file format for document exchange and print production. Supports image compression, but in some cases this can result in jagged edges in high quality images. It's better to adjust image size in Photoshop and if necessary apply the compression there, not when exporting to PDF.
EPS	Basically Postscript and can only be saved as one page per file. Not used much anymore in print production.
PSD	The Adobe Photoshop native file format. Supports layers and colour management, but might not be supported directly by all RIPs. Can however be embedded in designs and PDFs if they are generated by Adobe InDesign or Illustrator.
TIFF	Tagged Image File Format is a good rasterized format for single images, supporting ICC colour management but files may result in large files unless saved with some type of image compression, like LZW, ZIP or JPEG.
AI	AI Adobe Illustrator is a Intended for creation of vector graphics with infinite scaling but not supported directly by RIPs. They can however be embedded in PDFs, for example generated by Adobe InDesign.
PNG	PNG Portable Network Graphics is Often found recommended for DTG (direct to garment inkjet printing) as it supports transparency. Does however not support ICC colour management.

Other important formats

The eXtensible Mark-up Language (XML)

XML is a programming language with origins in typesetting mark-up. It defines rule sets for presenting information, adding tags to content, so that it can be processed according to associated instructions. XML, in part based on the Standardised General Mark-up Language (SGML), was originally developed for high volume electronic publishing and can describe pretty much any digital entity because it can represent arbitrary data structures. The language is also a powerful tool for telling different systems how they should treat a given datafile or content, so it is very popular with web services. Most software developers in the graphic arts have application programming interfaces that can process XML data, and XML is the backbone of the internet.

XML provides the graphical user interface on the internet that is the Worldwide Web. However internet interfaces are increasingly based on the JavaScript Object Notation (JSON), a lighter weight iteration of XML initially intended as a subset of JavaScript, a scripting language. JSON is a language independent data format used for data interchange, but without the oomph and overhead of XML.

Several other formats are important for graphic arts data management: HTML5, the Job Definition Format (JDF), the Colour Exchange Format (CxF) and the eXtensible Metadata Platform (XMP).

The Hypertext Mark-up Language (HTML)

HTML is probably the most important SGML application of all. It is used to structure information so that it can be presented on the Worldwide Web. The latest version is HTML5 which supports all previous HTML versions plus audio, video and all forms of multimedia and graphical content.

The Job Definition Format (JDF)

When JDF was introduced in 2000 it was considered the great saviour for print workflow automation. The format is an implementation of XML that specifies how to route files through workflows, initiating tasks such as PDF creation and colour data conversions along the way. It was designed for information exchange for processing job data efficiently, using metadata to

inform downstream tasks. For instance customer data from production can be added to an accounting system very easily, if both systems understand JDF. Achieving this however has been the biggest impediment to adoption of the format. Interoperability between technologies needs to be defined, which requires development effort. For many companies XML is easier to implement than JDF, although those companies who have gone down the JDF route claim excellent financial returns. ISO 20616 formalises the file format and is expected to be published in 2019.

The Exchange Job Definition Format (XJDF) introduced in 2018, is a means of overcoming some of the interoperability concerns associated with JDF. It makes it easier for diverse systems to cooperate, providing technology that connects stuff together for workflow automation.

The Colour Exchange Format (CxF)

CxF as specified in ISO 17972 is for sharing colour measurement and associated metadata electronically. The standard has several parts to suit different workflows, especially packaging production using spot colours. CxF can also be used to share colour information beyond print workflows, for example to electronic media systems, so it is very useful for ensuring colour appearance accuracy across media.

The eXtensible Metadata Platform (XMP)

XMP is a tool for embedding data about a resource, its metadata, into a document. The information can be as simple as the date on which an image was created or very complex, such as a record of every time the file was opened and how it was processed. The idea is to facilitate working with content whilst keeping track of details such as source, history and associated copyright restrictions, for instance. XMP and how XMP data is processed was published as ISO 16684 in 2012, but a revision was submitted for publication as ISO 16684 in January 2019. ISO is also working on an additional part for ISO 16684 for managing metadata across files, for instance for random content access, archiving and for machine learning applications.

Preflight Processes

Knowing whether production files will make it unscathed through a print media production workflow requires robust quality control procedures. Many such procedures are implementations of the data formats and tools described here. Preflight software should be selected for its suitability for the workflow and be able to identify and fix PDF errors without drama. Supposedly print-ready PDFs can still have missing fonts or RGB images instead of CMYK. Preflight checking software is fundamental to process automation and quality control for preproduction data management, ensuring that errors are fixed before fixing them gets expensive. Knowing how digital systems and apps work helps improve workflow management, resource utilisation and ultimately profits.

Further Reading

https://www.gwg.org/
https://www.iso.org/
https://www.ibm.com
https://opentextbc.ca/graphicdesign/
https://dzone.com/

JDF can be used to manage output to the printer and also send data to ensure that cutting is also accurate. ©Roland

Published by FESPA Limited
Holmbury
The Dorking Business Park
Station Road
Dorking
RH4 1HJ

t +44 1737 240788
f +44 1737 233734
e info@fespa.com
www.fespa.com

FESPA
profit for purpose