



the national archives

**Digital Preservation
Guidance Note:**

4

Graphics File Formats

Document Control

Author: Adrian Brown, Digital Archives Analyst

Document Reference: DPGN-04

Issue: 1

Issue Date: 9 July 2003

Contents

1	INTRODUCTION	4
2	TYPES OF GRAPHICS FORMAT	4
2.1	Raster Graphics.....	4
2.1.1	<i>Colour Depth</i>	5
2.1.2	<i>Colour Spaces and Palettes</i>	5
2.1.3	<i>Transparency</i>	6
2.1.4	<i>Interlacing.....</i>	6
2.1.5	<i>Compression</i>	7
2.2	Vector Graphics.....	7
2.3	Metafiles	7
3	RASTER GRAPHICS	8
3.1	Graphics Interchange Format (GIF).....	8
3.2	JPEG File Formats (JFIF and SPIFF).....	8
3.3	JPEG 2000 (JP2).....	9
3.4	PCX.....	9
3.5	Portable Network Graphics (PNG).....	9
3.6	Sun Raster (RAS).....	10
3.7	Tagged Image File Format (TIFF)	10
3.8	Windows Bitmap (BMP).....	11
4	VECTOR GRAPHICS	11
4.1	AutoCAD Drawing Format (DWG)	11
4.2	AutoCAD Drawing Exchange Format (DXF).....	11
4.3	Microstation Drawing Format (DGN)	12
4.4	Scaleable Vector Graphics (SVG)	12
5	GRAPHICS METAFILES.....	13
5.1	Computer Graphics Metafile (CGM)	13
5.2	Windows Metafile (WMF).....	13
5.3	Windows Enhanced Metafile (EMF)	14
5.4	WordPerfect Graphics Metafile (WPG).....	14

1 Introduction

This document is one in a series of guidance notes produced by the Digital Preservation Department of The National Archives, giving general advice and guidance on issues related to the preservation and management of electronic records. It is intended to be used by anyone involved in the creation of electronic records which may need to be preserved over the long term.

This guidance note provides advice on general issues which should be considered by the creators and managers of electronic records when selecting file formats for use with images. The National Archives does not specify or endorse the use of any particular image file formats for records which are to be transferred. Choice of file format should always be determined by the functional requirements of the record-creating process. However, record creators should be aware that long-term sustainability may be a requirement, both for ongoing business processes and archival preservation.

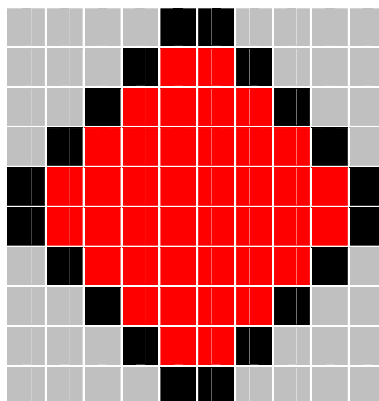
Advice on criteria for selecting file formats for long-term preservation is provided in Guidance Note 1 of this series. This document provides summary information about the most common graphics file formats in current use, with pointers to more detailed sources, in order to enable data creators, managers and archivists to make informed decisions. It is limited in scope to static images, and specifically does not include animation or video formats. It also does not include Page Description Language and presentation file formats such as PostScript and Portable Document Format, which may include image data.

2 Types of Graphics Format

Two distinct approaches exist for digitally encoding static images. These are known as raster and vector graphics. A further group of formats can store both raster and vector data within a single file, and are known as metafiles. A variety of file formats exist for encoding each type of image.

2.1 Raster Graphics

A raster image comprises a two-dimensional grid of pixels, each pixel having a specific colour value. A simple example is shown below:








The image data is usually stored as a series of scan-lines, each representing one row of the image grid. Each scan-line comprises sets of consecutive values representing the colour of each pixel in the row. These scan-lines may simply be stored contiguously within the file, or be aggregated into strips or tiles, which can speed up the decoding and decompression of the image.

A number of issues need to be considered when using raster images:

2.1.1 Colour Depth

Colour depth describes the number of distinct colours which can be displayed by a single pixel within a raster image, and is function of the number of bits allocated to each pixel. For example, pixels described by a single bit can only display 2 colours (represented by 0 or 1), whereas pixels using 24 bits can display 2^{24} (16,777,216) colours. The most common colour depths are as follows:

-  1-bit: Monochrome (2 colours)
-  4-bit: Greyscale or colour (16 colours)
-  8-bit: Greyscale or colour (256 colours)
-  16-bit: High Colour (65,536 colours)
-  24-bit: True Colour (16,777,216 colours)

True Colour is so-called because it corresponds approximately to the number of unique colours distinguishable by the human eye.

2.1.2 Colour Spaces and Palettes

Colours are defined by specifying a number of values; the scheme which defines the nature of these values is referred to as a *colour space*. The most common colour space is RGB, which defines colours in terms of their red, green and blue components. Typically, each component is stored as one byte, having a range of possible values between 0 and 255. Any colour can therefore be specified as a combination of red, green and blue values, for example:

Colour	Red	Green	Blue
Black	0	0	0
White	255	255	255
Blue	0	0	255
Yellow	255	255	0

Note that yellow is created by mixing red and green because RGB uses additive colours, i.e. those created by light rather than pigment. Subtractive (pigment) colours, as used for

example in printing, are usually defined using the CMYK colour space. This specifies colours in terms of their cyan, magenta and yellow components (the K represents black, which is defined separately for convenience). Another common colour space is HSV, which specifies colours through a combination of their hue, saturation and value. Like RGB, this uses an additive colour system.

The colours of individual pixels may be stored within a file in one of three ways, depending upon the colour depth:

- 📄 For 1-bit pixels, the colour is defined by either a 0 or 1, representing black or white.
- 📄 For colour depths up to 8-bits per pixel, indexed colours are generally used. Each pixel value refers to an entry in an index (otherwise known as a palette or look-up table), usually stored elsewhere in the file. The index relates these values to actual colour values for display. This allows the image data to be stored more efficiently, since the real colour values need only be stored once, rather than for each pixel. It also allows the colours used by an 8-bit file to be selected from a palette of 16,777,216 colours, even though only 256 separate colours can be used within the image.
- 📄 With colour depths above 8-bits, colours are generally stored as literal values for each pixel (for example, as RGB values). This is because any palette would be too large to represent a real saving in space.

2.1.3 Transparency



Raster images are normally opaque, but some formats allow regions of an image to be transparent. This is of use when overlaying images, or creating icons. This effect is usually achieved by allocating 8 bits per pixel to an *alpha channel*, which can be set to any level of transparency from 0 (opaque) to 256 (completely transparent). Transparency is only supported by certain image formats.

2.1.4 Interlacing

Raster images are typically stored as a series of consecutive scan lines. However, image data can also be interlaced, whereby the scan lines are stored out of sequence. In its simplest form, even rows might be stored first, followed by odd rows. The image would then be displayed in two passes, the first showing the even rows, and the second building up the complete image with the addition of the odd rows. Interlacing allows users to see an image before all the image data has been read, which can be helpful when the data is being downloaded from the Internet. It is only supported by certain file formats.

2.1.5 Compression

Various compression algorithms may be employed to reduce the physical size of a file. Graphics compression techniques work by compressing the image information within a file, and are entirely distinct from file compression methods, such as WinZip, which compress the entire file bitstream. Graphics compression algorithms fall into two categories:

-  Lossy compression (such as JPEG) achieves its effect at the cost of a loss in image quality, by removing some image information.
-  Lossless compression techniques (such as CCITT Group 4) reduce size whilst preserving all of the original image information, and therefore without degrading the quality of the image.

More detailed information and advice on graphics compression techniques may be found in Guidance Note 5.

2.2 Vector Graphics

Vector graphics formats store images as mathematical representations of image elements, such as shapes or lines. For example, a line segment might be defined in terms of the coordinates of its starting point, a direction, and a length. More complex shapes can be built up from simple shapes. Enclosed shapes can also be filled with colours. Some vector formats support 3-D objects as well, such as wire frame models. Vector formats are most commonly used in the field of Computer-Aided Design (CAD), since they are ideally suited to the creation of architectural and engineering drawings, maps, schematics, and charts. They also form the basis for 3-D modelling and animation, although this is beyond the scope of this Guidance Note.

Vector files can be easily manipulated, and rescaled without loss of quality. The size of a vector file is proportional to the complexity of the image (unlike raster images). Vector files do not usually support compression. However, vector file sizes are typically far smaller than the equivalent raster image.

2.3 Metafiles

A third category of file, known as a metafile, allows both raster and vector versions of an image to be stored within a single file. Metafiles are frequently developed to provide hardware- and software-independent methods of transferring image data. However, their use tends to be limited to very specific applications.

The following sections describe the most common file formats currently available in each category, including information about the developers, different versions of the format which may exist, the main features of the format, and sources of more detailed technical information.

3 Raster Graphics

3.1 Graphics Interchange Format (GIF)

The GIF format was developed in 1987 by CompuServe Incorporated, primarily for use on the Internet. The current version, GIF89a, was released in 1990. It supports colour depths from 1-bit (monochrome) to 8-bit (256 colour) and always stores images in compressed form, using lossless LZW compression. Other supported features include interlacing and transparency.

GIF is a proprietary format. In addition, the patent for the LZW algorithm is owned by Unisys Corporation, which has licensed its use by CompuServe. Despite the controversy which surrounded the application of license fees to GIF readers and writers, this format remains one of the most widespread in use on the Internet. The US patent expired in June 2003 and the UK patent will expire on 18th June 2004.

The extent to which it will be replaced by alternatives such as PNG (see 3.5) remains an open question.



The full technical specifications for GIF version 89a are available from CompuServe Incorporated.

3.2 JPEG File Formats (JFIF and SPIFF)


JPEG itself is not a file format, but rather an image compression algorithm developed by the Joint Photographic Experts Group in 1990. The original specification did not describe a file format for data exchange. However, the Independent JPEG Group and C-Cube Microsystems developed a JPEG File Interchange Format (JFIF) in 1992, which has become a *de facto* standard; this is what is commonly referred to as the JPEG file format. In 1996, Part 3 of the JPEG standard was released, containing the specification for an official file format, called SPIFF (Still Picture Interchange File Format). This is more complex than JFIF, and does support other compression schemes as well. It is designed to be interoperable with JFIF. The JPEG algorithm is also supported by a number of other raster image formats, including TIFF.

JFIF and SPIFF are 24-bit colour formats and use lossy compression, although a lossless extension to JPEG is also available (but not widely used). JPEG is particularly suited to the storage of complex colour images, such as photographs.

The arithmetic encoding extension of JPEG is subject to patent. A patent claim on the baseline JPEG Discrete Cosine Transform compression algorithm was made by Forgent Networks in 2002. This is disputed by the JPEG committee and has yet to be resolved, although the patent in question will expire in 2004.




The full technical specifications for the current version of the JFIF specification (version 1.02) are freely available from the official JPEG website (www.jpeg.org) and from C-Cube Microsystems.

 The full technical specifications for the SPIFF specification are freely available from the official JPEG website, as an extract from ISO/IEC 10918 Part 3.

3.3 JPEG 2000 (JP2)

JPEG 2000 is a replacement for the JPEG algorithm, developed by the ISO JPEG group in 2000. The JPEG 2000 standard defines a minimum file interchange format (JP2), in a similar manner to JFIF and SPIFF. JPEG 2000 provides for lossy and lossless compression, and uses wavelet compression to achieve higher compression rates with a lower corresponding reduction in image quality. It supports colour depths up to 24-bit (true colour) and is best suited to complex colour images, such as photographs.


JPEG 2000 may utilise some patented technologies, but is intended to be made available on a license- and royalty-free basis. Support for the format is now beginning to appear in a number of commercial software packages.

 The full technical specifications for the core JPEG 2000 format have been published as an international standard (ISO/IEC 15444 Part 1).

3.4 PCX

The PCX format was developed by ZSoft for its MS-DOS PC Paintbrush graphics application. In 1990, Microsoft licensed PC Paintbrush to integrate into Windows 3.0 (as Microsoft Paintbrush), and this resulted in PCX becoming one of the most widely-used raster formats for the Windows platform. However, its popularity has waned somewhat in recent years, following the advent of more advanced formats. The most recent version of the format is Version 5, released in 1991.


PCX supports 1-bit to 24-bit colour depths, and all image data is compressed using RLE. The format is still supported by a wide variety of applications.

 The full technical specification for PCX version 5 was published in the ZSoft *Technical Reference Manual, Revision 5*, in 1991. Zsoft have ceased trading and the specification no longer appears to be officially supported, although copies are widely available on the Internet.

3.5 Portable Network Graphics (PNG)

The Portable Network Graphics (PNG) format was developed by the PNG Development Group in 1996, to provide an open alternative to GIF and the associated licensing issues with LZW compression (see 3.1). It supports colour depths from 1-bit to 48-bit. Image data is always stored compressed, using a variation on the lossless LZ77-based Deflate compression algorithm which is not patented, and therefore free to use. Support for interlacing and transparency is also provided, and the format incorporates a number of error detection mechanisms.


The PNG specification is entirely public domain and free to use. This, combined with its rich feature set and burgeoning popularity, make it one of the most promising raster image formats currently available.

 The full technical specifications for PNG 1.0 have been published as RFC-2083, and as a W3C recommendation. Version 1.2 is intended to be released as an ISO standard (ISO/IEC International Standard 15948).

3.6 Sun Raster (RAS)

The Sun Raster format is the native raster format of the SunOS Unix operating system, and the most widespread Unix-specific raster image format. It was developed by Sun Microsystems Incorporated in 1990. It supports colour depths up to 32-bit (True Colour) and provides optional lossless RLE compression.

The format is supported by most UNIX-based imaging software.

 Information about the technical specifications for RAS may be obtained from Sun Microsystems Incorporated.

3.7 Tagged Image File Format (TIFF)

The TIFF format was originally developed by the Aldus Corporation, primarily for use in scanning and desk-top publishing. Aldus first published the specification in 1986. When Adobe Systems Incorporated purchased Aldus in 1994, they acquired the rights to the TIFF specification, and have maintained it since then. The current version of the specification (revision 6.0) was released in 1992.

TIFF supports colour depths from 1 bit to 24 bit (e.g. monochrome to true colour), and a wide range of compression types (RLE, LZW, CCITT Group 3 and Group 4, and JPEG), as well as uncompressed data. It should be noted that the implementation of JPEG compression described in Revision 6.0 is flawed: a revised design was issued as TIFF Technical Note #2, which should be followed instead.

TIFF also incorporates perhaps the most comprehensive metadata support of any raster format, allowing the addition of a wide variety of technical and resource discovery information to be included. TIFF is designed to be an extensible format, and new tags can be registered with Adobe. A number of official and unofficial extensions exist, but should be treated with care since support may not be widespread.

Adobe owns the TIFF specification, but makes it freely-available for use. The same licensing issues apply to the use of LZW compression in TIFF files as GIF (see 3.1). An unofficial TIFF extension has been developed using the same Deflate algorithm as the PNG format (see 3.5). However, this is not generally supported.

TIFF is one of the most flexible, popular and widely-supported raster formats in use today.

 The full technical specification for TIFF revision 6.0 is freely available from the Adobe website.

3.8 Windows Bitmap (BMP)

The BMP format (sometimes referred to as a Device Independent Bitmap) was developed by Microsoft as the native raster format of the Windows operating system. Versions of the format have therefore coincided with releases of Windows, the first version appearing in 1985 with Windows 1.0. Additional versions were developed for use with IBM's OS/2 operating system. The BMP format supports colour depths from 1-bit to 32-bit and provides optional lossless RLE compression (Windows 3.0 versions and later only).

Although proprietary, this format is free to use and is widely supported, due to its association with Windows. However, its popularity has declined in recent years with the advent of more advanced formats.



A formal technical specification for BMP has not been released by Microsoft, although information may be available through the Microsoft Developers Network Knowledge Base.

4 Vector Graphics

4.1 AutoCAD Drawing Format (DWG)

Autodesk's AutoCAD software is probably the most widespread Computer-Aided Design (CAD) package in current use. As a result, AutoCAD's native drawing format, DWG, has become a *de facto* standard for vector graphics. The DWG specification is revised with each release of AutoCAD: the original version being released with AutoCAD 1.0 in 1982, and the latest version being AutoCAD 2004.

DWG supports 24-bit colour depths (True Colour) and 3-D models.

The DWG format is proprietary to Autodesk and has not been released. Instead, Autodesk recommend the use of the DXF format for data exchange (see 4.2). Nonetheless, DWG has become a *de facto* standard for the exchange of CAD data, and is supported by most CAD packages, albeit with varying degrees of fidelity – according to current estimates there are in excess of 4 billion DWG files in existence worldwide.



The OpenDWG Alliance, formed by a group of interested vendors, is committed to making the DWG standard open. Reverse-engineered specifications are available from their website, but it must be stressed that these are not official.

4.2 AutoCAD Drawing Exchange Format (DXF)

The Drawing Exchange Format (DXF) was developed by Autodesk Incorporated to provide a method of data exchange with other CAD applications. The DXF specification is revised with each release of AutoCAD: the original version being released with AutoCAD 1.0 in 1982, and the latest version being AutoCAD 2004.

DXF files can be encoded either in 7-bit ASCII or binary. They support 8-bit colour depths (256 colours) and 3-D shapes, and are uncompressed. DXF is a complex format, and the quality and sophistication of its implementation in different applications varies considerably. The frequent changes to the specification can also cause compatibility problems. In particular, users must be aware that some applications may read a DXF file whilst skipping unsupported features. This can lead to the loss of information in a manner which may not be obvious to the user.

The DXF specification is owned by Autodesk, but is freely available to use. It is widely used for the exchange of vector data between different applications and platforms, but is best suited to relatively simple line data.



The full technical specifications for the most recent versions of the DXF format (currently Releases 13 through 2004) are freely available from the Autodesk website.

4.3 Microstation Drawing Format (DGN)

DGN is a native file format shared by Bentley Systems Incorporated's Microstation software and Intergraph Corporation's Interactive Graphics Design System software products. As such it is in widespread use for CAD applications. The Microstation software and the DGN specification were originally developed in 1984 by Intergraph, before being purchased by Bentley Systems.

The introduction of Microstation V8 in 2001 instituted the first substantial revision of the DGN format, and represented a move towards closer compatibility with the DWG format.

The DGN format is proprietary to Bentley Systems Incorporated, and its use tends to be limited to Microstation-based environments.



Information about the technical specifications for DGN may be obtained from Bentley Systems Incorporated.

4.4 Scalable Vector Graphics (SVG)

Scalable Vector Graphics (SVG) is a format for describing two-dimensional graphics using XML. It was developed by the SVG Working Group of the W3C in 2001. The current version is 1.1, released in 2003. SVG is primarily a vector format, although it does allow for the inclusion of raster images (JPEG and PNG). In this regard, it is closely associated with metafiles.

SVG supports 24-bit colour, and allows the creation of sophisticated dynamic and interactive graphics. Being entirely XML-based, it enjoys the advantages of extensibility, interoperability and flexibility. SVG can be easily manipulated and transformed using standard XML tools, or generated on-the-fly.

SVG is an open, non-proprietary format. It is rapidly becoming a major standard for vector imagery, particularly on the Internet. It is widely supported by all the major vendors, and a number of free viewers are available for on- and off-line viewing.



The full technical specification for SVG 1.1 has been published as a W3C recommendation. Version 1.2 is scheduled for release in 2004.

5 Graphics Metafiles

5.1 Computer Graphics Metafile (CGM)

The Computer Graphics Metafile format was developed by the International Standards Organisation (ISO) and the American National Standards Institute (ANSI) as an open, platform-independent format for the exchange of raster and vector data for technical applications. The original version of the specification was released in 1986. The current version (version 4) was released in 1999. CGM supports all colour depths, and provides for RLE, CCITT Group 3 and CCITT Group 4 compression. It has a very rich feature set and provides good metadata support.

The CGM standard allows three encoding methods: character-based, binary, and plain ASCII. Not all are supported by every CGM application. In addition, many vendors have implemented different CGM features, or added their own extensions. As a result, CGM has not always achieved true application or platform-independence. However, it is widely used for the exchange of technical design data within many industries.

A recent development is WebCGM, a CGM profile for use within web-browsers. This is being developed as an open standard by the CGM Open consortium. Although there are certain similarities with the SVG format (see 4.4), WebCGM is much more focused on technical graphics. Both are being taken forward as complementary standards by the W3C.



The full technical specification for CGM version 4 is available as an ISO standard (ISO/IEC 8632:1999).



The full technical specification for WebCGM 1.0 revision 2 has been published as a W3C recommendation.

5.2 Windows Metafile (WMF)

The Windows Metafile format was developed by Microsoft for use with its Windows operating system. WMF supports colour depths up to 24-bit, and does not offer compression. Embedded raster images are stored in Windows Bitmap (BMP) format (see 3.9).

WMF has been partially superseded by EMF, but is still supported by many Windows applications and, although proprietary, is free to use.



Technical information about the WMF specification is available from Microsoft.

5.3 Windows Enhanced Metafile (EMF)

The Enhanced Metafile format was introduced by Microsoft with the advent of Windows NT in 1993. It was designed as a more advanced successor to WMF, allowing device independence and improved embedding features. The format supports colour depths up to 24-bit, and does not offer compression. Embedded raster images are stored in Windows Bitmap (BMP) format (see 3.9).

EMF is supported by many Windows applications and, although proprietary, is free to use.



Technical information about the EMF specification is available from Microsoft.

5.4 WordPerfect Graphics Metafile (WPG)

The WordPerfect Graphics Metafile format was developed by the WordPerfect Corporation, primarily for use with its WordPerfect word-processing software. A WPG file can store raster and vector images, using colour depths up to 8-bit indexed colour. Versions created using WordPerfect 5.1 or later can store raster and vector data in the same file, but earlier versions can only store one or the other. Raster images are stored using RLE compression. Encapsulated PostScript data may also be embedded within a WPG file – any raster images contained within the EPS data are converted to the WPG bitmap format.

The WPG format is proprietary and now owned by the Corel Corporation. Its use tends to be limited to environments using WordPerfect products – it is not widely used as an interchange format.



Information about the technical specifications for WPG may be obtained from Corel Corporation.