



Vector images: A guide to good practice



22nd February 2023





The research leading to these results has received funding from the European Community's Horizon 2020 Programme (H2020-INFRAIA-2018-1) under grant agreement n° 823914.

For any questions regarding this Guide, please email: contact@ariadne-infrastructure.eu

PDF/A-1 (ISO 19005-1), created in MS Word (Version 2016) from the docx format.

Contents

1. Introduction to vector images	5
1.1 What are vector images?	5
Areas of application	6
1.2 Things to consider	7
2. Things to consider when creating vector images	8
2.1 General considerations	8
3. Archiving vector images	9
3.1 Deciding what should be preserved	9
3.2 Deciding how it should be preserved	9
Significant properties	9
Recommended file formats	10
3.3 Metadata and documentation	11
4. File format	12
5. Bibliography	20

Vector images: a guide to good practice

Acknowledgements: Based on Archaeology Data Service's Guides to Good Practice on Vector Images (<https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/>). The information has been reworked, updated and supplemented by the Swedish National Data Service (<https://snd.gu.se/en>) to be more appropriate to data of non-archaeological origin.

1. Introduction to vector images

Images appear in many contexts and may be used for different purposes. An image is usually a two-dimensional representation of something and may form the empirical data material in a context or be used as an educational and visual aid for example. There are two main image file formats used to describe digital images. These are usually referred to as vector images or raster images (these are sometimes also referred to as bitmaps). Both image file formats have their advantages and disadvantages. For example, the vector format handles enlargement better than the raster format (see Figure 1) but is not able to represent colour nuances as accurately (Figure 3). Vector images are commonly used in GIS and CAD, for example, while digital photographs are the most common type of raster image. This guide aims to provide guidance on what to consider when preserving two-dimensional vector images. For guidance on handling raster images, see our guide in the same series: *Raster images: a guide to good practice*.

The guide covers the most common file formats used for storing vector images and which formats are suitable for archiving. It will also discuss the archival strategies that may be used to ensure that the quality of the files is maintained. Although some basic technical descriptions will be given in the guide, it is not intended to provide a more in-depth account of the technical specifications of vector images. Three-dimensional vector objects are often part of complex 3D datasets, are stored in different formats from 2D data and will not be covered in this guide. Although vector graphics are common in GIS and CAD, for example, this guide does not focus specifically on the management of GIS or CAD datasets.

Vektor Raster

Figure 1: The difference between enlarged vector and raster images. The vector image remains sharp after enlargement, while the raster image is perceived as pixelated and blurred.

1.1 What are vector images?

Actually, it is misleading to talk about ‘vector images’ because it gives the impression that all vector images are four-sided images in the same style used to describe rasters. A vector image may consist of a rectangular image for which all elements are described with vector graphics, but it is equally likely that the image consists of one or more vector figures that are all stored in separate files. How vector graphics are used varies between applications. A general feature, however, is that vector images consist of defined, geometrical figures and are mainly used for simple illustrations such as logos, fonts, or certain types of maps. Unlike rasters, which consist of grids of pixels in which each pixel has its specific colour and location, vector graphics are not made up of pixels (see Figure 2). A digital photograph often contains too much colour information for a vector format to be appropriate (see Figure 3). Not because it is impossible to create the same detailed complexity using vector graphics, but because the file would be unnecessarily large, and the procedure is more time-consuming. Vector figures may consist of lines, polygons, and curves. To describe the shape of the figure, connected points are used, which is similar to joining the dots and seeing a picture emerge. The location of the points is determined by coordinates and mathematical formulas, allowing vector objects to be scaled without loss of quality. Therefore, vector graphics can be enlarged without the risk of the enlarged object being blurred or perceived as pixelated. Vector figures can also be manipulated in various ways (for example rotated and combined with other vector objects) and can

contain two- or three-dimensional geometry. In some cases, files may contain both vector and raster data.

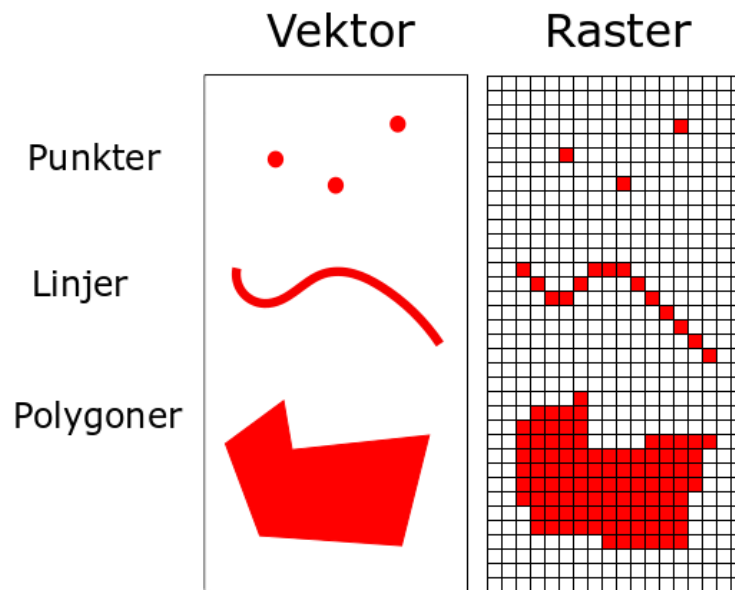


Figure 2: Representation of points, lines and polygons as vectors or rasters.

Areas of application

Vector images are common in research fields where maps and other spatial information are important to understand the study objects (for example in archaeology and climate research). Vector graphics may be created in many different contexts but are often used for two-dimensional images produced for publication purposes in articles or project reports. It is also more or less standard for logos to be made using vector graphics, as a single vector file can be used for everything from headers on paper documents and web pages, to advertising posters covering entire house walls (without losing resolution or requiring a larger file). In 3D modelling, almost only vectorized representations are used, except for surface texture which may consist of raster (for example from digital photographs).¹



Figure 3: A vectorized version (left) of a digital photo (right). Much of the detail and shades of colour is lost in the vectorized version.

¹ https://en.wikipedia.org/wiki/Vector_graphics. Accessed 21 February 2023.

1.2 Things to consider

The shape of vector figures is determined by the position of pairs of coordinates (X and Y) for different properties. In GIS applications, coordinate properties can be linked to real coordinates, and objects can then be related to a physical location. The vector format is well suited for map data that contain clear boundaries such as land borders, fields, rivers, and streets. In applications that do not aim to describe spatial information, customized, arbitrary coordinates are used to keep the shape and position of the vector figure the same on another computer, or in another application. The properties may be in the form of points (POINT or NODE), linked to lines (ARC or LINE) or connected to form defined surfaces (AREA or POLYGON). Attribute data linked to the individual properties/figures are maintained in a separate database.

2. Things to consider when creating vector images

2.1 General considerations

There are two main ways in which vector images can be created:

1. Generation by transformation of imported data values, for example GPS-measured measurement points with coordinates.
2. By manually placing curves and fields, thereby defining the shape of the figure, or by converting/extracting properties from a raster image (Coyne et al. 2007).

In the first case, these may be vector objects generated from measured values taken from, for example, a .txt or .csv file. In such a case, it is advisable to keep the original file of measured values with the vector file. The second case represents all situations in which the data creator has 'drawn' the figure themselves. An example of this is when you want to create vectorized maps from scanned historical maps. In the case of vectorization of a raster image, the original raster image needs to be stored with the vector file.

Vector files are sometimes accompanied by an associated database of important metadata. The metadata included may vary, but the database may, for example, contain information about the ID of different vector figures, what each figure describes (for spatial data, this may be designations such as 'industrial area,' 'marsh', or 'Korpen district'), coordinates and the like. If such a database exists, it is particularly important for it to be stored with the associated vector file as it contains the necessary information for future reuse.

In addition to the image itself, and any associated database, many vector image applications allow scripts (for example JavaScript, AppleScript, or VBScript in Illustrator) to be included in the file. Scripts in such files are intended to automate or simplify the creation or editing step but are not core components of the vector file. If such scripts are worth preserving, they should be stored and documented separately with the image.

3. Archiving vector images

3.1 Deciding what should be preserved

In 1. *Introduction to vector images*, it was mentioned that many vector images are created or incorporated in large datasets in CAD or GIS applications. These datasets often originate from a range of data collection techniques such as prospecting, total station/GPS surveying or laser scanning. If a vector figure originates from a file containing a number of different vector figures, for example a GIS file with associated attribute database, it is recommended to keep the original file intact instead of exporting all figures to separate files. This is because the original file contains all the necessary information and because it is convenient to have a single file instead of many. It is also possible to export individual figures later if required. However, there are times when it is better to store vector figures as separate files, but this varies from project to project. Regardless of the software, collection technique or purpose, a data creator should ensure that each file is accompanied by adequate documentation to be reusable.

Vector images are often created as part of large projects, usually for inclusion in a publication. In some cases, vector images may even be converted into raster images such as JPEG or TIFF before being included in a document. The data creator should evaluate the importance of the vector image in the workflow to determine whether the file in question has reuse value as a vector format. Where the image exists both in its original form (for example a GIS file containing many different vector figures) and as a raster image, it may be decided that the individual vector file is not retained and archived. The approach to be taken varies and must be decided by the data creator in each case.

3.2 Deciding how it should be preserved

Vector images should be organized logically and in a way that suits the specific project. Where metadata exist, it is suggested that they are stored in a simple text file (.txt or .csv) or XML file with the preserved image files. As for other file types, when archiving vector images, it is essential that the significant properties of the vector image are identified and maintained in the chosen format. The *Significant Properties of Vector Images* (Coyne et al. 2007) describes the properties that should be maintained when archiving vector graphics, and these will only be briefly described here.

Significant properties

The significant properties of a vector image revolve around the geometry of the image, the relationships between different figures and the appearance properties (colours, line types and thickness). These need to be kept intact when migrating between file formats and for archiving. The most significant properties of vector images are described below:

- Points, lines with text and/or embedded metadata
- Basic geometries/primitives (for example rectangles, circles, ellipses, etc.)
- Image structure (i.e., grouped objects and layers)
- Clipping and masks
- Transformations and coordinate systems (may be available in GIS and CAD vector files)

In addition to ensuring that these specific vector elements remain intact, there are a number of other properties that may also be added. Where standard conventions exist and have been followed (such as specific meanings associated with certain line colours or dash types), it is important that these properties and formatting are maintained in archived files. Such properties include:

- Line width, terminations of lines (these may sometimes be rounded) and join types

- Mitre limit (where two lines meet at an acute angle, the mitre limit is used to cut off the apex of the angle)²
- Patterns and offsets
- Colour
- Transparency
- Rendering³ (the process of representing complex digital vector images based on simple primitives)
- How the area that makes up the interior of a polygon figure is defined
- Gradient (transition from one shade to another)
- Text attributes such as font size and font.

Unfortunately, there is no easier and more objective way of comparing original files and migrated versions than to visually assess that all elements are rendered correctly. When converting a file to a new format, it is also important to check for any hidden information, such as transparent objects or layers, and assess their relevance for archiving. Linked data, such as embedded raster files or external fonts, may also be present in the original image and such data should remain intact when migrating between file formats. How successfully a vector file is preserved depends largely on the nature of the original, its complexity and whether it contains only vector data or whether raster data are also included. Such mixed files may require that different data types be separated and saved in different preservation formats.

Recommended file formats

The file formats described below are those recommended for the preservation of image files:

Format	Description
Scalable Vector Graphics, version 1.1 (.svg)	As an open XML-based standard, SVG is the recommended format for the archiving and preservation of vector graphics.
WebCGM 2.1 (.cgm)	WebCGM is recommended for preserving vector images where SVG is not appropriate or where images have been created in CGM format.
AutoCAD Drawing Interchange Format (.dxf)	DXF is the recommended format for preserving and disseminating CAD data, and the ASCII version of the format should be the first choice.
Encapsulated Postscript (.eps)	EPS is suitable for dissemination and storage of vector data. Supports both vector and raster data.
PDF/A (.pdf/a)	Although PDF/A is suitable for preserving vector graphics, the format is only recommended as a last resort. Although the format retains the appearance of the vector images, the possibilities for reuse are reduced because PDF/A can only be used as a distribution/viewing format.

² <https://stevencrawley.wordpress.com/2015/09/17/what-is-the-miter-limit/>. Accessed 29 August 2022.

³ <https://sv.wikipedia.org/wiki/Rendering>. Accessed 29 August 2022.

3.3 Metadata and documentation

As with other types of data, vector image metadata provide vital information about how the image(s) was(were) created. Apart from basic project-level metadata, metadata associated with the vector image should be stored so that their structure can be understood.

Properties	Description
File name	A unique identifier for the image, usually the file name.
Title	The title of the image or an appropriate caption.
Description	Description of the image.
Originator	Name of the data creator.
Date	Date on which the image(s) was(were) created.
Copyright	Who owns the files and how may they be used?
Keywords	Appropriate keywords to describe the file.
Provenance/relationship to other documents	Description of the context in which the image was created and how (if relevant) it is related to other files.
Conventions	If the file contains standard conventions (colours, layers, line styles, etc. with definite meaning), there needs to be an explanation of what these conventions are and what they describe.
File format and version	For example, SVG 1.1.
File size	Size of the file in bytes.
Software	Programme used to create the file, for example Adobe Illustrator.

4. File format

The tables below describe some common file formats for storing vector images and the formats that are recommended for archiving. Some of these formats may appear obsolete, but they are described here because it may be necessary to handle older file formats:

Adobe Illustrator Artwork	
File format/extension	AI/.ai
Format	Adobe Illustrator Artwork is a proprietary ⁴ file format owned by Adobe. The format is mainly used for two-dimensional, layer-based vector files.
Description	The AI format has changed considerably since it was created in 1987. Early versions were based on PostScript and 3D functions were only introduced later with Illustrator CS. AI files are created in the vector graphics editor Adobe Illustrator. The AI format stores data as either EPS or PDF and supports raster in addition to vector. AI is now supported, to varying degrees, by several applications not owned by Adobe. However, it should be noted that some applications that can open AI files convert them into raster, with the result that the images will lose some functionality and become blurred when enlarged.
Recommendations	Not suitable for dissemination or archiving.

Adobe PDF	
File format/extension	PDF/.pdf, .pdf/a, .pdf/e, .pdf/x
Format	PDF was developed in 1993 by Adobe and is an open, widely used format based on PostScript.
Description	The PDF format is designed primarily for cross-platform document sharing and provides documents that look the same both on screen and in printed form. PDF content may in some cases be ASCII text, but it is more common for files to be compressed and therefore binary. PDF is primarily a distribution format. At the same time, .pdf incorporates a number of variants of PDF that

⁴ **Proprietary file formats** are file formats that have restrictions (usually set by the owner) on how they can be used, modified, or copied.

	are intended for specific purposes. PDF/A is an ISO standardized (ISO 19005 - 1:2005) ⁵ archiving format.
Recommendations	PDF/A has been accepted as a viable format for archiving (for example by the Library of Congress ⁶) and has been reviewed and assessed by DigitalPreservationCoalition (DPC). ⁷ PDF/A files are deemed suitable for preservation, but owing to its limitations, the format is not suitable for all types of content. The format may be suitable for dissemination, but much of the functionality of the vector images will be lost.

AutoCAD	
File format/extension	DXF/.dxf
Format	DXF (Drawing Exchange Format) is a proprietary, but openly documented format developed by AutoDesk. The format is designed for CAD data.
Description	The DXF format is widely used for its compatibility for data transferred between different CAD applications, which is also the purpose of the format. The programme is well established and supported by many applications. DXF is the preferred format for preserving and disseminating CAD data, primarily the ASCII version of the format. AutoCAD is gaining more complex features and the changes to the specification may cause some compatibility problems. A data creator must therefore be aware that some applications that support DXF files may not support all special functions. It is currently uncertain whether DXF will continue to be the preferred format for CAD data in the future, or whether it will be replaced by other formats. ⁸
Recommendations	Suitable for dissemination and archiving of CAD files.
File format/extension	DWG/.dwg
Format	DWG is a proprietary CAD data format that was developed by AutoDesk back in the 1980s.
Description	DWG is a file format for CAD models and is the standard file format for AutoCAD, Intellicad and PowerCAD. DWG has thus become widely used, but

⁵ <https://www.iso.org/standard/38920.html>, last visited 29 August 2022.

⁶ <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>. Accessed 29 August 2022.

⁷ 'Preserving the Data Explosion: Using PDF' (Fanning 2008), written by DigitalPreservationCoalition (DPC). Accessed 5 September 2022.

⁸ <https://www.loc.gov/preservation/digital/formats/fdd/fdd000446.shtml>. Accessed 29 August 2022.

	AutoDesk has never published a specification for the format. The DWG format is an alternative to the DXF format for storing CAD data where file size or full compatibility may be an issue.
Recommendations	Suitable for dissemination, but not for archiving. For CAD files, the DXF format is preferred for dissemination and archiving.

CDR	
File format/extension	CDR/.cdr
Format	CDR is created in CorelDRAW and is a proprietary, binary, layer-based 2D format.
Description	CorelDRAW is a vector-based illustration application and a competitor of Adobe Illustrator. As of version X4 (14), the CDR file consists of a ZIP-compressed directory containing several files, including XML files. There is no published specification for CDR. Like AI, CDR can handle both vector and raster data.
Recommendations	Suitable for dissemination, but not for archiving.

Computer Graphics Metafile	
File format/extension	CGM/.cgm
Format	CGM is an open, freely available file format for two-dimensional raster/vector files. CGM is an ISO standard format (ISO/IEC 8632-1:1999). ⁹
Description	The CGM standard permits two encoding methods: character-based and binary. Not all CGM applications support both encoding methods. In addition, many suppliers have implemented various CGM functions or added their own extensions. As a result, CGM is not entirely application-independent or platform-independent. However, it has been widely used for the exchange of technical design data in some sectors. ¹⁰ There is no web support for the format.

⁹ <https://www.iso.org/standard/32378.html>. Accessed 5 September 2022.

¹⁰ <https://www.nationalarchives.gov.uk/documents/graphic-file-formats.pdf>. Accessed 5 September 2022.

Recommendations	Not suitable for dissemination or archiving.
File format/extension	WebCGM/.cgm
Format	WebCGM is an open, freely available format based on CGM.
Description	A subset of the CGM format, WebCGM has been developed in recent years by W3C for use on the internet. WebCGM is essentially the same format as CGM but has more functions and web support. In addition, WebCGM uses only binary encoding to avoid the problem that some types of encoding are not supported by all applications.
Recommendations	WebCGM is suitable for dissemination and archiving where SVG cannot be used or where the files were originally created in WebCGM.

Encapsulated PostScript

File format/extension	EPS/.eps, .epsf, .epsi
Format	EPS or Encapsulated Postscript is a proprietary file format developed primarily for vector data, but it may also be used for raster images.
Description	EPS can contain vector data, raster data, and fonts. One problem that can occur with EPS files is that the header often contains a version number, but if the version number is missing, some applications may reject the file. EPS is essentially a defined, standalone Postscript file that can be embedded in other documents. The EPS file consists of two parts: an (optional) low-resolution preview image and a Postscript part that can contain both vector and raster graphic commands. ¹¹ There will be no further development of the format. There are security risks with the format, which is why Microsoft and others have made it impossible to import the format into some of their products. ¹² It is recommended that the format is converted into SVG before archiving.
Recommendations	Can be used for dissemination, but not for archiving.

¹¹ https://en.wikipedia.org/wiki/Encapsulated_PostScript. Accessed 29 August 2022.

¹² <https://www.loc.gov/preservation/digital/formats/fdd/fdd000246.shtml>. Accessed 5 September 2022.

Macintosh PICT

File format/extension	PICT/.pict, .pic, .pct
Format	PICT, or Picture File, is a proprietary metafile format developed in 1984 by Apple.
Description	PICT can store both raster and vector images. The format was developed to be used between different Apple applications but has been replaced by PDF and Apple has not supported the format since 2009. A PICT file contains all the QuickDraw commands used to draw the image. Many Apple applications still support the format, and it is supported by some Windows applications.
Recommendations	Not suitable for dissemination or archiving.

Macromedia Flash

File format/extension	SWF/.swf, .fla, .swd, .flv, .swc, .swt, .flp
Format	SWF is a 2D vector-based animation format developed by Adobe and optimized for web browsers.
Description	SWF stands for Small Web Format and is an Adobe Flash format. To open the files, Flash usually needs to be installed on your computer. The format specifications for SWF are freely available, but some elements are still deemed to be closed. Flash files can be used for both vector data and various types of multimedia. There has been a lot of controversy surrounding Adobe Flash and the application is deemed to be unstable and a security risk. In recent years, both the SWF format and the Flash application have been used less and less.
Recommendations	Not suitable for dissemination or archiving.

Macromedia Freehand

File format/extension	AF/.af
Format	Macromedia Freehand is an illustration package that uses the file type AF, which is a type of raster vector hybrid.
Description	The file format has changed significantly between versions, which causes some problems when migrating early versions. As of 2007, both development and support of Freehand ceased, although Adobe Illustrator still supports AF files.
Recommendations	Not suitable for dissemination or archiving.

Micrografx Designer

File format/extension	DRW/.drw, .dsf
Format	DRW, or Drawn file, is a vector-based format suitable for technical illustrations, such as drawings, user manuals or other technical documentation. Owned by Corel.
Description	DRW is a generic format used by many different applications. Image data are stored in a vector format that uses lines (or paths), instead of points, to represent the shape of the object. The format handles colours poorly compared to many other vector file formats.
Recommendations	Not suitable for dissemination or archiving.

Microsoft Windows Metafile

File format/extension	EMF/.emf, .wmf
Format	EMF is a proprietary vector-based file format developed by Microsoft.
Description	EMF, or Enhanced MetaFile, is an improved version of the older WMF format ¹³ and supports 32-bit colour space instead of only 16-bit in its predecessor. EMF uses Graphics Device Interface (GDI) commands ¹⁴ to render an image and the format is supported by many image editing applications. EMF stores image data in an RGB format and does not support CMYK.
Recommendations	Not suitable for dissemination or archiving.

PostScript

File format/extension	PS/.ps
Format	PostScript is a programming language developed by Adobe to describe the appearance of images and text in a device-independent way.
Description	PostScript is used to describe how pages should look, either on a computer screen or when printed. A PS file can contain vector graphics, raster graphics and text. PS files can be printed directly by a PostScript printer without being opened in an application. ¹⁵ The format has a number of derivatives, notably EPS and PDF, which in recent years have increasingly replaced PostScript in terms of popularity.
Recommendations	Although the format has potential as a suitable preservation format, PostScript is no longer widely used and newer alternatives such as PDF/A present a more robust alternative.

¹³ https://en.wikipedia.org/wiki/Windows_Metafile. Accessed September 2022.

¹⁴ https://en.wikipedia.org/wiki/Graphics_Device_Interface. Accessed September 2022.

¹⁵ <https://fileinfo.com/extension/ps>. Accessed 7 September 2022.

Scalable Vector Graphics

File format/extension	SVG/.svg
Format	SVG ¹⁶ was developed by W3C and is an open, XML-based standard format used to describe 2D vector graphics.
Description	SVG image properties are defined in XML text files. Therefore, SVG files can be created and edited with any text editor, although it is more practical to create them with a drawing application that supports the format. Since the format is XML-based, it also supports lossless compression. SVG supports 24-bit colour space and supports vector, raster, text, and animations.
Recommendations	Suitable for dissemination and archiving

WordPerfect Graphics Metafile

File format/extension	WPG/.wpg
Format	WPG was developed by WordPerfect Corporation and is a 2D graphics metafile that can store raster and vector graphics or EPS data.
Description	The format is proprietary and is currently owned by Corel Corporation. A WPG file can store raster and vector graphics and handle up to 8-bit colour space. Versions created with WordPerfect 5.1 or later can store raster and vector data in the same file, but earlier versions cannot store both at the same time, only either raster or vector image data.
Recommendations	Not suitable for dissemination or archiving.

¹⁶ <https://www.w3.org/TR/SVG/>. Accessed 7 September 2022.

5. Bibliography

Coyne, M., Duce, D., Hoggood, B., Mallen, G., Stapleton, M. (2007). *The Significant Properties of Vector Images*. JISC.

http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf, last visited 7 September 2022