[CV-02-020] Raster Formats and Sources

Abstract

Raster data is commonly used by cartographers in concert with vector data. Choice of raster file format is important when using raster data or producing raster output from vector data. Raster formats are designed for specific purposes and have limitations in color representation and data loss. The simplest raster formats are just a single two-dimensional array of pixels, where multi-band raster datasets use additional data values to represent color or other data. The article covers considerations for the intended use of raster formats. Formats and resolutions appropriate for the web may not be appropriate for print or higher resolution devices. Several types of raster sources are available including single band measures, imagery, and existing raster maps or basemaps. The future of raster will evolve as more formats, sources, and computational improvements are made.

Keywords: cartography, Data Considerations, imagery, pixel, raster

Author & citation

Williams, C. (2019). Raster Formats and Sources The Geographic Information Science & Technology Body of Knowledge (4th Quarter 2019 Edition), John P. Wilson (Ed.).
DOI: 10.22224/gistbok/2019.4.11.

Explanation

1. Definitions
2. Introduction
3. Raster Formats
4. Raster Sources
5. Raster Map Production
6. Future Directions

1. Definitions

- **alpha channel**: a raster band used for transparency compositing that allows for blending of pixel content with digital content conceptually beneath the raster dataset
- **anti-aliasing:** an approach to minimizing visual artifacts caused by fitting vector graphics into a grid of pixels
- **bands**: a defined channel or subset of a raster dataset, for instance, the red channel of an RGB image
- **bit**: a portmanteau of binary digit, the smallest unit of storable information
- **bit depth**: the number of bits storable for a given data type
- **color model**: a mathematical model for representing colors such as RGB or CMYK
- **color space**: a definition of colors in a color model which allows for repeatable reproduction of color
- **dithering**: a technique for minimizing content loss during reduction of color depth by inserting pixel noise
- **gamut**: a subset of colors in a color model representable in a color space

- **LIDAR**: Light Detection and Ranging, a technology for collecting data using laser reflectance
- **lossless**: compression that does not result in the loss of data
- **lossy**: compression that results in data loss
- **pixel**: a portmanteau of 'picture element', the smallest unit of a raster
- **raster**: datasets defined as an array of pixels
- **resolution**: the number of distinct pixels that can be displayed in a measured space
- **vector**: datasets defined as geometric primitives rather than raster data

2. Introduction

Cartographers use **raster** and **vector** data sources to produce maps (Kimerling et al., 2009) and deliver maps in raster or vector formats. Raster formats use pixels at a known resolution to convey information, whereas vector formats use geometric primitives that are generally resolution independent. See **Vector Formats & Sources** for more information. In many cases, a map product may be a combination of raster and vector sources. When designing and producing maps, the quality, resolution, and color characteristics of raster sources should be considered. When producing raster output, the device target, such as print, screen display, and the map content, should be considered when making a choice of raster format to use. This entry covers the primary raster formats applicable to cartography, considerations when using them as well as raster source types that are used for map production.

3. Raster Formats

Raster formats share the fundamental building block of a **pixel** (picture elements), the smallest unit of a raster dataset**.** Raster datasets are an array of pixels typically in a rectangular grid (Wise, 2000). Each pixel has a specified number of **bits** (binary digit) that designate the information that can be stored per pixel. In many formats, the bit depth designates the richness of color and the image can be thought of as being constructed of color **bands** where each band represents a color channel of the color raster. For instance, a 24-bit RGB image has 8-bits each for red, green, and blue channels (see **Color Theory**). A 32-bit RGB image adds an extra **alpha channel** that allows for varying transparency in the image (Porter and Duff, 1984). Some data formats are single band with the **bit depth,** or the number of bits storable for a given data type, allocated to each pixel in that band. Elevation data is commonly distributed as 16-bit or 32-bit integer or floating-point raster images with the full value range available per pixel designating an elevation value in the specified unit. The choice of bit-depth for elevation data is dependent on the accuracy and range of the data.

There are numerous raster formats and specifications that leverage raster data. Raster data can be distributed in a format as simple as a text file or in more sensor specific binary formats (see **the Raster Data Model**). There are a few key formats used for cartography on a regular basis as both input and output formats discussed here, including PNG, JPEG, TIFF, and GIF. An understanding of the concepts used for storage and compression in these formats will provide a base of knowledge for understanding other raster formats and raster inclusion in vector formats such as Scalable Vector Graphics (SVG) or Portable Document

Format (PDF). File formats here do not cover the full range of possible raster data formats. Proprietary data formats like MrSID and open variations of a similar compression type are often used. See **the Raster Data Model** for specifics on these and other formats.

3.1 PNG

Portable Network Graphics (PNG) is a raster format which supports **lossless** compression, compression that does not result in the loss of data. Typically used to represent RGB or RGB images with an alpha channel, the PNG format also supports an indexed 8-bit encoding that can be used to efficiently store images with a small number of colors. PNG was originally developed to avoid patent issues with the popular GIF format and has become a standard image format for web usage. In cartographic workflows, PNG is well suited for raster representations of vector data that use few colors, compress well, and will be distributed on the web. The alpha channel capabilities allow for variable transparency and are useful for cases where multiple images are composited. PNG uses the DEFLATE compression algorithm (Roelofs, 2003) which is lossless and unencumbered by patents leading to near universal software support for PNG.

3.2 JPG

The JPEG Interchange Format, or JPEG, is the most common format leveraging the Joint Photographic Experts Group's method for storage of photographic images, which is a type of **lossy** compression, a compression that results in data loss. JPEG compression was designed for images with smooth variations of color, as occur in most photographs or raster datasets representing continuous phenomena. The JPEG compression algorithm is commonly used with a user specified compression ratio that gives the user some control over data loss. Data loss in JPEG compressed images can result in visual artifacts, especially at sharp color boundaries or when using large compression ratios (Bendell et. al, 2006). Sharp color boundaries often occur in images sourced from vector data. Therefore JPEG is not an appropriate image type for sharing the output of work that began as vector graphics, including many maps. Compression choices often must be made on an image-by-image basis due to varying content. Compression over 90% will result in higher quality images, but larger file sizes. Compression below 70% will result in smaller images, but lower quality images in many cases. Multiple edits of a single JPEG file are not recommended as data loss will be incurred with each save. Therefore, it is common to store source images in another format such as TIFF for preservation. JPEG does not support transparency and cannot have an alpha band. JPEG files are typically used for RGB or grayscale images but can be stored with a CMYK color profile for press production needs.

3.3. TIFF

Tagged Image File Format (TIFF) is a widely known generic container format for raster data. TIFF supports multiple band combinations, compression algorithms, and is extensible. GeoTIFF is a well-known extension of the TIFF format which embeds coordinate reference information into the TIFF file (Mahammad and Ramakrishnan, 2009). Consumers of GeoTIFF can use this information to place the raster dataset in geographic space. TIFF consumers that are not GeoTIFF-aware can ignore the information and continue to consume the TIFF. The versatility of the TIFF format and the geographic referencing possible via the GeoTIFF standard make TIFF a common delivery format of geographic information. One downside of TIFF is that it is not a supported raster format in most web browsers.

3.4 GIF

Graphics Interchange Format (GIF) is a raster image format limited to 8-bits per pixel. GIF images use the lossless Lempel-Ziv-Welch compression scheme (CompuServe, 1990). Graphics applications commonly drop colors through **dithering**, a technique for minimizing content loss during reduction of color depth by inserting pixel noise , or other techniques to fit into an 8-bit 256 RGB color palette required for GIFs. From a static image standpoint, GIF images have few advantages over the more flexible PNG format. The popularity of GIF is in its ability to support animations. Animation support in GIFs has some utility in cartography to share small animations in a widely supported format.

3.5 Comparison

Table 1 compares PNG and JPEG image formats for use in cartographic workflows. The two primary workflows considered here are maps coming from vector data sources and maps coming from photographic-style data sources, such as aerial or satellite imagery. Image formats all have tradeoffs, but Table 1 and Figures 1 through 4 show that PNG is generally better for output maps coming from vector data sources and JPEG is generally better for output maps coming from photographic-style data. Note that there’s not often a clear choice with maps that mix vector and raster data sources, but PNG is a better lossless choice when file size is not a concern.

Table 1. Comparison of PNG and JPG Image Formats for Cartographic Workflows

	PNG	JPEG
Vector data sources	The image contains sharp boundaries with no visual artifacts and continuous areas of similar pixels lead to smaller file sizes due to compression. See Figure 1.	The sharp boundaries of the image lead to visual artifacts and the image is poorly compressed, leading to large file sizes. See Figure 2.
Photographic-style data sources	There are no visual artifacts. The image contains many pixels of different color values that do not compress well and lead to larger file sizes. See Figure 3.	Visual artifacts are less noticeable and the pixel values of continuous color compress well for smaller file sizes. See Figure 4.



Figure 1: An 8-bit PNG of a vector-based map. File size 3.6 kilobytes. Image magnified 400% from source image of 96-DPI. The image has smooth transitions between colors from anti-aliasing but no visual artifacts. Source: author.



Figure 2: A JPEG using 85% compression, a medium-quality compression amount, of the same vector-based map shown in Figure 1. File size 18.3 kilobytes. Image magnified 400% from source image of 96-DPI. The image shows strong visual artifacts caused by the sharp color transitions in the image and is larger in size than the 8-bit PNG. Source: author.

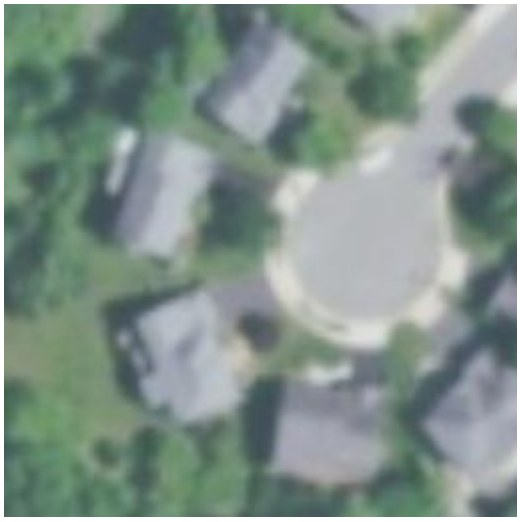


Figure 3: A 256 x 256 pixel image in PNG 24-bit format. The file is 131 kilobytes. Source: author.



Figure 4: The same image as in Figure 3, but in JPEG format with 85% compression. The file is 8.5 kilobytes. Source: author.

3.6 Resolution

Resolution of raster data is a key characteristic defining its quality and appropriateness for use. Raster **resolution** describes the number of distinct pixels that can be displayed in a measured space and is typically defined as dots per inch, or DPI, a measurement based on print quality of raster files. Raster datasets may describe resolution in terms of a pixel's geographic measurement representation (e.g., 1 pixel equals 30 cm). Dots in the DPI context refers to physical dots from printing or photographic processes and is not technically applicable to modern display technology. The term pixels per inch, or PPI, is often used interchangeably with DPI. Screen resolution has historically been lower than print resolutions, so screen-only use of raster data has lower demands on image resolution. With the advent of high-DPI mobile devices and monitors, resolution requirements of

images have increased and must be considered with cartographic design. For instance, a map produced for the web with only a low-DPI image available will be upscaled on high-DPI devices and appear less clear than if a high-DPI version were available. One solution to this issue is to create both low and high-DPI versions of the image and select the appropriate one for the device. Figure 5 shows a magnified standard 96-DPI image used for standard web display while Figure 6 shows a 192-DPI image commonly used for high-DPI web display. Note the lower resolution of Figure 5 leads to the visual appearance of pixels in this 200% magnification as a user would see using this resolution in a high-DPI display. The overhead of this approach has led to an increase in vector sources in design work. Vector sources and vector output formats can be realized at the native resolution of the display mechanism and do not suffer from this issue.

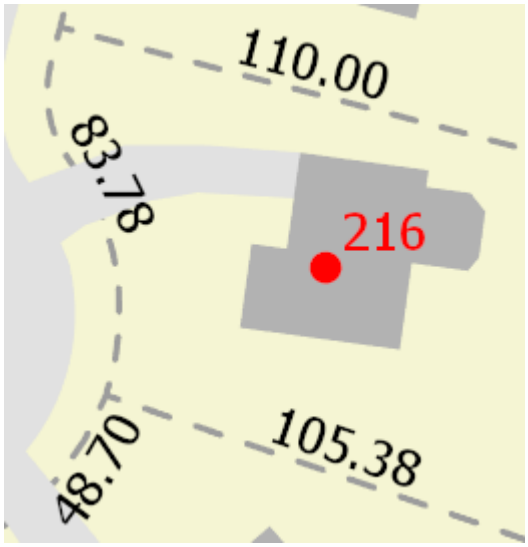


Figure 5: A 96-DPI image that needs to be upscaled when displayed on a high-DPI display. Source: author.

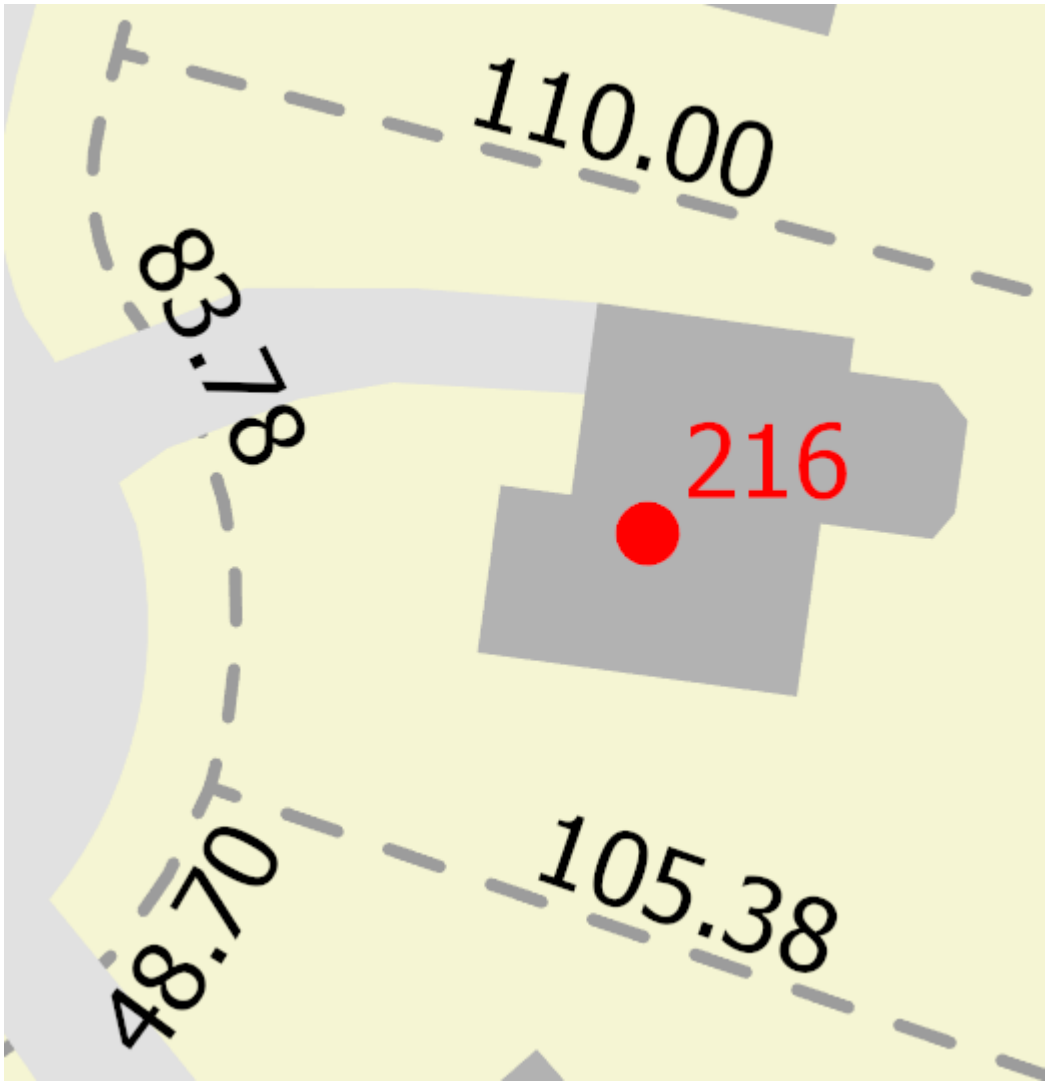


Figure 6: A 192-DPI image at its naïve resolution to show clarity of the image on a high-DPI display. Note it has 4 times the number of pixels as the 96-DPI image. Source: author.

3.7 Anti-aliasing

Vector artwork converted to a raster image must be fit to the raster grid. In low resolution cases, the grid fitting may result in visual artifacts. This is especially noticeable with angled lines where the filled raster pixels may appear like stairsteps. These artifacts are referred to as aliasing**. Anti-aliasing** techniques are designed to minimize the visual artifacts of grid fitting to low resolution displays (Freeman, 1974). While many anti-aliasing techniques exist, the simplest techniques are resampling of higher resolution images into lower resolution images, with the resampling resulting in gradations of color at sharp boundaries which smooth the overall appearance of the artwork. The Figure 7 below shows a magnified view of aliasing of a raster image generated without anti-aliasing techniques. Note the stair-like behavior and blocky appearance of text. Figure 8 shows anti-aliasing resulting from a 10 x 10 grid of pixels being resampled into 1 pixel for a smoother appearance. Note the blending of colors between the circle and gray building underneath to create a smoother transition between the circle and building as well as a more circular shape than the aliased case.



Figure 7: Vector graphics stored in a PNG without anti-aliasing. Image magnified 400% from an original 96-DPI source. Source: author.



Figure 8: Vector graphics stored in a PNG with anti-aliasing. Image magnified 400% from an original 96-DPI source. Source: author.

4. Raster Sources

Raster sources for cartographic work can be thought of in three categories: data fields,

imagery, and map products. Common raster sources and example locations to acquire them can be found in Table 2.

Table 2. Common Raster Sources and Example Locations

Raster Source	Provider	Example
Data fields	Government agencies and international consortiums. Often re-hosted by industry specific sites	A digital elevation model combining multiple sources is available for European Union countries via the European Environmental Agency.
Imagery	Satellite or aerial photography vendors and government agencies. Often re-purposed for general use by industry specific sites.	Landsat data can be explored and downloaded through the United States Geological Survey.
Map products	Government agencies and international consortiums.	Nautical charts for the United States are available from the NOAA Office of Coast Survey

4.1 Data fields

Raster datasets representing a single data field, such as digital elevation models, are common raster sources for cartographic work. Digital elevation models may be further processed into hillshades, contours, or hypsometrically-tinted surfaces (see **Terrain Representation**). While elevation is the most common data field raster cartographers work with, other phenomena including non-elevation interpolated surfaces generated from measurements may be represented in a similar format. Climate, weather, oceanographic, and atmospheric data expand on this by providing single band measurements in multi-dimensional formats, optimized for time series data measured or simulated at different altitudes or depths. **LIDAR** (Light Detection and Ranging, a technology for collecting data using laser reflectance) data sources are also often used to produce raster data for a single data field of specific light returns (see **Remote Sensing Platforms**). For instance, different return values may result in a digital terrain model and a digital surface model, with the surface model taking vegetation and human built objects into account.

4.2 Imagery

Raster data sources based on imagery are plentiful and may come from a variety of sources. Once only available with aerial photography, the widespread use of satellite imagery and drones have allowed for more widely available and frequently updated imagery (see **Unmanned Aerial Systems**). While we tend to think of imagery as what is visible to the human eye, sensors can be equipped to collect more than the visible spectrum with measurements of various ranges of infrared and thermal phenomena (see **Nature of Multispectral Imagery**). Image data sources may be single or multi-band. Modern sensors are often multi-spectral, but historic data sources may only be available in single band panchromatic imagery.

4.3 Map Products

Completed map products are often used as a source for additional map production. For instance, a topographic map may be combined with a hillshade to produce a new product with an alternative approach on representing terrain. Often the existing map product is in

the form of a basemap and the new cartographic content is simply overlaid in top (see **Basemaps,** forthcoming). This approach is common in web cartography where basemaps are designed specifically for this purpose. When using existing raster products, including scanned historical sources or pre-created tile sets, the resolution of the dataset may limit the final output product.

5. Raster Map Production

The final output product should be considered when creating a raster map product (see **Map Production and Management**). Approaches for the web are based on the RGB color model. Approaches for print are typically based on the CMYK color model used in traditional offset printing. CMYK may be augmented with additional spot colors as needed. Resolution is also a consideration: will the product be viewed on high DPI devices on the web? In the modern era this is almost always true. For print, the output printing device should be known and is typically higher resolution than the screen resolution.

5.1 Web

Graphics on the web generally use subtractive RGB model, and in many cases assumed to be in the sRGB **color space (**Pemberton and Pettit, 2018). While sRGB is a common color space, it has a limited color **gamut**, or a subset of colors in a color model representable in a color space. (see **Color Theory**). Recent developments with wide-gamut color spaces hope to address the limited range but are not widely supported by computer monitors. Therefore, the recommendation is to produce RGB images with the sRGB color space.

5.2 Print

Printing uses additive color models such as CMYK. In specialized circumstances, spot colors may be used to designate specific colors rather than relying on the four color CMYK process. Traditional map production of offset printing with process or spot colors has involved the creation of separate color images representing each plate that will be inked and used at print time. TIFF color separates were commonly used for this step and still are used in some cases today. Preparation went beyond color separates with specifications for how colors would be combined. Process colors are mixed using halftoning techniques that place dots in regular patterns, or screens, to produce the mixed color. These regular patterns can be distracting if not varied by plate, so cartographers managed screen angles for plates to produce the desired mix. With innovations in color management, the process is simpler and is handled through the delivery of color managed graphics to the press. This is especially true with many products being combinations of vector and raster sources with raster photographic sources commonly being RGB (King, 2001). These compound documents may be constructed of both CMYK vector graphics and RGB images. If both cases are color managed in a compound document, press software can produce process color separates for production using a raster image processing software. Halftoning screen angles can automatically be calculated based on the image content and the capabilities of the device. Offset printing often involves a proof step, where the output is manually inspected before going to full production. Updates to color management parameters and halftoning screen angles can be made at this time.

6. Future Directions

Improvements in video formats and video compression have led to the development of new raster formats, such as WebP (Alakuijala, 2012) and AV1 (Concolato, et al., 2019). These formats use both storage containers and compression algorithms developed for video with added benefits like wider **gamut** support (Concolato, et al., 2019). Slow implementation of these formats in software such as browsers hinders adoption, but they continue to be formats to watch. Collection of raster data through cameras and other sensors is on the rise. New formats will be developed to handle this data and image processing software will be adapted to handle them. Cloud computing and graphics processing units provide powerful computing capabilities that will continue to grow. Higher resolution raster data with better color fidelity will be available as sensors collect wide-gamut images and devices are able to display them.

References

Alakuijala, J. (2012.) WebP Lossless Bitstream Specification. Google.

Bendell, C., Kadlec, T., Weiss, Y., Podjarny, G., Doyle, N., McCall, M. (2016). High Performance Images: Shrink, Load, and Deliver Images for Speed. O'Reilly Media.

CompuServe (1990). Graphics Interchange Format.

Concolato, C., Klemets, A., and Kerr, P. (2019) AV1 Image File Format (AVIF). The Alliance for Open Media.

Dent, B. D., Torguson, J. S., & Hodler, T. W. (2008). Cartography: Thematic Map Design. 6th Edition. Boston: McGraw-Hill.

Freeman, H. (1974). Computer processing of line-drawing images. Computing Surveys, 6, 54-97.

Kimerling, A. J., Buckley, A. R., Muehrcke, P.C., and Muehrcke, J. O. (2009). Map Use: Reading, Analysis, Interpretation (6th edition). Redlands, California: Esri Press.

King, J. C. (2001). Why Color Management? Adobe Systems Incorporated.

Mahammad, S. and Ramakrishnan, R. (2009). GeoTIFF – A standard image file format for GIS applications. Geospatial World.

Pemberton, S., Pettit, B., (2018). Çelik, T., Lilley, C., Baron, L.D., eds. CSS Color Module Level 3. W3C. Section 4.2.1. RGB color values.

Porter, T. and Duff, T. (1984). Compositing digital images. ACM SIGGRAPH Computer Graphics, 18(3): 253-259.

Roelofs, G. (2003). PNG: The Definitive Guide (2nd ed.). O'Reilly Media.

Slocum T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). Thematic Cartography and Geographic Visualization (3rd edition). Upper Saddle River, NJ: Pearson/Prentice Hall.

Wise, S. (2000). GIS data modelling-lessons from the analysis of DTMs. International Journal of Geographical Information Science, 14(4), 313-318.