Geospatial imagery formats

File Formats

Common structures

World files

Some geospatial raster file formats store information about the alignment of the image in physical space within the primary data file. But some store it in a related but separate "world" file.

Examples: .tfw [for .tif], .j2w [for .jp2], .sdw [for .sid], .hdr [for .flt]

Pyramids

Rasters can be very large. To speed up display and exploration of datasets, geospatial software often generates a resampled, lower resolution version of the raster as a "pyramid" file. These could potentially be deleted without compromising the primary data files.

Examples: .ovr, .rrd, and .rde [for .tif in ArcGIS Desktop], .qvr [for .las in LP360]

Auxiliary files

Rasters often are structured as a grid where each cell contains a single value. Because of this supplementary information about the data may be stored in a related but separate "auxiliary" file.

Examples: .aux and .aux.xml [for .tif], .aux [for .jp2]

Major Formats

Processed Imagery

.tif

TIFF is a tag-based file format for storing and interchanging raster images. TIFF serves as a wrapper for different bitstream encodings for bit-mapped (raster) images.

GeoTIFF is format extension for storing georeference and geocoding information in a TIFF 6.0 compliant raster file by tying a raster image to a known model space or map projection

Extension	Description	Notes
.tif or .tiff	primary raster image	May or may not include location

		information in the header
.tfw	location data when not stored in the header of the .tif	If you open this in a text editor it will be 6 numbers
.ovr and .rrd	generated in ArcGIS to display large raster files more quickly (ie. a lower resolution image created by "pyramid resampling")	
.rde	If an .rrd file exceeds two gigabytes, a reduced resolution dataset external (.rde) raster data file is created as a supplementary file	
.tif.aux.xml	metadata about how .ovr/.rrd pyramid files were generated	
aux and aux.xml	stores any auxiliary information that cannot be stored in the raster file itself, such as: ● Color map ● Statistics, histogram, or table ● Pointer to the pyramid file ● Coordinate system ● Transformation ● Projection information	

.sid

MrSID Image Format (Multi-resolution Seamless Image Database) is a patented, wavelet-based file format designed to enable portability of massive bit-mapped (raster) images. LizardTech's proprietary MrSID format is commonly used for orthoimages in need of compression. Color images can be compressed at a ratio of over 20:1. LizardTech's GeoExpress is the software package capable of reading and writing MrSID format.

Extension	Description	Notes
.sid	primary raster image	
.sdw	contains georeferencing information that maps the raster image data in the SID file to real-world coordinates; contains the units of measurement for the image pixels as well as the coordinates of the upper-left hand pixel	

.jp2

[JPEG 2000 Part 1 (Core) jp2](link) is a compressed bitmap image saved in the JPEG 2000 (JP2) Core Coding format. It has the ability to wrap object types beyond JPEG 2000 core bitstreams, e.g., auxiliary opacity/transparency channels, color profile information, and other metadata.

Extension	Description	Notes
.jp2	primary raster image	May or may not include location information in the header
.j2w	location data when not stored in the header of the .jp2	
.aux	stores any auxiliary information that cannot be stored in the raster file itself, such as: - Color map - Statistics, histogram, or table - Pointer to the pyramid file - Coordinate system - Transformation - Projection information	

.jpg

[JPEG](link) is a commonly used method of lossy compression for digital images, particularly for those images produced by digital photography. The degree of compression can be adjusted, allowing a selectable tradeoff between storage size and image quality.

Extension	Description	Notes
.jpg	primary raster image	
.jgw	location data when not stored in the .jpg	

.img

There are [multiple imagery software and sensors](link) that save data with this file extension, including ERDAS IMAGINE, ENVI, Image Display and Analysis (IDA) and Planetary Data System (PDS).

[ERDAS IMAGINE .img/.ige](link) represent a proprietary, partially documented format for multi-layer geo-referenced raster images developed originally for use with ERDAS IMAGINE software. This format is used widely for processing remote sensing data, since it provides a framework for integrating sensor data and imagery from

many sources. A key capability of ERDAS_IMG is that it distinguishes between two types of raster layer: 1) Continuous: where the value of each pixel relates to a quantity measured in a continuous range, including aerial or satellite imagery, temperature, elevation, etc. and 2) Thematic: where pixels have been classified, or put into distinct categories

Extension	Description	Notes
.img	primary raster image	
.ige	produced using ERDAS image processing software if image size is greater than 2GB	linked to an .img file; open this file and it will point to the associated .ige

SRP – Standard Product Format. The ASRP and USRP raster products (as defined by DGIWG) are variations on common standard product format. ASRP (in a geographic coordinate system) and USRP (in a UTM/UPS coordinate system) products are single band images with a palette and georeferencing.

Extension	Description	Notes
.img	main raster image	
.gen	general information file	
.sou	source file	
.qal	quality file	
.thf	transmission header file	

.ecw
Enhanced Compressed Wavelet (ECW) is a proprietary (ERDAS), wavelet-based, lossy compression format, similar to JPEG 2000. (More information)

Extension	Description	Notes
.ecw	compressed raster image data	map projection information can be embedded into the ECW file format

.bil
Band Interleaved by Line is a binary raster file format for aerial photography, satellite imagery, and spectral data. The BIP data organization can handle any number of bands, and thus accommodates black and white, grayscale, pseudocolor, true color, and multi-spectral image data.

Extension	Description	Notes
.bil	a raster storage extension for single/multi-band aerial and satellite imagery	stores pixel information based on rows for all bands in an image
.hdr	a header file that describes the number of columns, rows, bands, bit depth and layout in an image	
.bsq	Band sequential format (BSQ) stores separate bands by rows.	

Legacy file formats

.adf

ESRI ArcInfo Grid (also known as ArcGrid) is a raster format which stores numeric values representing a geographic attribute (such as elevation or surface slope) as either a 32-bit signed integer or a 32-bit (single precision) floating point number. The native binary ESRI Grid format comprises a number of component files. These files are stored within an ArcInfo workspace, which may incorporate several grids and other geospatial datasets, such as coverages.

Extension	Description	Notes
dblbnd.adf or bnd.adf	boundary file containing the minimum and maximum x,y coordinates for an ESRI_grid	INFO format All grid BNDs are stored in double precision. This file will have corresponding .nit and .data files in the INFO directory of an ESRI_grid
hdr.adf	header file containing information about the ESRI-grid's cell resolution, type (integer or floating point), compression, and tile information	Binary format
log	log file containing a history about the creation of and alterations to an ESRI_grid	ASCII format
prj.adf	projection file storing the known map projection information of the ESRI_grid	ASCII format - optional
sta.adf	the statistics (minimum, maximum, mean, and standard deviation) of the cell values	INFO format Values in the statistics are floating point values.

		This file will have corresponding .nit and .data files in the INFO directory of an ESRI_grid.
vat.adf	the value attribute table (VAT) storing the attribute data associated with the zones of the grid	INFO format Can only exist for integer grids. This file will have corresponding .nit and .data files in the INFO directory of an ESRI_grid.
w001001.adf	the value of each cell in the first (base) tile of the grid	Binary format Subsequent tiles will be numbered sequentially
w001001x.adf	the index for the blocks in the tile	Binary format If the grid has more than one tile, the value and index .adf file names will increment correspondingly.
INFO subdirectory/ arcNNNN.dat	the relative path to the grid data file	INFO format
INFO subdirectory/ arcNNNN.nit	file structure and field definition information of the grid	INFO format For any particular grid, the NNNN values for .dat and .nit should match.
INFO subdirectory/ arc.dir	the data dictionary for the grid	

ESRI ArcInfo Grid interchange formats (.asc and .flt)

ESRI ArcInfo ASCII Grid and ESRI GridFloat Output File are interchange formats used primarily for exchange with other programs. They have very similar internal structure (with differences detailed in the section below).

ESRI ASCII Grid (.asc) is an output format created by the GRIDASCII command (from ArcInfo Workstation) and by the Raster to ASCII tool in ArcGIS for Desktop.

Extension	Description	Notes
.asc	holds a single measure for each cell in a rectangular grid	

ESRI GridFloat(.flt) is an output format created by the GRIDFLOAT command (from ArcInfo Workstation) or by the Raster to Float tool in ArcGIS for Desktop. Unlike the ESRI ASCII Grid (.asc), gridded data values are in binary form, typically resulting in smaller files. ESRI_GridFloat is a pair of files, with a simple text file with extension

.hdr that contains the same information as the first six lines of the equivalent ESRI_ASCII_Grid with one additional line (in addition to the primary raster data file).

Extension	Description	Notes
.flt	holds a single measure for each cell in a rectangular grid	IEEE floating-point 32-bit (aka single-precision) signed binary format
.hdr	defines the properties of the grid, such as the cell size, the number of rows and columns, and the coordinates of the origin of the rectangular grid	text file

Raw Imagery / Returns

.las / .laz

LAS is a file format for the interchange of 3-dimensional point cloud data. It was primarily designed for the exchange of LIDAR (Light Detection and Ranging) point cloud data, but supports the exchange of any 3-dimensional x,y,z tuplet data. LAZ, a losslessly compressed variant of LAS.

Extension	Description	Notes
.las	point cloud data	open, binary format
.laz	compressed point cloud data	
.qvr	pyramid file created to improve display of .las data in LP360	

Indexes

.shp

ESRI Shapefile is a format that stores non-topological geometry and attribute information for vector spatial features in a data set.

Extension	Description	Notes
.shp	geometry	required

.dbf	attributes	required
.shx	index	required
.prj	projection	important
.xml	metadata	important
.gpj	projection file created when shapefiles are generated in QGIS	optional
.cpg	codepage for identifying the character set used	optional - recommended if file includes non-ASCII characters
.sbx	spatial index of the features	optional
.sbn	spatial index of the features	optional
.atx	attribute index that was created for each shapefile or dBase table in ArcCatalog for ArcView GIS 3.X	legacy - replaced with a new form of indexing in ArcGIS 8

.gdb

ESRI File Geodatabase (FGDB Specification) is the primary data storage model for ArcGIS software products and services. It is a container for spatial and attribute data, enabling storage of many different types of GIS data (both vector and raster) within its structure. The database is implemented as a folder of binary files in the file system managed by the user's operating system. The folder may be on the local computer or a network-accessible file system and may have many files, including two or more tables for each dataset within the database and other supporting files.

Extension	Description	Notes
.gdbindexes	lists the indexes that may exist on certain fields of a .gdbtable	This only applies to FileGDB v10 .gdbindexes : v9 .gdbindexes have a different (and more complicated) structure.
.gdbtable	describe fields and contain row data	an header, a section describing the fields, and a section describing the rows
.gdbtablx	contains the offset of the rows of the associated .gdbtable file.	
.spx	spatial index for the geometry field of a .gdbtable	
.prj		
.freelist	contain the offset to the holes (rows deleted, or old updates) in	optional, deleted if the FGDB is compacted

	the associated .gdbtable file	
.atx	contains indexes for a field of a .gdbtable	the values that the field takes in the .gdbtable are listed in ascending order with the associated FID.

.mdb

ESRI Personal Geodatabase is an option for data storage for a single-user Esri geodatabase that is implemented as a single Microsoft Access file. It has been replaced in functionality and accessibility by File Geodatabases.

Extension	Description	Notes
.mdb		Legacy - these cannot be opened in ArcGIS Pro

Types of imagery

Stereo imagery

multiple overlapping images collected as a sensor flies along a flight path

Orthoimagery

georeferenced images of the Earth's surface that have been collected by a sensor and then had image object displacement removed (by correcting for sensor distortions and orientation as well as terrain relief)

Digital orthophoto quadrangle (DOQ and DOQQ)

A subset of orthoimagery. The DOQ format was developed by the United States Geological Survey (USGS) and used for orthorectified digital aerial photographs and satellite images from the late 1980s through 2006. DOQs are cast to the UTM projection and referenced to the NAD 27 or NAD 83 datum. DOQs may be black and white (B/W), natural color, or color-infrared (CIR) images. DOQ refers to both the 1-meter ground resolution quarter-quadrangle digital orthophoto (3.75- x 3.75-minutes in extent) and the one or two-meter ground resolution quadrangle digital orthophoto (7.5- x 7.5-minutes in extent). The acronym DOQQ is used for quarter-quadrangle orthophotos when the distinction from the quadrangle DOQ orthophotos is necessary. Digital Orthophotos (DOQ & DOQQ) - Library of Congress

Orthomosaic

orthoimages that have been edge-matched and color balanced to create a seamless mosaic