

Extracted from:

Desktop GIS

Mapping the Planet with Open Source

This PDF file contains pages extracted from Desktop GIS, published by the Pragmatic Bookshelf. For more information or to purchase a paperback or PDF copy, please visit <http://www.pragprog.com>.

Note: This extract contains some colored text (particularly in code listing). This is available only in online versions of the books. The printed versions are black and white. Pagination might vary between the online and printer versions; the content is otherwise identical.

Copyright © 2008 The Pragmatic Programmers, LLC.

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior consent of the publisher.



The book you're reading is still under development. As part of our Beta book program, we're releasing this copy well before we normally would. That way you'll be able to get this content a couple of months before it's available in finished form, and we'll get feedback to make the book even better. The idea is that everyone wins!

Be warned. The book has not had a full technical edit, so it will contain errors. It has not been copyedited, so it will be full of typos and other weirdness. And there's been no effort spent doing layout, so you'll find bad page breaks, over-long lines with little black rectangles, incorrect hyphenations, and all the other ugly things that you wouldn't expect to see in a finished book. We can't be held liable if you use this book to try to create a spiffy application and you somehow end up with a strangely shaped farm implement instead. Despite all this, we think you'll enjoy it!

Throughout this process you'll be able to download updated PDFs from your account on <http://pragprog.com>. When the book is finally ready, you'll get the final version (and subsequent updates) from the same address. In the meantime, we'd appreciate you sending us your feedback on this book at <http://books.pragprog.com/titles/gsdgis/errata>, or by using the links at the bottom of each page.

Thank you for being part of the Pragmatic community!

► **Andy Hunt**

Data Formats

One of the challenges in working with GIS software, whether it be proprietary or open source, is making sense of the many data formats you encounter. Let's take a look at some of the common formats you will encounter so you can get an idea of what's out there. We'll also look at where these data formats come from, some of the conversion options, and lastly how to choose a standard data format for your mapping projects.

6.1 Common Formats

So far, we've indirectly discussed a number of formats, including shapefiles, GeoTIFF, grids, PostGIS, and GRASS vector and raster. If you are a casual or intermediate user of OSGIS software, odds are you are going to be using only a few data formats on a regular basis. Typically this means working with the following:

- Shapefiles (.shp)
- GeoTIFF or TIFF with world files (.tif, .tfw)
- JPEG with world files (.jpg, .jpw)
- GPS data (.gpx)

In fact, these are pretty common, and you can accomplish a lot with just these formats. Other vector formats you might run across during your search for data include the following:

- ArcInfo Binary Coverage
- ArcInfo Interchange File (.e00)

- MapInfo (.tab, .mid, .mif)
- SDTS, a data transfer standard for both vector and raster data¹
- Topologically Integrated Geographic Encoding and Referencing (TIGER) data, used and distributed by the U.S. Census Bureau
- Digital Line Graphics (DLG)

There are a lot of raster formats you might encounter, including the following:

- ERMapper Compressed Wavelets (.ecw)
- Erdas Imagine (.img)
- Digital Elevation Models (.dem)
- JPEG 2000 (.jp2, .j2k)
- Multi-Resolution Seamless Image Database (MrSID) (.sid)
- GTOPO30, a global digital elevation model (DEM) derived from a number of raster and vector sources

Some of the formats are open, meaning that they have a published specification and you can use it to write applications and utilities that work with the format. Others are closed, requiring you to use the vendor-provided API. Of course, this is a concern only if you want to write your own applications and utilities. If you are content with using the OSGIS applications available, someone else has done the hard work for you.

Although it's not important to understand these formats to use them, it does help to know a bit about them so you can determine whether your favorite OSGIS software supports the format. In case you haven't realized it yet, the GDAL/OGR library supports a huge range of vector and raster formats—see Section [A.2, GDAL/OGR](#), on page [298](#) for lists of additional vector and raster formats you might encounter.

The good news is if you are using OSGIS software that uses GDAL/OGR for accessing raster and vector data (such as GRASS or QGIS), then you have access to most, if not all, of the formats listed.

In Section [11.2, Using GDAL and OGR](#), on page [195](#), we'll look at using the GDAL/OGR utilities to get information about our data as well as convert and transform both raster and vector layers.

1. See <http://mcmcweb.er.usgs.gov/sdts>

Web-Deliverable Data

Another “format” you’ll encounter is data deliverable over the Web. This category of data is often referred to as W*S. The moniker W*S is attached to standards for delivering geospatial data over the Web and includes Web Mapping Service (WMS), Web Features Service (WFS), and Web Coverage Service (WCS). A good chunk of the web mapping applications you might use in your browser get some or all of their data from a W*S service.

Many of our desktop applications include support for at least WMS. This allows us to include data from across the Internet in our mapping projects. The good thing is you don’t have to understand the standard or how it works; you just use it and get good data for free.

If you want to get real technical information on WMS, WFS, and WCS, you can find standards on the Open Geospatial Consortium website.²

6.2 Choosing a Standard Format

You might be wondering about a standard format for all your projects. This isn’t strictly necessary, although it might make your life easier. Assuming your OSGIS software can handle a multitude of formats, there may be no reason to convert.

Reasons to Standardize

There are some valid reasons to standardize on a data format for your raster and vector data. In reality, you’ll probably still have some data that’s not in your standard format. Let’s take a look at a some reasons why you might want to convert to a standard format.

Data Management

GIS data can have several unsavory characteristics—and we’re not talking about accuracy or quality. As you begin to work with data, transform it, analyze it, and so forth, you’ll find the following:

- It propagates rapidly.
- It grows and hides in places you never expect.
- Unchecked, it rapidly becomes unmanageable.

2. <http://www.opengeospatial.org>

Some (OK, us) have gone so far as to call GIS data an illicit drug.³

You may be wondering how converting to a standard format will improve data management. Well, it's not a silver bullet, but it can aid in creating a logical structure for storing your data. For example, if all your vector data is in shapefiles and you can create a nicely organized directory structure, be it by theme or by project, then your ability to find and use the data you need increases.

Some find that data management is improved by using a spatial database, although there are perhaps better reasons to use one. By storing your vector and/or raster data in a spatial database, you provide one point of entry for all your data needs. There is no question of which server or directory you need to search to find the data you need. We'll look at spatial databases in more detail in Chapter 7, *Spatial Databases*, on page 107.

Another example is GRASS, which uses its own format for storing both raster and vector data. Data in GRASS is organized by location and mapset, making it easy to structure your data collection in a way that can be more easily managed.

Is improved data management alone a reason for converting to a particular format? Probably not, especially if you're talking about converting between file-based formats such as shapefiles and GeoTIFFs. The chief considerations are making the data discoverable, accessible, and usable. If your formats of choice include file-based data, you should create a structured logical directory layout and naming convention and adhere to it. This will make managing your data much easier and prevent the multiplication of duplicate data sets. Another important management tool is metadata that documents each of your datasets. If you want to take the formal approach and create metadata in a format that others will understand, use the standard.⁴ At the very least, you should include a text file describing the data, its origin, and the processing steps used to create it. Fortunately, the metadata standard includes all these components, so you may find it's worth using.

Improved Functionality

If you want to do more than display and edit spatial data, then conversion to gain improved functionality is certainly a worthwhile consider-

3. <http://spatialgalaxy.net/2006/03/29/gis-data-is-an-illicit-drug/>

4. <http://fgdc.gov/metadata>

ation. For example, although QGIS is great at displaying a wide range of raster and vector data and is even able to digitize and edit features, it lacks the ability to perform common GIS operations. Harrison's simple analysis of bird sightings using a buffer (Figure 1.3, on page 22) requires something with geoprocessing capability. In this case, Harrison could have converted his digitized lakes (which likely began life as a shapefile) into a PostGIS or GRASS layer. Both of these give him the functionality he needs to create a buffer. In the case of PostGIS, it's done with a fairly simple SQL query. In GRASS, he can use the `v.buffer` command or the buffer module in the QGIS-GRASS toolbox.

PostGIS is a good example of a reason to convert. Not only does it improve data management by giving you a “portal” to your data, but it has been certified as OGC compliant and provides the spatial functions specified in the Simple Features Specification for SQL. This means that not only can we display and edit PostGIS data in an application-like QGIS, we also get a whole batch of spatial functions that we can use to query the relationships between features, transform between projections, and create new features. If you find that your work requires more than just simple viewing and editing, then conversion is worth considering.

When software is certified as OGC compliant, you can be assured that it adheres to established standards and can interoperate with other compliant software.

Enhanced GIS Capabilities

The other reason to convert is to gain enhanced GIS functionality. You're probably asking what the difference is between this and the improved functionality aspect we just covered. You can view it as a progression to more powerful and perhaps complex GIS operations. We're pretty much talking about GRASS and its bountiful assortment of vector- and raster-processing tools. Since GRASS stores data in its own format, we need to convert our existing data in order to take advantage of the tools.

Examples of the type of capabilities we're talking about include the following:

- Line-of-sight analysis
- Union and intersection of layers to create a new layer
- Merging raster data
- Mathematical operations on grids
- Contouring

We'll dive into some of these topics later in Chapter 10, *Geoprocessing*, on page 159.

6.3 Conversion Options

So, you decide to convert some or all your data to a new format. The next question is, what tools are available to do the job? Fortunately, in the OSGIS world, conversion between formats is not only commonplace but easy as well.

If you choose to migrate all your data to PostGIS or GRASS, it's not a problem. Both provide the routines to import your data and export it should the need arise.

GRASS Conversion

GRASS provides both vector and raster import/export functions for a nice range of formats. To give you an idea of the capabilities, here is a partial list of the import commands and formats supported by GRASS:

r.in.arc

Converts an ESRI ARC/INFO ASCII raster file (GRID) into a (binary) raster map layer

r.in.ascii

Converts ASCII raster file to binary raster map layer

r.in.aster

Imports, georeferences, and rectifies an ASTER image

r.in.gdal

Imports a GDAL-supported raster file into a binary raster map layer

r.in.srtm

Imports Shuttle Radar Topography Mission (SRTM) .hgt files into GRASS

r.in.wms

Downloads and imports data from WMS servers

v.in.ascii

Creates a vector map from ASCII points file or ASCII vector file

`v.in.db`

Creates new vector map (point layer) from database table containing coordinates

`v.in.dxf`

Converts AutoCad DXF files to GRASS format

`v.in.e00`

Imports an ArcInfo export file `.e00` to GRASS format

`v.in.garmin`

Downloads waypoints, routes, and tracks from a Garmin GPS receiver into a vector map

`v.in.gpsbabel`

Downloads waypoints, routes, and tracks from a GPS receiver or a GPS ASCII file into a vector map using formats supported by `gpsbabel`

`v.in.ogr`

Converts OGR-supported formats into a GRASS vector map

You can see from the list of commands that there are a lot of options for getting your data into GRASS. In fact, we didn't list all of them for you, just some of the major ones. For exporting data out of GRASS, there are also a lot of options. We won't list them here, but in case you're curious, the commands are all of the form `r.out.*` for rasters and `v.out.*` for vectors.

PostGIS Conversion

If you choose PostGIS, it supports the loading of data using SQL and the importing/exporting of shapefiles using `shp2pgsql` and `pgsql2shp`. We'll take a look at these two utilities in Section 11.4, *PostGIS*, on page 214. If your source data isn't in shapefile format, you can still import it—you just need a little extra power. In this case, you need to use the Swiss Army knife of conversion tools.

The Swiss Army Knife

Just as you wouldn't go out into the wilderness without your Swiss Army knife (or maybe bug spray), venturing into the world of data conversion without the GDAL/OGR utilities is not advised. These utilities provide conversion between file-based vector and raster formats, as well as spatial databases.

The Pragmatic Bookshelf

The Pragmatic Bookshelf features books written by developers for developers. The titles continue the well-known Pragmatic Programmer style, and continue to garner awards and rave reviews. As development gets more and more difficult, the Pragmatic Programmers will be there with more titles and products to help you stay on top of your game.

Visit Us Online

Desktop GIS's Home Page

<http://pragprog.com/titles/gsdgis>

Source code from this book, errata, and other resources. Come give us feedback, too!

Register for Updates

<http://pragprog.com/updates>

Be notified when updates and new books become available.

Join the Community

<http://pragprog.com/community>

Read our weblogs, join our online discussions, participate in our mailing list, interact with our wiki, and benefit from the experience of other Pragmatic Programmers.

New and Noteworthy

<http://pragprog.com/news>

Check out the latest pragmatic developments in the news.

Buy the Book

If you liked this PDF, perhaps you'd like to have a paper copy of the book. It's available for purchase at our store: pragprog.com/titles/gsdgis.

Contact Us

Phone Orders:	1-800-699-PROG (+1 919 847 3884)
Online Orders:	www.pragprog.com/catalog
Customer Service:	orders@pragprog.com
Non-English Versions:	translations@pragprog.com
Pragmatic Teaching:	academic@pragprog.com
Author Proposals:	proposals@pragprog.com