

Potential Barriers to Use of Regionalized Datasets and Some Thoughts on Overcoming Them

Donald L. Schrupp, Wildlife Ecologist
Denver, Colorado



**OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011**



Considerations

- * Finding & Assessing Datasets
- * Retrieving Datasets
- * Using Datasets
 - Documentation / Metadata
 - Coverage & Completeness
 - Projections and Re-projection
 - Data Extraction and Conversion
 - Formats and Schemas
- * Compute Resources
 - Software
 - Hardware

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011



Finding and Assessing Datasets

The Google Approach:



The Clearinghouse/Portal Approach (Feds):

<http://geo.data.gov/geoportal/catalog/search/searchpage.html>

<http://irma.nps.gov/>

<http://fsgeodata.fs.fed.us/index.html>

<http://www.blm.gov/nstc/gis/GISsites.html>

<http://datagateway.nrcs.usda.gov>

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



**OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011**



Finding and Assessing Datasets

The Clearinghouse/Portal Approach
(States/Academia/NGOs):

State of Colorado: <http://ndis.nrel.colostate.edu/>

Colorado State University: <http://gis.colostate.edu>

NGO: <http://www.rmbo.org/public/monitoring/mapmcb.aspx>

(Commercial):

ESRI: <http://resources.arcgis.com/>

Others ?

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011



Finding and Assessing Datasets

* Metadata

What is the coverage of the dataset (spatially) ?

What is the coverage of the dataset (is it complete) ?

Is a detailed data dictionary available ?

Are data Views available for multi-table datasets ?

Is the dataset well documented ?

Is an accuracy assessment provided ?

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Retrieving Datasets

- * Are CD/DVDs Available for the Regionalized Dataset ?
- * Is the dataset available for FTP download ?

Are various data layers available individually ?

Is a MapServer available with WFS/WMS support
for 'previewing' the data

Does the download site support 'Clip/Zip/Ship'
if you only need a subset area of the dataset ?

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations: OVERVIEW

Extent of Coverage(s)

Data Table Schemas / Linked Tables Views

Projection Differences

Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

Software Options

Proprietary ArcGIS / ENVI / ERDAS Imagine

Open Source GRASS GIS / Quantum GIS /
OSSIM / MDB Tools / uDIG

Hardware Platforms

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them

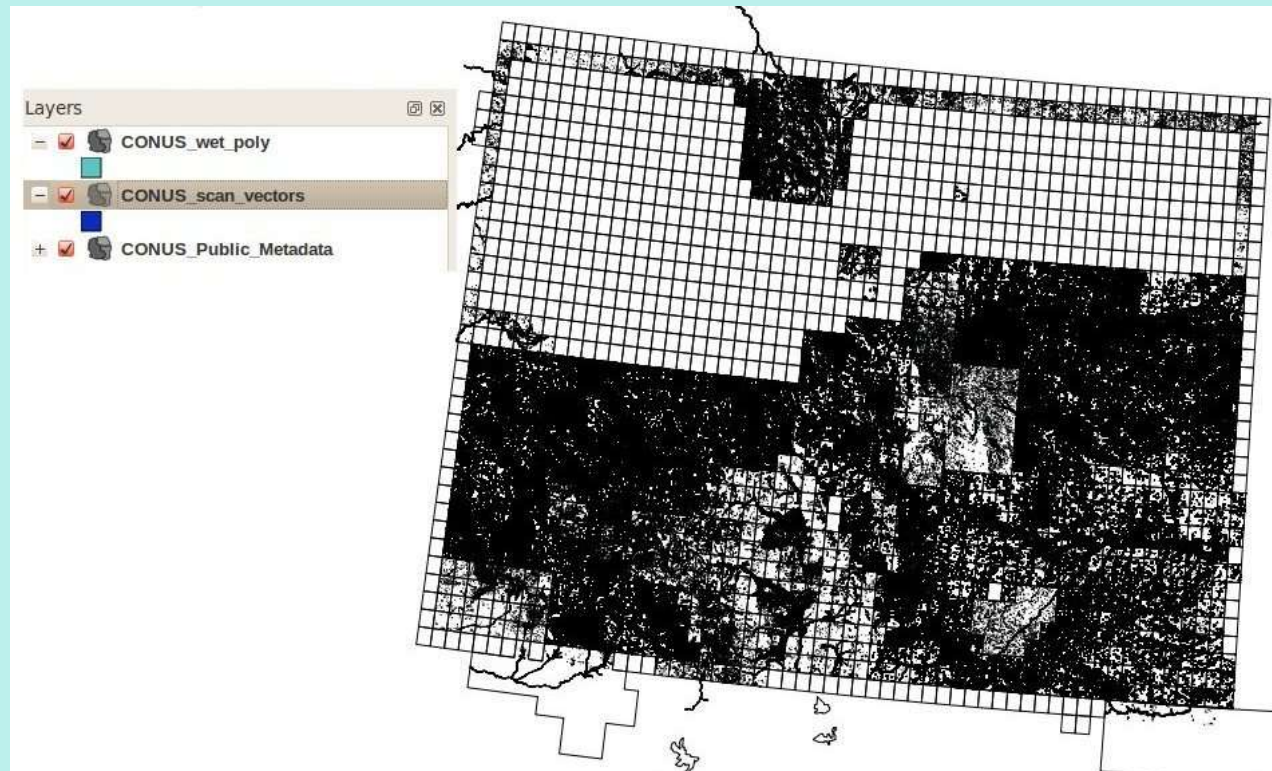


OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations: Extent of Coverage(s)



Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



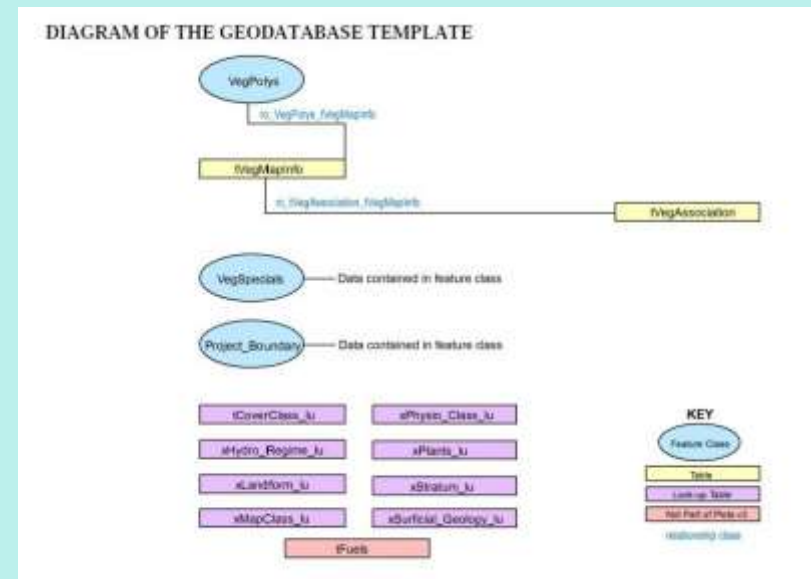
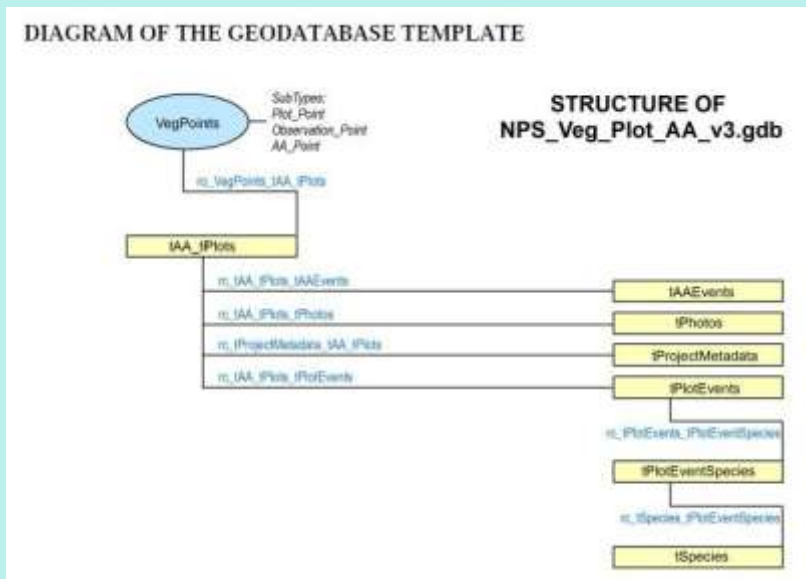
OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations:

Data Table Schemas / Linked Tables Views



Example: NPS Vegetation Plots Data for Rocky Mountain National Park

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations:

Projection Differences

- * A 'true' Regional dataset would have all data layers provided in the same projection; with the 'standard' specified in the metadata
- * If the user is assembling their own 'Regional' dataset from multiple sources, varied data layers may need to be reprojected to a common projection system (as later analysis using on-the-fly reprojection could consume large amounts of compute resources)

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

A number of GIS datasets are being prepared using proprietary software, and consequently are posted for subsequent use in these proprietary formats. An example of this for raster datasets are the USDA – Natural Resource Conservation Services NAIP Imagery datasets, which are downloadable as county mosaics in MrSID format (.sid).

To overcome this barrier: Lizard Techs 'mrsidgeoinfo' and 'mrsidgeodecode' utilities can be used to extract subsets areas of the imagery as GeoTIFFs for subsequent tiling

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

A number of GIS datasets are being prepared using proprietary software, and consequently are posted for subsequent use in these proprietary formats. An example of this for vector datasets are many agencies vector datasets that are created with ArcGIS and subsequently posted for download in either 'Personal Geodatabases (.mdb) or File Geodatabase (.gdb) formats.

To overcome this barrier: The open source Geographic Data Abstraction Library's 'OGR Simple Features Library' (OGR) can translate .mdb file to shapefiles; and .gdb files to shapefiles, (IF the FileGDB API is obtained from ESRI and compiled in to the 'ogrinfo' and 'ogr2ogr' utilities).

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

When many agencies vector datasets that are created with ArcGIS are subsequently posted for download in the 'Personal Geodatabases (.mdb) format, their additional 'linked' tables are stored in Microsoft Access database format.

To overcome this barrier: A suite of open source utilities called 'MDB Tools' is available at: <http://mdbtools.sourceforge.net> 'to document data stored in the MDB format used in Microsoft Access database package, and to provide a set of tools and applications to make that data available on other platforms'.

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



**OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011**




Using Regional Datasets: Considerations

General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

Extracting data from
.MDB Tables and
making the data avail-
able to other formats:



MDB Tools
Unlocking Your Data... <http://mdbtools.sourceforge.net/>

What is MDB Tools?

The MDB Tools project is a effort to document the MDB file format used in [Microsoft's Access database](#) package, and to provide a set of tools and applications to make that data available on other platforms.

Specifically, MDB Tools includes programs to export schema and data to other databases such as [MySQL](#), [Oracle](#), [Sybase](#), [PostgreSQL](#), and others.

Also included is a SQL engine for performing simple SQL queries. The 0.5 release includes an updated GUI interface (screenshot is available [here](#)). A sparse but functional ODBC driver is included as well.

MDB Tools currently has read-only support for Access 97 (Jet 3) and Access 2000/2002 (Jet 4) formats. Write support is currently being worked on and the first cut is expected to be included in the 0.6 release.

For more information check our [FAQ](#) and our [Installation Guide](#). Also see our [SourceForge Page](#) for forums, bugtracking, CVS, et al.

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



**OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011**



Using Regional Datasets: Considerations

General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

The 'MDB Tools' utilities are:

MDB_ARRAY

MDB_EXPORT

MDB_HEADER

MDB_HEXDUMP

MDB_PARSECSV

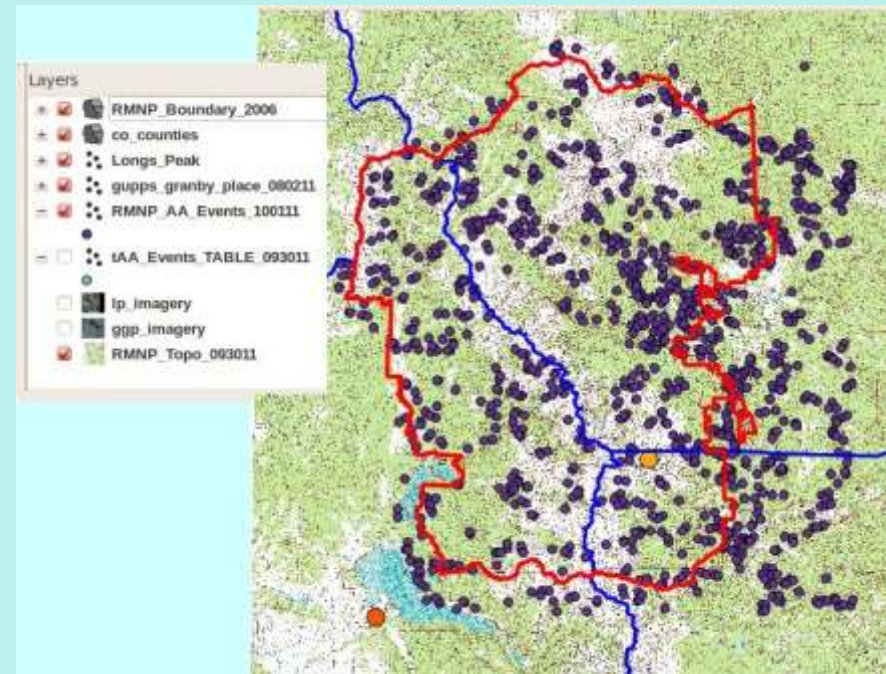
MDB_PROP

MDB_SCHEMA

MDB_SQL

MDB_TABLES

MDB_VER



Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



**OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011**

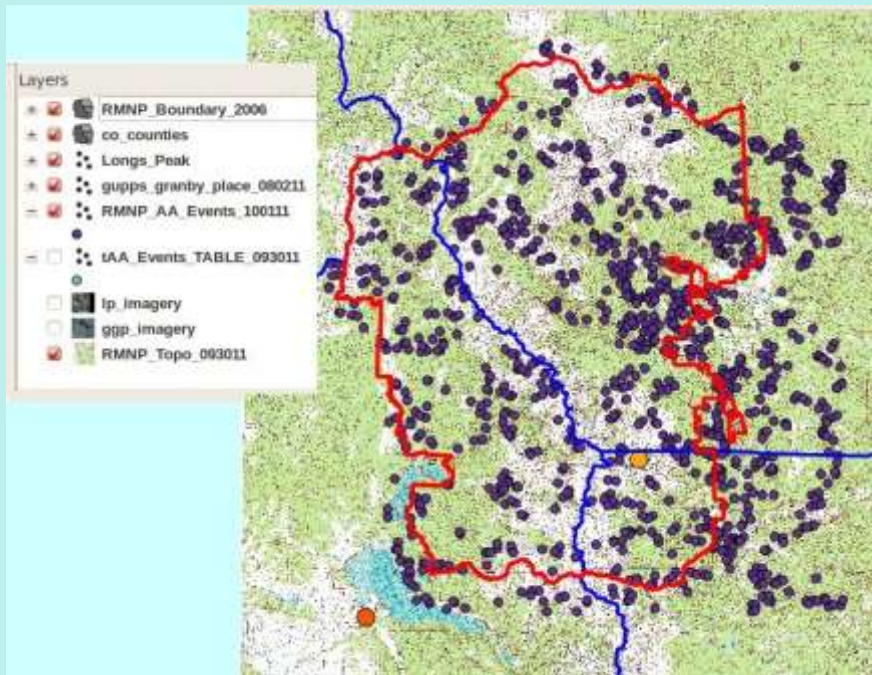


Using Regional Datasets: Considerations

General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS



Using the MDB Tools utilities, of MDB_TABLES, ,MDB_SCHEMA, MDB_EXPORT, followed by MDB_PARSECVS combined with the 'DELIMITED TEXT' Plug-in in Quantum GIS, and its 'Export to Shapefile' function, the Rocky Mountain National Park Vegetation Plots point data was extracted from the original .mdb data file.

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



**OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011**



Using Regional Datasets: Considerations

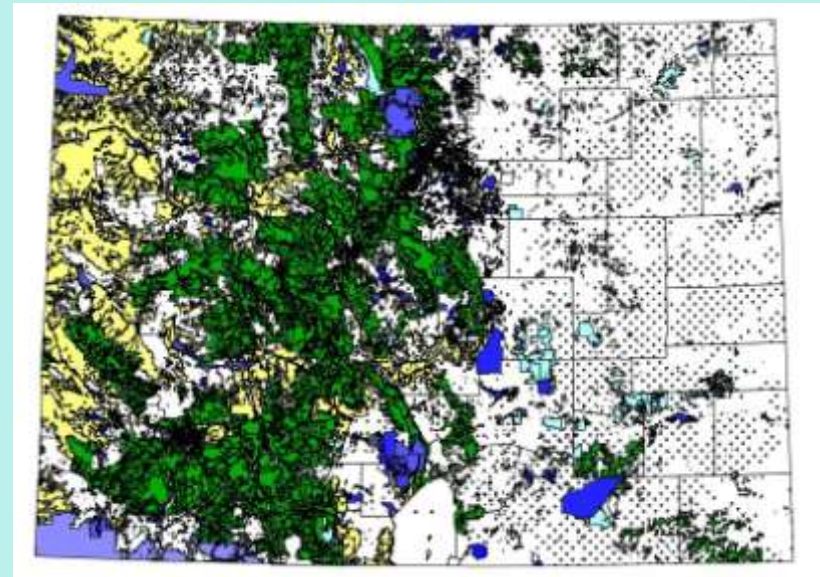
General Considerations: Format(s) – Proprietary or Open

Raster: .sid [MrSID] / .img / .bil

Vector: .shp (& .dbf) / .mdb / .gdb / PostGIS

Using the 'PGeo' driver in the GDAL/OGR utilities 'ogrinfo' and 'ogr2ogr' the Personal Geodatabase (.mdb) for the Colorado Managed and Protected' I((COMaP) lands dataset was converted to shape-file format.

Note: the ogrinfo and ogr2ogr utilities in the Windows version of FWTools worked, whereas the older Linux version would not extract the polygon data.



Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

In Summary:

Compute Resources: Software Options

Proprietary	ArcGIS / ENVI / ERDAS Imagine
Open Source	GRASS GIS / Quantum GIS / uDIG OSSIM / MDB Tools / GDAL-OGR

For efficient data capture and robustness of geo-processing capabilities may data generating agencies use proprietary geographic information system and remote sensing software.

With some effort, open source software can be leveraged to make cost effect use of Regional geospatial datasets.

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Using Regional Datasets: Considerations

General Considerations:

Compute Resources: Hardware Options

Both proprietary and open source geographic information system and remote sensing software can run adequately on modern laptop and desktop computers when moderately intensive compute tasks are performed.

Major geo-processing and analysis of large Regional geo-spatial datasets will benefit greatly from use of more powerful workstation and server class computing hardware.

Massive data storage capacity is a necessity. More is better.

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011



Regional Dataset Conversion Considerations

Options

- * If doing it 'in-house': Invest in the right 'Tools'
- * If not doing the conversions in house, consider:
Supporting Service Bureau efforts
Partnering with: Agencies, Universities, NGOs
Contracting with Commercial Enterprises

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference St. Louis, Missouri
& Business Meeting October 17-20, 2011



General Recommendations

- * Support the Development of Well Documented Regional Datasets, with Robust Metadata
- * Within Reason, 'Keep It Simple'; Data Complexity Can Limit Data Usability by Others
- * Support Open Standards; Avoid Locking Up Data Under Proprietary Constraints
- * Evaluate Partnerships and Contract Options
- * Develop and Contribute Geo-processing Utilities and Applications; and Documented Methodologies

Potential Barriers to the Use of Regional Datasets
and Some Thoughts on Overcoming Them



OFWIM Annual Conference & Business Meeting **St. Louis, Missouri**
October 17-20, 2011

