GPL Document Number 2004-01

GIS Layer and Attribution Naming White Paper

Prepared by the
Oregon GIS Program Leaders (GPL)

June, 2004

_Oregon Geographic Program Leaders
Department of Administrative Services • Department of Agriculture • Department of Forestry • Department of Water Resources
Department of Parks and Recreation • Department of Transportation • Employment Department • Department of Geology & Mineral Industries
Division of State Lands • Oregon Watershed Enhancement Board • Department of Land Conservation & Development
Department of Environmental Quality • Department of Revenue • Oregon University System
et al.

Oregon GIS Program Leaders (GPL)

The Oregon GIS Program Leaders (GPL) group was established to support inter-agency collaboration and cooperation between State, Federal, and local government agencies and private organizations. The GPL serves as technical advisors for the Oregon Geographic Information Council (OGIC) and for state agency GIS development and implementation efforts. The GPL promotes the coordinated development, use, sharing, and dissemination of geographic data.

For more information about the GPL group, please contact:

Cy Smith, Statewide GIS Coordinator
Oregon Geospatial Enterprise Office
1225 Ferry St SE, U120
Salem, Oregon 97301-2556

Telephone: (503) 378-6066
Facsimile: (503) 378-5200
Internet (electronic mail): cy.smith@state.or.us
World Wide Web: http://www.gis.state.or.us

or

Bob Harmon, GPL Chair (2003-2005)
Oregon Water Resources Department
725 Summer Street NE, Suite A
Salem, Oregon 97301-1271

Telephone: (503) 986-0866
Internet (electronic mail): robert.c.harmon@state.or.us
World Wide Web: http://www.wrd.state.or.us

CONTENTS

GIS File Naming Convention

This GIS layer and attribute naming white paper is created to:

1) promote consistency in GIS layer and attribute (variable) naming,
2) provide guidance to data stewards and data contributors related to data communications,
3) provide a clearer understanding of the information in the layers/data set via the layer name. and,
4) facilitate the identification of GIS data sets between data users.

This document will serve as a guideline for groups submitting geospatial data to the Oregon Geospatial Enterprise Office/Geospatial Data Clearinghouse (GEO/OGDC), which are to be posted subsequently to the Spatial Data Library for Internet access and download. These guidelines do not supersede any required and/or highly recommended data components contained in the Federal Geographic Data Committee (FGDC) metadata standard (and, by extension, the Oregon Metadata Standard).

The primary objective is to promote a consistent file naming convention for shared, geo-referenced data sets. This guideline proposes clear filename creation methods in order to minimize confusion, errors, and unnecessary support when GIS data are exchanged between users. Two competing objectives need to be balanced: the need to make the data set name easily recognizable, and the need that the data file name be as short as possible for use in various software systems (and that data variable names not be truncated during data exchange and format changes, which could unintentionally create non-unique variable names).

Background

In its role of providing access to shared GIS data resources via the Internet, the GEO staff has seen the need to develop a standardized guidance document for the naming of GIS data layers and associated data attributes. GEO framed the issue, and then asked for input from the Geographic Program Leaders (GPL) group. GPL agreed that the issue was important, and formed a subcommittee to look at the problem and suggest possible solutions. This document is the result of their work. And while the suggestions described below are good practice generally, these naming conventions should be used for data layers exchanged between data stewards (state agencies, et al.) through the Spatial Data Library at GEO/OGDC. The suggested conventions are not required for internal agency use, but, if used, would promote improved GIS data communication within the organization.

Data sharing technologies are slowly developing, and the coordinated development of GIS data is developing in a parallel manner. The following is a general guidance strategy for naming geospatial data for Oregon data exchange between agencies. It is platform independent. The primary goal of the naming convention is to provide guidance to data stewards for a consistent means of keeping GIS data layer names to a reasonable length (without sacrificing readability and comprehension), but which will allow fast recognition of primary data sources and facilitate on-going data updates (version tracking).

Naming Convention

- Theme names must be descriptive but concise. Coverages, shapefiles and other formats must conform with a 10.3 filenaming structure (that is, CXXXXXXXXX.XXX, where C is a character and X is alphanumeric, for a total of 13 characters and one period separating the filename from the extension). If the three characters that form the filename extension are not recognizable or commonly known, then a decoding instruction must accompany the file for the transfer. For example, it the filename is “Ed12345.zi1” instead of “Ed12345.zip” in order to attach the file to an email, the instruction to substitute a “p” for the “1” must accompany the file. Standard filename extensions are listed at: http://www.computeruser.com/resources/dictionary/filetypes.html

  If abbreviations for GIS themes are necessary, a common set of abbreviations should be used. The committee suggested coming up with a master list of theme name abbreviations starting with the USGS DLG abbreviations. Detailed examples in “Agency/Business Matter” section below and additional information in Appendix C provide common GIS abbreviations and acronyms, as well as links to on-line abbreviations and acronyms.

- All themes names must be 10 characters or less in length, with 3 characters reserved for the extension. This specification is directly related to the export requirements for creating an ESRI shapefile. The length of fieldnames within a GIS data theme should be limited to 10 characters due to the same export considerations.
- We have discussed the possibility of dedicating a data theme's common name (e.g., roads, geology) to the Framework layer, which represents the "best available statewide data" for that theme. That data theme would only be provided to the GEO/OGDC Spatial Data Library by a recognized data steward. In contrast, non-Framework data themes and layers should be identified with the scale for which the data was compiled in the name (i.e., roads100, geology2m).
- If two themes exist at the same scale, then the Framework theme will get the common theme name. For non-framework themes, the second theme will have an abbreviation added (suffix) to it showing the agency of origin (i.e. veg250, veg250odf).
- In cases where the same theme is created on a yearly basis the last two digits of the year will be attached to the theme (i.e. citylim02).

Naming Components

The filenames for GIS themes that are to be distributed through GEO/OGDC are composed of up to four components. These components are added to the filename in the descending level of importance illustrated below. The components include: a brief description of the data set, the scale at which the data were compiled, the agency responsible for the compilation, and the year of its compilation/validity. Not all of these components are necessary in each file name, and the placement is important to recognizing the critical information. In order to utilize all of the field components, the description may need to be limited to three characters (e.g., if a non-steward agency provided a city limits file to GEO/OGDC, the filename might be "lim12kWRD04," which would translate to "2004 city limits at a scale of 1:12000 from the Water Resources Department"). Examples of abbreviations for each naming component are listed in the following paragraphs.

Naming Fields

Field Component	Type	Length
- Description	Text	3-10 (depending on the need for add'l components below)
- Scale	Number/text	0-3
- Agency	Text	0-3
- Year	Number	0-2
File Name	File Name	3-10
Extensions	File Type	3

In addition, several agencies recommend that "internal" GIS data files are stored in major subfolders using the Oregon Framework Themes groupings. A similar approach is used the by USDA – NRCS (see Appendix A for other state/agency approaches to naming GIS files, and Appendix B for information about Oregon state agencies' efforts in this regard). The following is an alphabetical listing of the Oregon Framework Themes (as of June 2004):

- Administrative Boundaries
- Bioscience
- Cadastral (Private Ownership and Public Ownership)
- Climate
- Cultural & Demographic
- Elevation
- Geodetic Control
- Geoscience
- Hazards
- Hydrography
- Land Cover/Land Use

- Ortho-imagery
- Transportation
- Utilities

Additional information on the Framework efforts in Oregon can be obtained at:
http://www.gis.state.or.us/coord/FIT.html

Details for each of the file name components are described below:

Description (Topical Subject Matter)

The primary component of the GIS data theme name should be recognizably related to the topical nature of the data. This schema will assist in classifying or grouping the data into major types of data elements. For statewide framework themes these should be descriptive names and can be up to 10 characters. Examples would include "Geology" for statewide Geology features, "CityLim" for statewide City Limits polygons, etc. This component of the file name cannot be more than 10 characters, and in the case of a framework theme no scale and/or agency would be specified (since these components are assumed to be already defined in Framework). Two characters are available for year, if it is important to distinguish annual versions of the Framework data sets. In that case, the descriptive component of the theme filename must be only 8 characters in length.

It is critical that the descriptive component should be as intuitive as possible, in order to reflect the data contained in GIS layer/coverage. Historically, GIS data filenames were kept short due to limitations within the software used. But with recent advanced in software functionality, layer naming does not have the same length restrictions. However, for data exchange using the ESRI shapefile format and between operating systems/hardware platforms, long file names can be problematic and difficult to type and/or recognize programmatically. Software systems are especially sensitive to the presence of spaces and special characters in the filenames for GIS data sets.

Standard, recognizable abbreviations exist for GIS data and application extensions. Where these standard abbreviations are available, the name should conform to the naming standards of the agency creating it. Appendix C provides additional information on common GIS abbreviations and acronyms. In addition, the USGS Digital Line Graph (DLG) abbreviations should be used wherever possible. FGDC-compliant metadata will contain additional information concerning the full description of the GIS data, how it was derived, and who to contact for questions.

Scale

This component of the GIS filename contains the cartographic resolution (or generalized cartographic accuracy) of the data layer, stated in standard ratio scale units. The scale is represented in thousands in order to conserve space in the filename. For example, 250 would be the filename scale component for 1:250000 data sets, and 24 would be the filename scale for 1:24000 data. Data in the millions should have an ***M*** added to the name (e.g. 2M would represent 1:2000000). The data resolution/accuracy should only be part of the name if all features in the layer are of similar (or better) resolution/accuracy. In other words, the data scale must be stated as the "worst" of the compiled data components. Base level data from Framework does not need to include the scale in the file name. And again, the FGDC-compliant metadata will contained additional information regarding scale.

Scale	
24	1:24,000
100	1:100,000
250	1:250,000
500	1:500,000
2M	1:2,000,000

Note that this is not a *scale* designator related to absolute accuracy of the spatial location. The scale designation represents a cartographic (or printout display) scale. It is an "accuracy/resolution indicator" that generally describes the data content. The resolution or accuracy, as well as intended scale (if known), should all be described more fully in the metadata of each GIS data layer.

Agency/Business Matter

An agency or organization name has the potential to be incorporated into each GIS layer filename. This is only required when multiple sources need to be differentiated. Three-letter or four-letter abbreviations should be used in order to maintain the file length specification below the 10-character limit. Listed below are abbreviations for agencies whose GIS data is currently hosted at GEO/OGDC in the Spatial Data Library:

Federal agencies
ACE US Army Corps of Engineers
BLM US Bureau of Land Management
EPA US Environmental Protection Agency
USC US Bureau of the Census
USFS US Forest Service
USGS US Geological Survey
USCS US Soil Conservation Service

State agencies
DEQ – Oregon Department of Environmental Quality
DHS – Oregon Department of Human Services
DOT – Oregon Department of Transportation
GEO – Oregon Geospatial Enterprise Office (formerly SSCGIS)
GMI – Oregon Department of Geology and Mineral Industries
IOE – Inside Oregon Enterprises (Dept of Corrections)
LCD – Oregon Department of Land Conservation and Development
NHP – Oregon Natural Heritage Program
ODF – Oregon Department of Forestry
ODA – Oregon Department of Agriculture
OED – Oregon Employment Department
PRD – Oregon Department of Parks and Recreation
WEB – Oregon Watershed Enhancement Board
WRD – Oregon Water Resources Department

Unique county abbreviations for Oregon are listed in Appendix C.

Date Information

When needed to uniquely identify or help clarify information about a given GIS layer, a date indicator should be added to the end of the filename. Two digit years will be used (i.e., 1995 would be presented as "95", and 2002 as "02"). The descriptive name of a layer of City limits in 1996 might be Citylim96. FGDC-compliant metadata will contain additional information regarding date of the data represented.

Additional Data/Information

With advances in GIS and data storage software, many components of GIS data that were historically incorporated into the filenames is no longer necessary. The topology of the data (point, line, polygon, route, grid, image, etc) no longer needs to be incorporated into the filename, but this information should be documented thoroughly in the metadata.

Separators

The file name must always be followed by a "." (period) and then the appropriate extension representing the file type(s). For example, an ESRI shapefile is comprised of three filetypes, and will have a .shp, .shx and .dbf extension to the filename (at a minimum). ESRI Arc/Info exchange files will have an ".e00" extension.

Underscores can also be used as a separator ("_") in the descriptive portion of the file name (as a substitute for a space), but this needs to be used sparingly – due in part to the 13-character limit to the total filename length.

Example names

The following example names follow this standard:

Name	Description	Areal Extent	Scale	Agency	Year	Field Length
Roads100	Roads	Statewide	1:100000	N/A - see metadata	N/A - see metadata	8
Geology2M	Geology	Statewide	1:2000000	N/A - see metadata	N/A - see metadata	9
City02	City Limits	Statewide	Framework	N/A - see metadata	2002 Annual Update	6
veg250	Vegetation	Statewide	1:250000	N/A - see metadata	N/A - see metadata	6
veg250odf	Vegetation	Statewide	1:250000	Oregon Dept of Forestry	N/A - see metadata	9
Emp250ED02	Employment	Statewide	1:250000	Oregon Employment Dept	2002 Annual Update	10
AddrYamh01	Address Points	Yamhill County	Framework	N/A - see metadata	2001 Annual Update	10

(Please note the non-standard agency abbreviation in the Employment Dept GIS file listed above. "ED" was used instead of "OED" to keep the filename no longer than ten characters.)

Metadata

Detailed FGDC-complainant Metadata is expected to be associated with all datasets and will provide additional information and detail on data files.

Standard Acronyms and Abbreviations

Standard acronyms should be used in the layer names. Standard abbreviations should also be used within layer names where appropriate (See Appendix C for recommendations).

Each data layer from an external source will be given at least a three-letter prefix that represents the source entity according to the following rules:

1. All prefixes shall have three alphanumeric characters
2. All source-entity-assigned layer names that exceed 10 characters shall be truncated or abbreviated to satisfy the 13-character limit for filenames.

File Name Tracking

The subcommittee discussed the creation of a master table (database) containing the file name and critical information relating to the data sets, which could be maintained and hosted by the OGDC/GEO. This lookup table would be available for building library documentation and minimizing replicated work between agencies. It could contain the following:

- File name
- Brief description of the data
- Scale
- Data Source/Steward

This could be posted at GEO and be used on local data repositories. At this time, however, that commitment has not been finalized.

Appendices

Appendix A – Guidance from other states/Agencies

Several other organizations have established GIS naming conventions. This data serves as a resource for developing this guideline:

Minnesota DNR has a guideline for images at:
http://files.dnr.state.mn.us/aboutdnr/bureaus/mis/gis/gis_spec/image_storage_guidelines.pdf

Vermont has standards at:
http://www.vcgi.org/techres/standards/partiii_section_e.pdf

Appendix B – Guidance from Oregon Agencies

Oregon Department of Forestry has some file naming information in their operations manual:

http://www.odf.state.or.us/divisions/administrative_services/services/gis/pdf/gisops72600.pdf

Appendix C – Common Abbreviations

Several on-line acronym/abbreviation libraries exist.

http://www.lib.berkeley.edu/EART/abbrev.html - national-online GIS abbreviation library at UC-Berkley

http://www.fema.gov/pdf/fhm/frm_gsay.pdf - FEMA/DFIRM official acronyms/abbreviations

http://www.geo.ed.ac.uk/agidexe/acronyms - GIS diction/list of acronyms at the Association for Geographic Information (Great Britain)

Unique County Abbreviations for Oregon (4 characters)

County	Abbreviation
Baker	bake
Benton	bent
Clackamas	clac
Clatsop	clat
Columbia	colu
Coos	coos
Crook	croo
Curry	curr
Deschutes	desc
Douglas	doug
Gilliam	gill
Grant	gran
Harney	harn
Hood River	hood
Jackson	jack
Jefferson	jeff
Josephine	jose
Klamath	klam
Lake	lake
Lane	lane
Lincoln	linc
Linn	linn
Malheur	malh
Marion	mari
Morrow	morr
Multnomah	mult
Polk	polk
Sherman	sher
Tillamook	till
Umatilla	umat
Union	unio
Wallowa	wall
Wasco	wasc
Washington	wash
Wheeler	whee
Yamhill	yamh