



**US Army Corps
of Engineers**
Construction Engineering
Research Laboratories

USACERL Technical Report 98/66
April 1998

GRASS to ARC/INFO Data Conversion

by

Kelly Dilks

Edward Delisio

Mark Gibb

Chris Rewerts

Eric Bader

Gail Taylor

This guide is targeted towards individuals that use geographic information system (GIS) technology. It describes several techniques for translating data from Geographic Resources Analysis Support System (GRASS) format data files to the ARC/INFO data format used with Environmental Systems Research Institute, INC (ESRI) products. The primary conversion routines discussed here are a result of a collaborative effort between the U.S. Army Construction Engineering Research Laboratories (USACERL) and ESRI. GRASS is a public domain geographic information system originally developed by USACERL. This report is intended to be used as a reference during the data conversion process; it describes the two data formats and contains tips that may facilitate the conversion process.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE April 1998	3. REPORT TYPE AND DATES COVERED Final	
4. TITLE AND SUBTITLE GRASS to ARC/INFO Data Conversion			5. FUNDING NUMBERS 4A162720 A896	
6. AUTHOR(S) Kelly Dilks, Edward Delisio, Mark Gibb, Chris Rewerts, Eric Bader, and Gail Taylor				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Construction Engineering Research Laboratories (USACERL) P.O. Box 9005 Champaign, IL 61826-9005			8. PERFORMING ORGANIZATION REPORT NUMBER TR 98/66	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) ODEP ATTN: DAIM-ED-N 600 Army Pentagon Washington, DC 20310-0600			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES Copies are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This guide is targeted towards individuals that use geographic information system (GIS) technology. It describes several techniques for translating data from Geographic Resources Analysis Support System (GRASS) format data files to the ARC/INFO data format used with Environmental Systems Research Institute, INC (ESRI) products. The primary conversion routines discussed here are a result of a collaborative effort between the U.S. Army Construction Engineering Research Laboratories (USACERL) and ESRI. GRASS is a public domain geographic information system originally developed by USACERL. This report is intended to be used as a reference during the data conversion process; it describes the two data formats and contains tips that may facilitate the conversion process.				
14. SUBJECT TERMS GRASS ARC/INFO Geographic Information Systems			15. NUMBER OF PAGES 42	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR	

Foreword

This study was conducted for the Installation Spatial Technology Advisory Board (ISTAB) and ODEP Under Project 4A162720A896, "Environmental Quality Technology," work Unit T38, "Land Management Technology Integration." The Technical Monitor was Vic Diersing DAIM-ED-N.

The work was performed by the Resource Mitigation and Protection Division (LL-R) of the Land Management Laboratory (LL), U.S. Army Construction Engineering Research Laboratories (USACERL) and Implementation Services, Environmental Systems Research Institute, Inc. (ESRI). The USACERL principal investigator was Kelly M. Dilks. Eric Bader works for ESRI, the remainder of the authors work(ed) for USACERL during the project. Robert E. Riggins is Chief, CECER-LL-R; William D. Severinghaus is Operations Chief, CECER-LL; and William D. Goran is Technical Director, Conservation, CECER-TD. The USACERL technical editor was Gloria J. Wienke, Technical Resources.

COL James A. Walter is Commander and Dr. Michael J. O'Connor is Director of USACERL.

Contents

SF 298	1
Foreword	2
1 Introduction	5
Background	5
Objective	6
Approach	6
Scope	6
Mode of Technology Transfer	7
2 GRASS Data Structure	8
GRASS Raster Data	8
GRASS Vector Data	9
GRASS Site Data	9
3 ARC/INFO Data Structure	11
Coverages	11
Grids	12
Tables	12
Images	12
Tins	13
Coverage Features	13
Coverage Extent (BND)	16
Grid Cells	16
Storing Attributes in the INFO Database	17
Item Definitions	17
Supported Item Types	18
The INFO Database	18
4 Automated Conversion Process	20
Steps for Automated Conversion	20
Things to Keep in Mind During the Process	22

5	Quality Assurance and Quality Control Procedures	24
	Steps for Quality Control Check	24
	Reviewing the Output of the Check.....	25
6	Summary	28
	References	29
	Appendix A	30
	Appendix B	31
	Appendix C	32
	Polygon data conversion using the GRASS command v.out.arc	32
	Line data conversion using the GRASS command v.out.arc	35
	Polygon data conversion using the GRASS command v.out.dlg	37

Distribution

1 Introduction

Background

The Geographic Resources Analysis Support System (GRASS) is a public domain image processing and geographic information system (GIS) that was initially developed by the U.S. Army Construction Engineering Research Laboratories (USACERL) during the late 1980's and early 1990's. The public domain GRASS computer software was initially developed to assist military installation environmental resource managers manage military lands when commercial software with the necessary capabilities was not available, had limited functionality, or was too expensive to implement across many military sites. Public domain GRASS has been used by both the public and private sector to assess environmental impacts, evaluate site suitability, detect change over time, manage resources, and model the effects of environmental phenomena across the landscape.

As the GIS technology matured in the commercial sector, and surpassed the capabilities available in public domain GRASS, it became apparent that USACERL should not spend additional federal resources competing with commercially available GIS software. USACERL, in conjunction with the Installation Spatial Technology Advisory Board (ISTAB), decided to halt public domain GRASS software development. The last full USACERL release of public domain GRASS was in the Spring of 1993. Many organizations have taken the public domain GRASS software code and enhanced its functionality beyond what was possible with federal dollars. Baylor University, Blackland Research Institute, GPZ Technologies, Logiciels et Applications Scientifiques, (L.A.S.) Inc., and the National Oceanic and Atmospheric (NOAA) Geophysical Data Center have all released various public domain and commercial versions of GRASS on a variety of computer platforms.

At the same time public domain GRASS development ceased at USACERL, efforts were made to ensure the established GRASS user community would have alternate avenues in place to fulfill their GIS requirements. Considerable effort was made to ensure the millions of dollars spent on the creation of GRASS data would not be wasted. Documented and tested routines would be developed to assist former GRASS GIS users in converting their data to some of the more popular commercially available GIS packages.

USACERL signed a Cooperative Research and Development Agreement (CRADA) with Environmental Systems Research Institute, Inc. (ESRI), the developers of the ARC/INFO GIS and ArcView product suites. This agreement

allows USACERL and ESRI to work collectively to assist in the transition of public domain GRASS to commercially available GIS products developed by ESRI. This document describes the results of the partnership between ESRI and USACERL. Prior to the development of these data conversion routines, several work-around programs existed – but these were not well documented, or were only partial solutions that did not convert all the data, often requiring significant editing and post-processing.

This document describes in detail the process to convert GRASS data files to an ARC/INFO data format for use in various ESRI-developed GIS software products. It contains sections on the data architecture of both GRASS and ARC/INFO, and tips and hints that may be used during the data conversion process. The data conversion ARC Macro Language (AML) scripts and associated UNIX scripts that accompany this document were tested on GRASS version 4.1, updates 2 and 3, and ARC/INFO versions 7.03, 7.1.1 and 7.1.2. ESRI is currently working on a GRASS data reader for ArcView for those users who do not have GRASS and ARC/INFO available to them on a UNIX machine. Information on ARC/INFO is provided courtesy of ESRI, Redlands, CA.

Those interested in the conversion of GRASS data to Intergraph formatted data files should refer to the USACERL technical report *GRASS-Intergraph Data Conversion Guide*, USACERL ADP Report N-92/01, April 1992.

Objective

The objective of this report is to document procedures and issues relating to the conversion of GRASS formatted data to the ARC/INFO data format of ESRI software products, and provide tips to facilitate the conversion process.

Approach

The primary data conversion technique identified in this report is a result of collective research between USACERL and ESRI technicians who developed these ARC Macro Language (AML) programs, and UNIX scripts. Other methods exist for the conversion of GRASS data to ARC/INFO formatted data. Some of these manual and semiautomated routines are in the appendices.

Scope

This report documents the conversion of GRASS Version 4.1, updates 2 and 3 to ARC/INFO version 7.03, 7.1.1 and 7.1.2. The quality control scripts were tested using ArcView, version 3.0b. The AML and associated scripts require both GRASS and ARC/INFO installed and executable from a UNIX machine. This is necessary because the AML and associate scripts make calls to the GRASS,

ARC/INFO and UNIX libraries. If you do not have simultaneous access to both packages, Appendix C describes how a manual method can be used to convert the data.

Mode of Technology Transfer

This report will be distributed in a limited number of hard copies with an accompanying CD-ROM containing related electronic files. The report will also be posted to USACERL's web page at: <http://www.cecer.army.mil>.

The information in this report will assist the GIS user in the technical transfer of USACERL's GRASS-formatted data files to a data format in use by a wide sector of the GIS community using commercial software available from ESRI.

2 GRASS Data Structure

The GRASS database makes use of the UNIX hierarchical directory structure. The setup and arrangement of GRASS data is similar to a filing cabinet with several drawers. The entire filing cabinet contains data about a particular geographic area. In GRASS this area is called the LOCATION, and covers a specific geographic area of the Earth, and contains map layers in a variety of themes. Like the filing cabinet that has several drawers, a LOCATION can contain several or N number of MAPSETS. MAPSETS are subdirectories under a LOCATION. Each MAPSET has a unique owner associated with it, controlled by UNIX permission settings. Each owner or GRASS user can view and use data in any MAPSET, but can only alter data owned by the current user in the current MAPSET. MAPSETS contain files and subdirectories that are associated with the data for the particular geographic area of interest. GRASS MAPSETS contain all data types – raster, vector, site, and image files. GRASS file names do not adhere to the DOS 8.3 file naming standard (ISO 9660). Filenames can be of any length, and do not require a 3-character extension separated by a dot (.).

GRASS Raster Data

Raster data types are used to depict information that has an aerial extent, such as a soils map containing soil type, elevation, satellite imagery, or land use. GRASS raster data are stored as a matrix of grid cells. Each grid cell covers a known, square area that is typically geographically referenced to the Earth. The entire area of the GRASS LOCATION is made up of a grid of rows and columns of raster cells, though data may not always exist for the entire area. The size of the grid cell, or pixel, is user specified, depending on the scale and resolution of the data. A low-resolution map might use a cell with sides representing 100 meters on the ground. Each raster cell is assigned a single attribute value called the category number. The following is a list of the directories that may be found in a GRASS MAPSET that relate to GRASS raster data:

Directory Name	Contents
cell	raster data files
cellhd	header files for raster data files
cats	category information for raster data files

colr	color tables for raster data files
colr2	secondary color tables for raster data files
cell_misc	histogram range for raster data files
hist	history information for raster data files

GRASS Vector Data

Vector data types can be used to depict linear features such as roads, transmission lines, boundaries, or stream networks. GRASS vector data, known generically as lines, can represent linear features or area edges. Area edges are lines that form closed polygons, or areas, that can be further processed into raster polygons. Lines are linear features that do not close and create areas. All vector data is stored as a series of x,y geographic coordinate pairs. Each vector map feature (either line or area edge) is assigned a single integer attribute value called the category number. In previous vector conversion routines, it was often the category information that was lost during the data conversion process. There are multiple files associated with a single vector theme, each being stored in different directories associated with vector data. The following is a list of the directories that may be found in a GRASS MAPSET that relate to GRASS vector data:

Directory Name	Contents
dig	binary vector data files
dig_ascii	ASCII vector data files
dig_att	vector category attribute file
dig_plus	vector topology file
dig_cats	optional vector category file
reg	digitizer point registration

GRASS Site Data

GRASS site data depict single point locations that have neither area nor length. Helicopter pads, firing points, nesting sites, or single tree locations can be examples of GRASS site data, depending on the scale of the data. This data type is stored as the single coordinate pair describing the location of the feature, and can be followed by a category value and a category label for the feature. GRASS

site data is easily exported for use with other programs. GRASS site files are located in a directory called `site_lists` in the user's GRASS MAPSET.

3 ARC/INFO Data Structure

The source for the information contained in this section on ARC/INFO Data Structure was taken from ArcDoc, Version 7.1.2, the online help documentation contained in ARC/INFO Version 7.1.2.

ARC/INFO GIS uses a hierarchical directory structure. Data sets are organized into ARC/INFO workspaces, which are directories that contain one or more geographic data sets, a local INFO database, and other supporting data.

Coverages

Coverages represent the fundamental data source for ARC/INFO. A coverage contains a set of features, each represented by a feature class such as arc, node, label point, annotation or polygon. The coverage supports the georelational model – it contains both the spatial (location) and attribute (descriptive) data for geographic features. The georelational model is the fundamental data model used in ARC/INFO.

A coverage is stored as a directory containing a set of files, with corresponding data in an INFO database identified as the INFO directory. The directory name is usually the coverage name. INFO is the relational database product integral to ARC/INFO.

The combination of feature classes present in a coverage depends on the geographic phenomena being represented. Each feature class stores attribute information in a corresponding feature attribute table. These tables reside in the workspace INFO database where the coverage resides.

Three topological concepts are used to define features: arc-node, left-right and area definition. Arc-node topology defines the connectivity of arcs; arcs connect at nodes. Polygons are defined using left-right and area definition topologies. A polygon is defined as an ordered set of connected arcs, with the constraint that the first and last arcs must connect (area-definition). For each arc, the left and right polygons are identified (left-right). Arcs also have a direction (to-from) that is defined by the location of the beginning and ending node of an arc.

Grids

Grids are ARC/INFO's raster data structure used to represent categorical data, and the GRID program extension is used for raster analysis of grid data sets. Each grid represents a spatial variable. Both raster images and maps can be stored in GRID.

Grid-based systems divide the world into discrete uniform units called cells. Every cell represents a certain specified area of the earth and is given a value to correspond to the feature or characteristic that is located at or describes the site. Location is not defined as an attribute but is inherent in the storage structure.

Grid systems treat points, lines, polygons and surfaces, and their locational structures, the same way; as cells in a grid. Analysis and computation using grids are generally very fast. Once registered, computing or deriving a value for an output cell from two or more input cells is a matter of direct value computation. No geometric detection, topology building, and error checking is necessary.

The uniform cells are organized into a Cartesian matrix consisting of rows and columns. A row identifies all cells equidistant from the top or bottom boundary of a grid. Columns identify all cells equidistant from the left or right boundary of the grid. Each Cartesian matrix is called a grid. Every cell in a grid has a unique row and column identifier.

A grid is similar to an ARC/INFO coverage. A grid is stored in an ARC/INFO workspace. The grid, like a coverage, is stored as a separate directory with associated tables and files that contain specific information about the grid.

Tables

Descriptive attributes of geographic features are stored in rows of a table. Each attribute is stored in a field or item, with one record (or row) of attributes for each feature. In this way, feature attribute tables can be related or linked to geographic features. The columns contain values for particular attributes, such as area, soil-type, drawing symbol. ARC/INFO manages three kinds of attribute tables: feature attribute tables, INFO data files, and external attribute tables from a relational database management system (RDBMS) such as ORACLE.

Images

Images store photographs in rows and columns as a set of cells called pixels. Images represent two types of information: map images and descriptive images. Map images can be aerial photos or satellite imagery. Picture images are items such as photos and scanned documents.

Tins

TIN, or triangulated irregular network, is the data structure used to represent surfaces and the data model for the TIN software extensions to ARC/INFO. Tins are useful for representing surfaces that are highly variable, and contain discontinuities and breaklines. The main components of a tin are triangles, nodes, and edges. Nodes are locations defined by x, y, and z values (xyz) from which a tin is constructed. Triangles are formed by connecting each node with its neighbors. Edges are the sides of the triangles. The exact structure of a tin is based on certain triangulation rules that control tin creation.

Coverage Features

The following are just some of the ARC/INFO coverage features and are the features that are used when converting data between the ARC/INFO and GRASS software products.

Points

Points represent geographic features that have no area or length, or features that are too small for their boundaries to be apparent for the given input map scale. A single x, y coordinate and an internal sequence number describe each point. In ARC/INFO, points are stored in a LAB file.

A point attribute table (PAT) is used to hold the attributes for the points. There is one record in the PAT for each point. The record is related to the point by the sequence number. At a minimum the PAT contains four items:

AREA	Holds the area of a polygon. The value is 0 for points.
PERIMETER	Holds the perimeter of a polygon. The value is 0 for points.
<cover>	(where <cover> is the name of the ARC/INFO coverage) Internal sequence number (i.e., the record number) of the point feature in the LAB file.
<cover>-ID	User-assigned feature ID for each point

Arcs

Arcs represent both linear features and the borders of areal features. Linear features represented by arcs can have length, but no area (e.g., elevation contours), or can be long narrow features whose width is not apparent at a given map scale (e.g., streets). Each linear feature may be made up of many arcs. Nodes indicate the endpoints and intersections of arcs. In addition, nodes can represent point features, which connect segments of a linear feature.

Arcs are stored in two coverage files: ARC and AAT. The ARC file contains one record for each arc. Each record contains the arc's user-id, location and shape information defined as a series of x,y coordinates, the from-node and to-node, and the left and right polygon numbers. Descriptive data about arcs are stored in an arc attribute table (AAT). There is one record in the AAT for each arc in the coverage. The record is related to the feature by the internal sequence number stored for each arc. At a minimum, the following items are contained in the AAT:

FNODE#	Internal sequence number of the from-node
TNODE#	Internal sequence number of the to-node
LPOLY#	Internal sequence number of the left polygon; set to 0 if the coverage does not contain polygons.
RPOLY#	Internal sequence number of the right polygon; set to 0 if the coverage does not contain polygons.
LENGTH	Length in coverage units.
<cover>#	Internal sequence number (i.e., the record number) of the arc in the ARC file.
<cover>-ID	User-assigned feature ID.

The from-node number (FNODE#) and the to-node number (TNODE#) identify which areas are connected (share a common node). The left polygon number (LPOLY#) and the right polygon number (RPOLY#) identify which polygons are contiguous (share a common arc).

Polygons

Polygons are used to represent area features. A polygon is defined by a series of arcs comprising its border and by a label point positioned inside the polygon. The user-id is assigned to the label point.

Polygons are stored topologically using left-right topology stored in the ARC file. The polygon arc list (PAL) file contains a list of all arcs and nodes defining each polygon's boundary. There is one record in the PAL for each polygon in the coverage. The CNT (centroid) file stores the label point numbers for each polygon. Label point coordinates are stored in the LAB file. Polygons require at least one label point in order to associate attributes. Descriptive data about polygons is stored in a polygon attribute table (PAT). There is one record in the PAT for each polygon, which is related to the polygon using the polygon's internal sequence number. At a minimum, the PAT contains four items:

AREA	Holds the area of a polygon.
PERIMETER	Holds the perimeter of a polygon.
<cover>#	Internal sequence number of a polygon.
<cover>-ID	User-assigned feature ID for each polygon.

Nodes

Node coordinates are not stored explicitly within a coverage. Instead, node locations are stored as a part of each arc – as the arc’s beginning and ending vertices. Internal numbers of nodes are automatically assigned and stored as part of the arc information in the ARC file. When an AAT is built for a coverage, it contains the FNODE# and TNODE# items.

When nodes are used to represent point features, descriptive data is stored in a node attribute table (NAT). There is one record in the NAT for each node. The record is related to the node by the node’s internal sequence number. At a minimum, the following items are contained in a NAT:

ARC#	Internal sequence number of one of the arcs that connects at the node location. If more than one arc shares the node, the arc with the lowest internal number is used. This allows the x,y coordinate for the node to be read from the arc’s record in the ARC file.
<cover>#	Internal sequence number of the node.
<cover>-ID	User-assigned feature ID. When an NAT is initially created, node IDs are automatically set equal to the node’s internal sequence number.

Tics

A tic is a registration or geographic control point for a coverage. TICs allow coverage coordinates to be registered to a common coordinate system. All tic information for a coverage is stored in the TIC file. It contains the following items:

IDTIC	The user-id for each tic
XTIC	The tic’s x-coordinate
YTIC	The tic’s y-coordinate

Coverage Extent (BND)

Coverage extent represents the outer boundary of a coverage. It is the minimum bounding rectangle that defines the coordinate limits (extreme minimum and maximum coordinates) of coverage arcs and label units, and by definition, polygons, route-systems, and regions.

The BND is typically used to set a map extent for coverage drawing and display operations. It is often used as a default map extent for quick coverage display. Many spatial processes use the BND to determine whether one coverage overlaps another and to sort coverage features by location for processing.

All extent information for a coverage is stored in the BND file, which contains the following items:

XMIN	The x-coordinate of the coverage extent's lower-left corner
YMIN	The y-coordinate of the coverage extent's lower-left corner
XMAX	The x-coordinate of the coverage extent's upper-right corner
YMAX	The y-coordinate of the coverage extent's upper-right corner

A coverage containing no arcs or label points (or a single label point will have an undefined BND).

Grid Cells

A grid cell is a discrete, uniform cell unit. Every cell represents a specific area on the earth. Each cell is given a value to correspond to the feature or characteristic that is located at or describes the site. An integer value is normally associated with each grid, which defines the group, class, or category the cell belongs to.

All integer grids include an INFO table called the value attribute table (VAT). The VAT always contains at least two items: VALUE and COUNT. VALUE is the value assigned to the cells in the grid, and COUNT is the number of cells in the grid that are assigned that value. Any number of additional items representing other attributes of the group, class, or category can be added or related to the VAT. There are no feature-IDs in a VAT.

Storing Attributes in the INFO Database

An ARC/INFO coverage can contain both the spatial and attribute information. Feature class attributes are stored in feature tables (i.e., AAT, NAT, PAT, and TIC).

Feature attribute tables are generated by ARC/INFO when you create a feature class topology. For each feature in the coverage feature class there is one record in the feature attribute table. The attribute tables contain a mandatory set of attribute items required by ARC/INFO.

In INFO, a column in an attribute table is called an item. These mandatory items vary between feature classes. You can then add additional items to the feature attribute table to hold the information you require to record about features in your database.

There is always one record in the feature attribute table corresponding to each feature in the coverage. Both the spatial information used to define the coverage feature and the corresponding record in the feature attribute table contain the feature number so that a one-to-one correspondence is maintained between the feature and its record in the feature attribute table. Even though the records in a feature attribute table maintain a one-to-one correspondence with coverage features, one-to-many and many-to-many relationships can be managed between the feature attribute table and corresponding tables.

Item Definitions

The specification of the format for each record in the data file is referred to as the item definition. Each record can be up to 4,096 characters (bytes) long. Any number of items can be defined for the data file. Items are defined by their name, the data type, the number of characters (or bytes) used to store values, a display width, and (for real numbers) the number of decimals you wish to display. INFO uses the following conventions to define the format of each item in a data file:

Item Name	Any name with up to 16 alphanumeric characters.
Item Width	Number of spaces (or bytes) used to store item values.
Output Width	Number of spaced used to display the item values.
Item Type	The data type of the item.
No. of decimals	The number of digits to the right of the decimal place for item types that hold decimal numbers.

Supported Item Types

The following INFO file item types are supported:

- | | |
|---|---|
| B | Whole numbers stored as binary integers (width of 2 or 4 bytes only). The maximum value for width of 2 is 32,767: for width of 4 is 2,147,483,647. |
| C | Character (width up to 320 alphanumeric characters) |
| D | Dates in the form DD/MM/YY or DD/MM/YYYY (item width is fixed at 8 and stored internally as YYYYMMDD). |
| F | Decimal numbers stored in internal floating-point representation (width of 4 or 8 bytes only). A 4-byte width is single-precision real (7 digits of precision), and 8 bytes is double precision (14 digits of precision). |
| I | Integers stored as 1 byte per digit (width from 1 to 16, maximum value possible is 2,147,483,647). |
| N | Decimal numbers stored as 1 byte per digit (width from 1 to 16). |

The INFO Database

Both feature attribute tables and related INFO files are stored in INFO databases. The INFO database is a file-based system. Each ARC/INFO workspace contains an INFO database directory; thus a multi-workspace ARC/INFO database contains many INFO databases.

Feature attribute tables must be stored in an INFO database located in the coverage workspace. In ARC/INFO, a workspace is a directory that contains a set of coverages and their INFO subdirectory. The INFO subdirectory contains all of the feature attribute tables for those coverages plus any other associated INFO files.

One important fact in the use of INFO is the user's view of INFO; sets of INFO files are managed as a unit in an INFO database. An INFO directory is a collection of INFO files stored in a single-user workspace. These files are only accessible from the ARC/INFO modules and the INFO database program.

Any selected set of INFO file records or related records will, by default, be returned in the order that the records were inserted into the INFO file. If you

wish to change this order, then you have to sort the INFO file. INFO files can be sorted on one or more item values so that records are returned in a predictable order. **Since feature attribute tables are always ordered by the cover#, sorting these files by cover# will corrupt your coverage.**

4 Automated Conversion Process

To facilitate the transition from GRASS GIS to ARC/INFO GIS, USACERL teamed with ESRI professionals to develop the AML and associated scripts found in Appendixes A and B on the accompanying CD-ROM.

The AML and associated scripts require both GRASS and ARC/INFO installed and executable from a single UNIX machine. This is necessary because the AML and associated scripts make calls to GRASS, ARC/INFO, and UNIX libraries. If the required configuration is not available at your site, you may go through the procedures in Appendix C, or contract the work out to someone who has the necessary tools available.

Steps for Automated Conversion

1. On a UNIX machine, change to the GRASS MAPSET directory containing the data you wish to convert.
2. Start ARC/INFO and create a workspace called `arc`, then quit ARC/INFO.
3. At the machine prompt type “`setenv GISDBASE (your database directory)`”. Your GISDBASE is the directory that you type when entering GRASS on the DATABASE line.
4. Start GRASS, using the LOCATION and MAPSET that contains the data to be converted.
5. Remove all MASKS from the GRASS data you wish to convert.
6. Double check the current region (and resolution, if converting raster data), either set the region from the map to be converted or set to default, and exit GRASS.
7. Copy the AML file and the five associated `.sh` files from the CD-ROM into the `arc` directory of your MAPSET. The AML scripts must be in this directory since the GRASS programs are written to look for the `arc` directory in the MAPSET.
8. Start ARC/INFO.

9. Type `pwd`. If you are not in your `/GRASSMAPSET/arc` directory use the `w` command in `ARC/INFO` to get into the `arc` directory.
10. Start the conversion AML by typing: `&run gras2arc.aml`.
11. Pick the data type you wish to convert to `ARC/INFO`.
12. When prompted, enter the GRASS database directory. This is the same directory you entered previously as your `GISDBASE`. If you cannot remember the path, look up at the current variable settings, and the `GISDBASE` should be listed on your screen.
13. The AML will list which `MAPSETs` are available in this `LOCATION`. Choose the one you are currently in for best results.
14. The AML will ask if you own this `MAPSET`. If the answer is no, the AML may terminate due to `UNIX` directory ownership permission settings. If you do not own the `MAPSET` and do not have read or write permission in the directories, this AML will not execute. You may want to have root change ownership of the `MAPSET`, or use the GRASS commands to copy the data files you wish to convert into a `MAPSET` you own (you must exit `ARC/INFO` to copy the GRASS maps).
15. The AML will list the `vector/raster/site` files available in the `MAPSET` depending on which type of data layer you chose earlier. Enter the name of the file you wish to convert. (Note: You **cannot** convert raster data without having the Arc Grid module.)
16. If you are converting vector data, the AML will ask if this is a line or polygon coverage. Enter the type of data you wish to convert. If there are both lines and polygons in the coverage, choosing `poly` is the best option and the AML will build for both labeled lines and polygons when in `ARC/INFO`.
17. The AML will ask you for an `ARC/INFO` coverage name for the converted data. The AML will test to ensure the entered coverage name is a valid `ARC/INFO` file name, and will allow you to re-enter a name if an invalid file name is given.
18. The AML will now execute and create your coverage.
19. You may convert multiple coverages in a row. The scripts may “bail out” after repeated use in a single session. This may be overcome by exiting `ARC/INFO` and restarting `ARC/INFO`.
20. The AML may leave temporary files behind in the `MAPSET/arc` directory. These files can easily identified and deleted.

Things to Keep in Mind During the Process

Data Conversion Basics

Always have a backup copy of your data before beginning any data conversion. In general, it is good practice to maintain a regular back-up schedule to recover data that may be lost under any circumstance.

Document everything you do. Documenting each step of the process not only allows you to reproduce the result later, but it also provides a record for error checking.

Make sure you have ample free disk space in the directory where the data conversion process will be performed. Some of the GRASS output files generated during conversion can be large and the resulting ARC/INFO coverages even larger.

For additional information on data conversion projects, we suggest reading *GIS DATA CONVERSION, Strategies, Techniques and Management* edited by Pat Hohl for Onword Press, 1998.

To do in GRASS

Take some time to preview the GRASS layer within the GRASS software. GRASS layers are ported out of the software one layer at a time and are then generated into ARC/INFO coverages. Upon completion of this process, check the newly generated ARC/INFO coverage for data conversion errors, data deletion, and for quality control/quality assurance measures. One good way to check the data conversion results is to make check plots within GRASS before converting the data, and compare with check plots made within ARC/INFO after the conversion is complete.

Quality Assurance/Quality Control

GRASS polygon or area layers generated into ARC/INFO coverages may need additional quality control. Many GRASS data layers created via a digitizer may have slivers, or small polygons that were not caught when the data was created, and were ignored by GRASS programs. Once the data is in ARC/INFO format, these polygons show up having a very small area, and no attributes. One easy way to determine if your data layer has slivers is to look at label errors. Once you have determined you have slivers, the easiest way to locate them is to view the data in ArcView. After the data is converted, export it to ArcView, look at the map table, do an incremental sort, and the unlabeled slivers will show up at the top of the table. You can use the zoom-to tool to find the slivers to be edited. When using ARC/INFO, a similar procedure may be done using ArcTools in the edit mode.

In ARC/INFO

Before beginning the process, design your ARC/INFO database. Define the workspace name and the contents of the workspace. Develop a data dictionary that will specify the contents of the workspace and determine the name of each coverage to be generated. Determine what attributes will be incorporated into each coverage and consider how the INFO files for each coverage will be used. This will help in defining the field name, type, width, number of decimal places, and the output format.

Projection information is vital to GIS data. Chances are the maps will not need to be re-projected as part of the conversion process. However, each coverage should include the `prj.adf` file that serves to document the projection. The `prj.adf` file can be created using ArcTools or with ARC/INFO commands.

If you wish to project your data using ArcView, remember it can only project from decimal degrees. If the data are properly projected before being viewed in ArcView, just set the units under view properties. It is easiest to project the data before export or transfer to ArcView since ArcView can only project from decimal degrees.

The conversion procedure outlined here does not create a bounding box for the converted coverages. The ARC/INFO command `rebox` can be used to create the bounding box for the coverage. The `rebox` command defines the `bnd` as the smallest area that will contain the all of the data in a coverage. The need for a coverage extent is described in Chapter 3.

To be able to use the ARC/INFO data in ArcView, or on another operating system, or for transport to another file system, you may wish to export the data. This can be done using the `export cover` or `export grid` commands in ARC/INFO. These commands will create a file with a `.e00` extension (if multiple export files are created, they will have incremental extensions up to `.e99`).

The sizes of the fields in the INFO tables are part of the AML. If you find your attribute information is being truncated, use a UNIX editor to edit the character string length. The number of spaces in the character string length was kept small to save on storage space in the INFO directory. If you have more than one different theme of information as an attribute label, you may want to consider separating the information into various tables and use the ARC/INFO `relates` to link the information.

NOTE: The AML and associated scripts require both GRASS and ARC/INFO to be installed and executable, running on a single UNIX machine. This is necessary because the AML and associated scripts make calls to the GRASS, ARC/INFO and UNIX libraries.

5 Quality Assurance and Quality Control Procedures

As is true with all data creation or conversion projects, quality assurance and quality control (QA/QC) procedures must be developed to ensure user confidence in the data. Many of the necessary QA/QC procedures cannot be automated and must be done manually. As part of the testing of the data conversion AML and associated scripts, a procedure was developed to check on the area of the polygons before and after conversion to ARC/INFO. The scripts described in this document are located in Appendixes A and B on the attached CD-ROM. They must be copied to a directory on your UNIX drive before executing any of the scripts.

This chapter describes one methodology available to quality check the polygon vector maps after conversion from GRASS to ARC/INFO format using the AML and associated conversion scripts developed by USACERL and ESRI. The QA/QC scripts perform several comparisons between the original GRASS map and the converted ARC coverage in order to identify:

1. Differences in polygon areas by category value.
2. Unlabeled polygons in the ARC cover. In testing the conversion it was found that the GRASS vector maps may contain very small slivers where lines overlapped. While GRASS ignored these data errors, ARC converts them into small, unlabeled polygons. Once identified they are easy to edit and correct in ARC/INFO.
3. Attribute values from the ARC cover that do not appear as a category value in the GRASS map.
4. Category values from the GRASS map that do not appear as an attribute value in the ARC cover.

Steps for Quality Control Check

The Quality Control scripts require two base information files as input. One provides the base information from the original GRASS map, the other the information about the converted ARC cover.

To produce the base information for the GRASS map:

1. The summary information for the GRASS map can be produced only through the raster functions. You must therefore first use `v.to.rast` to create a raster version of the vector polygon map. Make sure the region is set and that no MASK is present.
2. Create a file of raster map categories and the area of each category in square meters with:

```
r.stats -za GRASSMAPNAME > FILENAME
```

To produce the information from the converted ARC cover:

1. Read the coverage as a theme into an ArcView view.
2. Under the View menu, and the Properties submenu, set the map units and the distance units to meters, then click OK.
3. Double click on the legend so you can make the following changes.
 - a. Set the Legend Type to Unique Value.
 - b. Set the Values field to Attribute, then click Apply.
4. Open the table for the polygon coverage. Selecting the Theme menu to access the Table option will accomplish this.
5. Export the table. First make sure the table has focus. Under the File menu, execute the Export option. Make sure you set the export format to Delimited Text and save as a file.

The required files now exist and the scripts can run. The script will prompt for input and output files. NOTE: The scripts require a program called 'nawk' that is found with the Solaris Operating System, and may not run properly with other operating systems.

Start the script by typing in the path to the directory that contains the scripts and the startup script name, `poly_qc`. The output file that is created contains two sections, as shown in Table 1.

Reviewing the Output of the Check

Table 1 is an example output file from the QA/QC check. The output is formatted as a report with two sections.

The first section reports the area (square meters) shown for each category, followed by the difference (in square meters) between the ArcView and GRASS areas. The last column reports the percent difference between the ArcView and GRASS areas. The percent difference is calculated by using the formula:

$$\text{Percent} = (\text{abs}(\text{difference}) / ((\text{ArcView area} + \text{GRASS area}) / 2)) * 100$$

Since the methods used by GRASS and ArcView to calculate areas are not the same, we can't expect an exact match; however the areas should be reasonably close. The QA/QC script checks for percent differences greater than 5%, and highlights those categories in the report for further verification and investigation. The 5 percent is an arbitrary value and may be changed by the user by editing the script.

The base information file from ArcView reports the area for every polygon as a unique entity. However, the GRASS `r.stats` program reports area as a sum for each category. Unless each polygon in the GRASS map has a unique category, we cannot do a direct comparison to the ARC/INFO counterpart of the map. Therefore, in the area comparison, the ArcView file is processed so the areas for all polygons of the same category are summed. Keep this in mind if the output identifies a problem that needs to be investigated; be sure to query all polygons that make up a given category.

The second section of the report produces a summary that highlights: (1) the polygons in the ARC/INFO coverage without any attributes, (2) the categories in the ARC/INFO map not found in the GRASS map, and (3) the categories in the GRASS map not found in the ARC/INFO map. A successful data conversion will report no summary information for each of these tests.

Table 1. QA/QC Report.

<u>Cat</u>	<u>ArcView area</u>	<u>GRASS area</u>	<u>Difference</u>	<u>Percent</u>
1	622461.38	620000.00	2461.38	0.40
2	919849.07	913200.00	6649.07	0.73
3	339074.29	339200.00	-125.71	0.04
<u>SUMMARY</u>				
ARC polygons with no attributes:				
Categories in the ARC File not found in the GRASS File:				
Categories in the GRASS file not found in the Arc File:				
<u>Sample ArcView Table</u>				
0.00048	0.27853	409	0	0
0.3351	9.96722	352	0	0
0.05282	2.88405	460	0	0

6 Summary

This document describes a few methodologies for converting GRASS GIS data to ARC/INFO data format. It provides consistent results and has been tested by USACERL, ESRI, and installation personnel without causing additional data errors. The conversion of the GIS data is only the first step in the migration from Government off-the-shelf GRASS to Commercial off-the-shelf ESRI products. The QA/QC procedures included in this document are only the first step in quality checking the GIS data.

References

GRASS Version 4.1 User's Reference Manual, USACERL, Spring 1993.

GIS Data Conversion: Strategies, Techniques, and Management, ed. Pat Hohl, High Mountain Press, 1998.

Harmon, Victoria, Douglas M. Young, and Melissa A. Records, *GRASS-Intergraph Data Conversion Guide*, Technical Manuscript N-92/01, USACERL, April 1992.

ArcDoc, Version 7.1.2. 1982-1997 Environmental Systems Research Institute, Inc., Redlands, CA.

Appendix A

Appendix A is on the attached CD-ROM. The following 6 files are used for the data conversion processes discussed in Chapter 4, Automated Conversion Process:

gras2arc.aml

exsites.sh

exvects.sh

exvects2.sh

fgfiles.sh

lgfiles.sh

Appendix B

Appendix B can be found on the attached CD-ROM. It contains 3 scripts file for use with Chapter 5, Quality Assurance and Quality Control Procedures.

`poly_ac` – gets inputs and run the quality checks

`sum_cat.awk` – sums the area totals of like categories in ArcView

`compare.awk` – compares the ArcView and GRASS files

Appendix C

There are many manual methods for converting GRASS GIS data to ARC/INFO GIS data. This appendix outlines two of the most common methods used by U.S. Army personnel in the conversion of GRASS data to ARC/INFO data. None of the methods described below automate the quality assurance or quality control measures during the data conversion process. Manual editing and additional ARC/INFO commands may be required before the data is ready for analysis in ARC/INFO. If converting complicated polygon data, you may wish to output the data from GRASS as a line and then follow both the line and polygon options into the same coverage in ARC/INFO. This will decrease the amount of editing necessary in ARC/INFO.

Polygon data conversion using the GRASS command `v.out.arc`

Start the GRASS software and get into the MAPSET that contains the data you wish to convert.

```
cd to $LOCATION
```

From within GRASS run the `v.out.arc` command using the polygon option. This will create three files, `out.pol`, `out.lab`, and `out.txt` where “out” is the name you specified as the output file name.

Exit GRASS

Start ARC/INFO

Create a new, or get into an existing workspace

Copy all of the `v.out.arc` files from your GRASS MAPSET into your workspace directory

Execute the `generate` command, syntax is as follows.

```
generate (covername)
```

```
input out.pol
```

```
polygons
```

```
input out.lab

points

quit

build (covername) poly

list (covername.pat)
```

Now you must use your UNIX editor to make a few changes to the out.txt file so it can be used to attribute the ARC/INFO data. The space field separator must be deleted and a comma separator inserted in its place. If you are using the text editor, vi, the command is :%s / / , and must be repeated three times. If you do not have any blank spaces in your labels in the .txt file, you can use :%s/ / ,/g once to replace the blank spaces. If there is an END at the bottom of the file, you must delete it.

Now you are ready to create and define the attribute tables for your coverage. Invoke the tables command in ARC/INFO. Your line should look something like this:

```
Define (covername).pat2

Item name: ROW#

Item width:

    Item output width:

    Item type:

    Item name: (covername)-CAT

    Item width:

    Item output width:

    Item type:

    Item name: (covername)-ID

Item width:

    Item output width:

    Item type:
```

Item name: (covername)-AAT

Item width:

Item output width:

Item type:

Select (covername).pat2

Items

List

Add ROW# (covername)-CAT (covername)-ID (covername)-ATT from
out.txt

List

Quit

The item width, item output width, and item type are explained in
Chapter 3 of this document.

joinitem (covername).pat (covername).pat2 (covername).pat3 (covername)-ID
(covername)-ID ordered

Invoke the tables command

Rename (covername).pat (covername-old).pat

Rename (covername).pat3 (covername).pat

NOTE: If there are labeled lines in your polygon data, you should also go through the line option below to ensure all of the information in your GRASS data layer gets converted to ARC/INFO. This polygon conversion technique has had problems converting polygon data that is not topologically pristine. If you view the data in ArcEdit, and it appears to be "webbed," or additional lines have been created, you are better off using the automated or dlg methods for your data conversion. Additionally, after you generate your polygon coverage, and see many zero length errors, you may have a lot of editing to do in ARC/INFO.

Line data conversion using the GRASS command `v.out.arc`

Start the GRASS software and get into the MAPSET that contains the data you wish to convert.

```
cd to $LOCATION
```

From within GRASS run the `v.out.arc` command using the `line` option. This will create two files, `out.lin` and `out.txt` where “out” is the name you specified as the out file name.

Exit GRASS

Start ARC/INFO

Create a new, or get into an existing workspace

Copy all of the `v.out.arc` files from your GRASS MAPSET into your workspace directory

Execute the `generate` command, syntax is as follows.

```
generate (covername)
```

```
input out.line
```

```
lines
```

```
quit
```

```
build (covername) line
```

```
list (covername.aat)
```

Now you must use your UNIX editor to make a few changes to the `out.txt` file so it can be used to attribute the ARC/INFO data. The space field separator must be deleted and a comma separator inserted in its place. If you are using the text editor, `vi`, the command is `:%s / /`, and must be repeated three times. If you do not have any blank spaces in your labels in the `.txt` file, you can use `:%s/ /,/g` to replace the blank spaces. If there is an `END` at the bottom of the file, you must delete it.

Now you are ready to create and define the attribute tables for your coverage. Invoke the tables command in ARC/INFO. Your line should look something like this.

```
Define (covername).aat2
```

```
Item name: ROW#
```

```
Item width:
```

```
Item output width:
```

```
Item type:
```

```
Item name: (covername)-CAT
```

```
Item width:
```

```
Item output width:
```

```
Item type:
```

```
Item name: (covername)-ID
```

```
Item width:
```

```
Item output width:
```

```
Item type:
```

```
Item name: (covername)-ATT
```

```
Item width:
```

```
Item output width:
```

```
Item type:
```

```
Select (covername).aat2
```

```
Items
```

```
List
```

```
Add ROW# (covername)-CAT (covername)-ID (covername)-ATT from out.txt
```

```
List
```

Quit

The item width, item output width, and item type are explained in Chapter 3 of this document.

```
joinitem (covername).aat (covername).aat2 (covername).aat3 (covername)-ID  
(covername)-ID ordered
```

Invoke the tables command

```
Rename (covername).aat (covername-old).aat
```

```
Rename (covername).aat3 (covername).aat
```

Polygon data conversion using the GRASS command v.out.dlg

Some GRASS data files create instances where major line webbing or shifting of polygon lines to occur during the conversion from GRASS to ARC/INFO. There are three choices in this situation: (1) Use the automated methods described in this report, (2) Edit the arcs in ArcEdit, check all labels, and make check plots to verify that no lines were shifted during the conversion, (3) Follow the methodology described below.

In GRASS, execute the `v.out.dlg` command

Exit out of GRASS

In ARC/INFO, execute the `dlgarc` command. Your input line should look something like this: `dlgarc optional out.dlg (covername)` (there are additional options you can add to this if it fits your data, see `dlgarc` command usage).

```
Build (covername) poly
```

Execute the ARC/INFO `additem` command. This additional item will represent the data attribute. For example, it could look something like: `filename-att`. Make sure the width is large enough to accommodate your attribute length, and if the attribute is text, you enter it as a character).

```
Build (covername) poly
```

You are going to need to re-label all of your polygons. While this option ensures the integrity of the arcs that compose the polygons, the labels are lost in the data conversion process, and only the label points are converted. You may use ArcEdit or ArcTools to re-labels to your polygons.

This method can be very labor intensive. The automated method may be a better option for converting GRASS polygon data to ARC/INFO data.