

UC Santa Barbara

GIS Core Curriculum for Technical Programs (1997-1999)

Title

Unit 32: Managing Digital Libraries

Permalink

<https://escholarship.org/uc/item/7668f22h>

Authors

Unit 32, CCTP
Minter, Tim

Publication Date

1998

Peer reviewed

UNIT 32: MANAGING DIGITAL LIBRARIES

Written by Tim Minter, St. Johns River Water Management District,
Palatka, Florida

Context

Managing digital GIS data libraries can facilitate efficient and reliable access to validated, standardized, and documented GIS data. Responsibility for managing enterprise GIS library databases often rests with specifically assigned individuals who possess full understandings of the processes used to develop and prepare the data for acceptance into the library. However, individual users may employ this skill to organize their project data and share it with other users.

Library management techniques have been developed for geospatial vector and raster data types. Techniques for both data types can rely on tiling structures which are used to subset spatially the area of interest from the rest of the data within the layer before supplying the data subset to the user. A subset of the data can be served to and processed by the client quickly and efficiently. Carefully planning the tiling structures aids developing, updating, maintaining, and expanding the library.

The following example presents a typical vector data application of tile-based geospatial data library management skills.

Example Application

St. Johns River Water Management District (SJRWMD), which is responsible for water quality, water quantity, natural resource protection, and flood control within an approximately 12,600 square mile region in northeast and central Florida, has formally implemented a Geographic Information System to facilitate its data collection, analysis, and presentation projects. In support of the GIS, a set of standard base data layers has been identified and is to be managed within a digital library called the GIS Library

Database. All GIS users in the agency are to be given easy, efficient, and reliable access to validated, standardized, and documented geospatial data layers within this database. GIS data sets specific to departmental users are to be organized within the same structure as the GIS Library Database, but will be managed within their respective departments.

The scientists in the Department of Water Resources want to build a library in which all finished data layers are to be managed. The initial data layers are Submerged Aquatic Vegetation, Water Quality Sampling Sites, and Project Areas of Interest. The existing enterprise GIS Library Database structure and GIS data management strategies provide the following guidelines for inserting data layers into the Water Resources GIS Library Database.

1. Data layers must be technically correct.
2. Data layer content and spatial properties must have been validated.
3. Data layer standard documentation must be present.
4. Available tiling schemes are based on regular and irregular polygons.
 - USGS Quadrangle Boundaries (regular)
 - Political Unit Boundaries (irregular)
5. Available system locations for GIS library data layers are currently distributed among the users, but centralization is planned.
6. Sharing data with staff and external requestors is encouraged.

Project Managers, GIS Analysts, and GIS Technicians are responsible for developing the data according to standard operating procedures. The Water Resources GIS database custodian will be responsible for verifying that each data layer has been fully validated, standardized, and documented before posting it in the department's library database. The District GIS database administrator designs, tests, and assists in propagating the library structure in the departments. The administrator is also responsible for assisting users to access and utilize the data in an informed manner.

Learning Outcomes

The following list describes the expected skills which students should master for each level of training, i.e. Awareness/Competency/Mastery.

Awareness:

The learning goals of this section are to achieve a general understanding of digital library design and implementation, the benefits to users and a working knowledge of digital library management vocabulary.

Competency:

The learning goals of this section are to develop the ability to create a tile structure, add a data layer, and then access it as a user.

Mastery:

The learning goals of this section are to be able to design, implement, and maintain a digital, geospatial library.

Preparatory Units

Recommended:

1. Unit 7 - Using and interpreting metadata
2. Unit 30 - Validate databases
3. Unit 31 - Manage database files
4. Unit 34 - Pre/Post overlay tasks / Types of overlay operations
5. Unit 35 - Point/Line in polygon operations
6. Unit 41 - Using boolean search techniques
7. Unit 47 - Prepare materials for on-screen visualization

Complementary:

1. Unit 20 - Use text editors
 2. Unit 26 - Edit point data
 3. Unit 27 - Edit linear data
 4. Unit 28 - Edit polygon data
-

Awareness

Learning Objectives:

1. Student can define basic terms relating to digital library management.
2. Student can explain the types of operations that relate to digital library management.
3. Student can explain digital library management schemes which are unique to particular data models, i.e. vector and raster.
4. If software will be used, student can describe the commands in the software which relate to managing digital libraries.
5. Student can explain how data users benefit by the availability of managed digital libraries.
6. Student can describe a real application of digital library management and its implementation.

Vocabulary

- area of interest
- client/server
- data access
- data layer

- data set
- data theme
- raster data
- system architecture
- tiling system
- topology
- vector data

Topics

1. Unit Concepts

Data management capabilities and methods may vary widely among software systems.

- Name of the digital library management software
- Type(s) of data managed by the software
- Procedures employed by the software
- Commands or menu choices and their respective parameters
- Pre-processing steps which facilitate library management procedures

Competency

Learning Objectives:

1. Student will be able to organize vector data in a library structure.
 - Create a new map library.
 - Create a new layer in the map library.
 - Add data to the new layer.
2. Student will be able to organize raster data in a library structure.
 - Create a new image catalog.
 - Add data to the catalog.

Generic List of Tasks

- For vector data:
 - Choose tile structure
 - Choose computer system location for data
 - Create a new map library
 - Create a new layer in the map library
 - Add the data into map library
 - Test for successful completion
- For raster data:
 - Choose computer system location for data
 - Create image catalog

- Add image information to catalog
- Test for successful completion

Insertion Tasks

1. Read the instructions provided with the management software.
2. View descriptions of the layer to be inserted. Check to be certain that it has passed all verification tests.
3. Look at the layer using on-screen visualization capabilities. Compare the layer location with the tile structure location.
4. Prepare a list of the tiles in which the data will reside.
5. Test for adequate disk space in the intended system location and in the data processing location.
6. Insert the data.
7. Test for successful completion.

[Example Implementation: Creating a new map library](#)

Mastery

Learning Objectives:

1. Student will be able to identify appropriate and inappropriate situations for including GIS data in digital libraries.
2. Student will be able to specify hardware and software parameters necessary to support digital libraries.
3. Student will be able to provide training, consulting and troubleshooting services to others in relation to digital library management.

Generic List of Tasks

1. Assess and organize a given superset of GIS data layers for management within digital libraries.
 2. Design hardware and software environment for managing the digital library.
 3. Design standard operating procedures and training documentation which describe specific steps necessary to implement and maintain the digital library.
-

Follow-up Units

Suggested:

1. Unit 52- Implementing and Managing GIS
2. Unit 53 - Communicating about and distributing GIS products
3. Unit xx - Geography for Technical GIS
4. Unit xx - Computer Skills for Technical GIS

Created: May 14, 1997. Last updated: October 5, 1998.

Example Implementation: Creating a new map library

This example procedure describes how to create a new map library using the ESRI ARC/INFO Librarian database management module. For the purposes of this example the new map library will be named **WRLIB** (Water Resources Library). This specific process was developed in direct reference to the ARC/INFO online documentation (ARCDOC) for Librarian management. This process is designed for use in a computing environment in which the machinery are SunSparc 20s and Sun Ultras, the operating system is Sun Solaris 2.5 (Unix), the user interfaces are Common Desktop Environment (CDE) and OpenWindows operating on an FDDI token ring network using NFS mounting protocol. Users with PCs can access the environment through Hummingbird X-Windows software.

Actions for which ARC/INFO commands are available are capitalized,

- e.g. BUILD all coverages before MAPJOINing them.

Unix commands are prefaced by a typical Unix prompt naming convention of '<machine name>.<user name>% ' <command> {arguments}',

- e.g. abaco.tim% cp oldfile newfile
-

1. Choose a Tile Structure

Most Department of Water Resources project work is performed in large scale areas of interest. The GIS datasets created for, and resulting from these projects are best stored in a large scale tiling scheme. Index12 is based upon the USGS Quarter Quad, or 3.75' quadrangle, and is the best available tiling scheme for the new map library.

- USGS Quadrangle Boundaries
 - Index12: USGS 3.75' Quadrangle Boundaries
 - Index24: USGS 7.5' Quadrangle Boundaries
 - Index100: USGS 15' Quadrangle Boundaries
 - Index250: USGS 30' Quadrangle Boundaries
- Political Unit Boundaries
 - County Boundaries
 - State Boundary

2. Add an environment variable to the user interface environment files

The tile index coverage (see next step) stores a system pathname to each tile workspace directory. The value stored in the index coverage can be the name of an environment variable. Using an environment variable resolves the problems related to actual storage device location of the data. If the data location changes, just reset the system variable. Using this technique can allow structural alterations to be effected without interrupting user access to the data. A well-managed system will utilize standard master files which set a basic environment for all users. The environment variable should be set in these master files. You may need to consult with your system administrator to use this technique. In the directory where the master environment files reside, add the variable setting to the '.login' (Unix OpenWindows) & '.profile' (Unix Common Desktop Environment) files

```
setenv WRLIB /gislib/librarian/database/wrlib_ref/
```

3. Create the tile index coverage

Since data layers will continue to be added to this library for the foreseeable future, and since the spatial extent of the future data layers is as yet undetermined, the spatial extent of the tile coverage will cover all of UTM Zone 17 from east to west (this zone includes the entire peninsula of Florida) and from south to north will include the area from the Florida Keys to the area covered by the Floridan Aquifer (Georgia and South Carolina). Creating the tile index coverage to cover an area large enough to include features which will only possibly be mapped can eliminate difficulties associated with re-tiling the library in the

future at a minor cost in development time and disk space. The name of this coverage will be INDEX, and it will reside in the Library Reference Workspace.

- o The tile coverage must be extremely clean and well developed.
- o The tile coverage must adhere as appropriate to the SJRWMD standard general [vector data characteristics](#).
- o One or more items in the INDEX.PAT must cause each polygon to be unique.
 - eg, QUAD-NAME, QUAD-NUMBER, QUARTER or something similar
- o In Arc: ADDINDEXATT INDEX
 - This command adds two items to the INDEX.PAT; TILE-NAME and LOCATION
- o In INFO, populate the TILE-NAME item with values from the unique-value item(s) using CONCATENATE.
 - CONCATENATE TILE-NAME FROM 'tile',QUAD-NUMBER,QUARTER
- o In INFO, populate the LOCATION item with the correct path, then the TILE-NAME. Instead of a specific system path string, the system environment variable set in a previous step is inserted in the LOCATION item at this step. When a user accesses the data, the '\$' sign prompts the operating system to use the value stored in the variable rather than the string which describes the variable.
 - CONCATENATE LOCATION FROM '\$WRLIB',TILE-NAME
 - Example INDEX.PAT developed from the USGS 3.75' quad tile coverage INDEX12. The items included after LOCATION are user-defined items intended to facilitate future use of this index coverage.

INDEX.PAT

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	ALTERNATE NAME	INDEXED?
1	AREA	8	18	F	5	.	-
9	PERIMETER	8	18	F	5	.	-
17	INDEX#	4	5	B	-	.	-
21	INDEX-ID	4	5	B	-	.	-
25	TILE-NAME	32	32	C	-	.	-
57	LOCATION	128	128	C	-	.	-
57	QUAD-NAME	40	40	C	-	.	-
57	WMD#	4	4	I	-	.	-
229	QUARTER	2	2	C	-	.	-
231	USGS#	7	7	C	-	.	-
238	UTM_ZONE	2	2	I	-	.	-
240	ADD_TILES	1	1	I	-	.	-

4. Identify tiles in INDEX for which Library Tile Workspaces will be created:

The boundaries within which SJRWMD has responsibility are drawn politically utilizing existing state, county, and public land survey system boundaries. These politically drawn boundaries were based as closely as possible to physical surface water basin system boundaries. Most Dept. of Water Resources projects are based upon and bounded by surface water basins. ARC/INFO coverages for the SJRWMD boundary and the associated surface water boundaries exist. The surface water basins boundaries coverage will be used to create the initial Library Tile Workspaces.

- o CLIP INDEX SWBASINS TILECOV NET 0.0001
- o FREQUENCY TILECOV.PAT TILECOV.TILE-NAME
TILE-NAME

- end
- end
- o In Arcedit:
 - edit INDEX
 - editfeature poly;de poly;draw
 - relate add
 - Relation Name: tilere1
 - Table Identifier: tilecov.tile-name
 - Database Name: info
 - INFO Item: tile-name
 - Relate Column: tile-name
 - Relate Type: linear
 - Relate Access: ro
 - Relation Name:
 - sel tile-name ne "
 - resel tile-name = tilere1//tile-name
 - nsel
 - calc tile-name = ";calc location = "
 - relate drop
 - Relation Name: tilere1
 - Relation Name:
 - save;quit;BUILD INDEX POLY

5. Create the Library Tile Workspace directory

Library Tile Workspaces can reside on any disk which is shared on the network. In this case, the Dept. of Water Resources GIS Database Administrator has determined that the most appropriate disk drive for this library is a 9gb drive on a machine named Luna.

- o abaco.tim% cd /net/luna/data
- o abaco.tim% mkdir wrlib_tile

6. Create the Library Tile Workspaces

- o abaco.tim% cd wrlib_tile
- o abaco.tim% arc
- o Arc: librarian
- o Librarian: BUILDTILES <pathname to index coverage>/<index coverage>

7. Create the new map library

- o abaco.tim% cd /gislib/librarian/database
- o abaco.tim% arc
- o Arc: librarian
- o Librarian: CREATMAPLIBRARY wrlib <pathname to index coverage>/<index coverage>
transactional

8. Create the Master Tic Coverage

The Master Tic coverage will allow future data entry from a digitizing tablet if necessary. In this case, the document from which data is converted will need to possess quadrangle corners or other features which are coincident with 3.75', 7.5', 15', or 30' quadrangle corners. The coverage must be named MASTERTIC, and it will reside in the Library Reference Workspace.

- o abaco.tim% cd /gislib/librarian/database/wrlib_ref
- o abaco.tim% arc
- o Arc:CREATE MASTERTIC INDEX
- o Arc: WORKSPACE /net/luna/data/wrlib_tile
- o Arc: librarian
- o Librarian: VISIT arc copy /gislib/librarian/database/wrlib_ref/mastertic
(This copies the mastertic file to all of the tiles)
- o Librarian: VISIT arc rebox mastertic

(This leaves only four TICs which correspond to each tile)

9. Use **GRANT** in librarian to give users access (**Manager, Operator, Browse, Look, None**):

- Librarian: LIBRARY INDEX24UTM
- Librarian: GRANT tim manager library
- Librarian: GRANT dla manager library
- Librarian: GRANT \$rest browse library

10. Add data to the library [*outdated link has been removed*]

GIS Library Data Development

Generalized Technical Specifications for Vector Data

These technical specification elements are specific to the St. Johns River Water Management District GIS environment. They were developed in reference to the Mapping and Geographic Information System Standards and Conventions Committee's Mapping and GIS Standards Report, October 1993. Only a subset of the listed elements are appropriate for any given GIS data development effort. The necessary subset is to be determined by District staff responsible for developing GIS data. Information Resources Department, GIS Program Management Division staff can be consulted for assistance in determining which elements are necessary for a given GIS data development effort, but final responsibility for the data definition lies with the project manager.

Documentation

Complete and correct FGDC compliant documentation must accompany the geospatial dataset.

Positional Accuracy

- All coordinates shall be in the Universal Transverse Mercator (UTM) projection, zone 17, referenced to the North American Datum (NAD) of 1983, 1990 adjustment.
- Where appropriate, all data shall be tiled by NAD83 USGS quadrangle boundaries
- Projection information must be present and correct in each coverage projection file and be accessible through the Arc command DESCRIBE.
- Define the maximum allowable RMS error for source document registration (0.003 - 0.008, depending on source data accuracy, medium, etc.). Actual RMS value for each coverage shall be noted. If the defined RMS value cannot be met, a note explaining why must accompany the coverage.
- TIC x,y coordinates must match their respective registration coverage TIC x,y coordinates exactly.
- When more than one data layer is required for the end product, the coverages' respective TIC ids and x,y coordinates must match each other and the registration coverage exactly.
- All point features shall be located at the center of the symbol or feature on the source document which depicts the feature.
- Line features must be converted within one-half feature width of the feature represented on the source document (no daylight criterion).

Attribute Accuracy

- The attribute tables must contain their ARC/INFO coverage default items and contract specified items in the correct order and be defined correctly.
- There shall be no superfluous attribute tables
- All attribute items shall be defined correctly
- All attribute items shall exist in the specified order
- There shall be no superfluous attribute items
- All items shall be coded correctly

Logical Consistency

- All files must import from ARC/INFO EXPORT format files correctly
- No edit masks shall be present
- Coverages shall exist in DOUBLE precision
- Correct topology (POINT, ARC, POLYGON, LINK, NET, ROUTE, REGION topology) shall be present
- During the entire data development process, the ARC/INFO CLEAN routine shall not alter any feature components so that they do not meet the specified positional accuracy characteristics.
- Tolerance values for completed coverages shall be set to:
 - Fuzzy 0.0001
 - Dangle 0.0001
 - Tic Match 0.0001
 - Edit 1/100 of the greatest distance defined in the coverage .BND file
 - Node Snap 0.0001
 - Weed 0.0001
 - Grain 0.0001
 - Snap 0.0001
- All coverages shall be flawlessly coordinate edgematched to adjacent coverages so that a MAPJOIN and DISSOLVE will leave no edgematching, coding, or data capturing errors. Since product development often occurs in partial sets of coverages, each subsequent set shall be flawlessly coordinate edgematched to all previous sets.
- In case the source documents do not edgematch, the data developer shall notify the source document developer for resolution.
- Contiguous, identically coded polygons shall not exist.
- Coding combinations which by proximity or definition comprise a logical inconsistency in reference to real-world features shall not exist (this occurrence represents attribute miscoding and/or feature mapping errors).
- All features shall exist wholly within the registration coverage tile bounding arcs.
- No superfluous pseudo nodes shall exist.
- No superfluous tics shall exist.
- There shall be no duplicate features.
- No erroneous sliver polygons shall exist.
- No label errors shall exist.
- The universe polygon shall not possess a label.
- No dangle errors shall exist.
- No intersection errors shall exist.
- Any errors as defined here-in which occur as a result of the CLEAN process must be

rectified prior to coverage completion.

Completeness

The final product shall only be described as complete and fully accepted when all of the contract criteria are met. The completeness report shall include descriptions of the:

- relationship between the objects represented and the abstract universe of all such objects
- exhaustiveness of the spatial and taxonomic properties of the objects in the data set (were all kinds accounted for? were all specified features captured?)
- descriptions of the procedures used for testing, the test dates, and the test results