GIS Core Database Revision:

GIS Data Resource Site Specifications

August 24th, 2002

Robert Maki
GIS Database and Infrastructure Supervisor
Minnesota DNR
Management Information Services Bureau

Minnesota DNR
GIS Data Resource Site Specifications

Robert Maki
GIS Data and Infrastructure Supervisor
MIS-GIS

1.0 Introduction

This document provides technical specifications for GIS Data Resource Sites at the Minnesota Department of Natural Resources (DNR). Data Access Sites are part of the Revised GIS Core Database Architecture,[1] an information system initiative which seeks to created an integrated GIS data storage, access, distribution, and maintenance environment. The "Revised Core" initiative updates the original DNR GIS Core Database Architecture, which was established in 1995.

This document is intended for a technical audience, primarily system designers and application developers. Reviewers seeking a general understanding of the architecture may wish to consult other sources.[2] The specification has been prepared early in the Revised Core development project because several yet to be developed subsystems depend on the specification.

Data Resource Sites are encapsulated data, application, and user documentation environments that define the range of data resources available to users. Consolidating these resources into a single suite of structured locations facilitates data distribution and update, and software version control.

Data Resource Sites have several key characteristics, including: 1) may draw upon a wide variety of physical data sources which are encapsulated through a set of system level metadata structures, 2) includes not only data but also methods for accessing data, thereby enabling applications which do not require users to navigate file systems and data sources to load data into their application environment, 3) are organized, in part, around specific business units with an eye toward supporting business-specific applications.

The document consists of several sections as follows:

Section Number	*Section Title*	*Section Description*
Section 2:	Background	Historical context for the document

[1]Minnesota DNR, Management Information Services Bureau, "Revised Core Database General Architecture Design", March 2000

[2]Specifically: Minnesota DNR, Management Information Services Bureau, "GIS Core Database Architecture Revision", 1999.

Section Number	*Section Title*	*Section Description*
Section 3:	Upper Level Directory Standards	Top level file server standards for Data Resource Sites
Section 4:	Application Data Storage	Comprehensive treatment of application data storage specifications on Data Resource Sites
Section 5:	System Level Metadata	Specifications for metadata components that support application development around resource sites. This includes file format descriptions and physical organization.
Section 6:	Applications	Physical storage of applications and application support data. Serves both Windows NT and UNIX-based environments.
Section 7:	Process	Description and rationale for a dedicated processing area of the site.
Section 8:	Wrap-Up	Parting remarks.
Appendix 1	Operating Environment Description	Description of baseline assumptions, and hardware/software domains associated with the site specifications.

2.0 Background

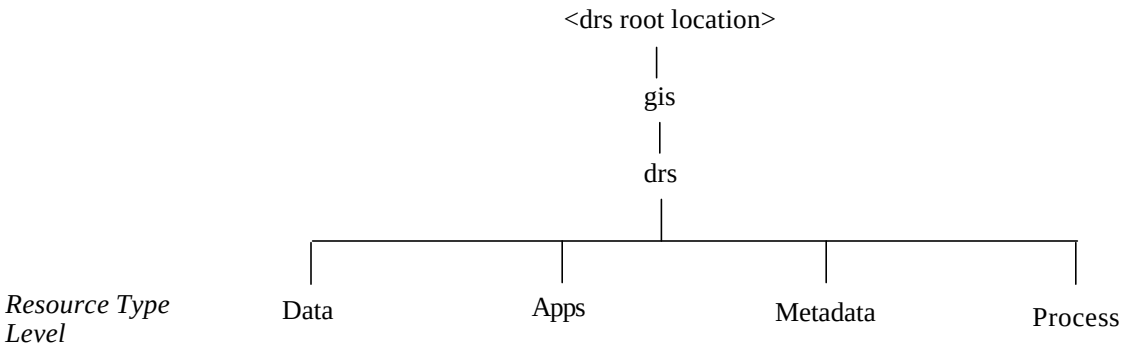
Experienced DNR Core Database users will recognize similarities between this document and the "DNR GIS Data Storage Standards", first published in 1995. The specifications presented here both compliment and replace the original specifications. The original standards will be retained as part of the department database administration environment, while this new specification forms the basis for the data resource sites which most users and business applications will interact with.

3.0 Upper Level Directory Standards

This section outlines standards and specifications for physically organizing a Data Resource Site file system. The first subsection outlines the top-level specifications. Later subsections provide detailed specifications for the various types of information which constitute a Data Resource Site, including: 1) data, 2) applications, 3) metadata, and 4) process.

3.1 Top-Level Directory Specifications

Top-level directory structures of Data Resource Sites include A root location followed by two levels named with “gis” and “drs” key words, followed by Resource Type Identifier level, as follows:



3.2 Data Resource Site Identifier {this section has been withdrawn}

3.3 Resource Type

The resource type identifier indicates the major site component stored in the various branches of the filesystem below the resource site root. Four components exist at this level. These are listed below.

Resource Type	*Resource Description*
data	GIS filesystem based data resources storage
apps	UNIX and NT-based applications, including ARCVIEW extensions. (Subdirectory structures to be determined)
metadata	System level metadata used by applications
process	Designated scratch space for application processes. Intended to isolate temporary files and system process files from data and application portions of the file system.

4.0 Application Data Storage

Application data constitute the full range of data types available to users, including: 1) fileserver-based GIS data, 2) database server resident GIS data, and 3) traditional (non-spatial) data served via a database server in a client-server environment.

4.1 General Specifications For All Data Types

A Data Resource Site stores and presents data in a "layer" paradigm, where individual data sets exist independently, but are designed to be combined in map displays or analysis tasks. Each layer has an access methods associated with it. Certain specifications are applicable to all data types, regardless of the methods in which the data are stored and accessed. These are described in this section, and include the following specifications: 1) layer naming, 2) layer description, 3) coordinate storage, and 4) thematic classification.

4.1.1 Layer Names

Layers are discrete data sets that are referenced by way of a standardized data name regardless of the environment within which data are stored. Layer names adhere to a set of constraints and structural rules.

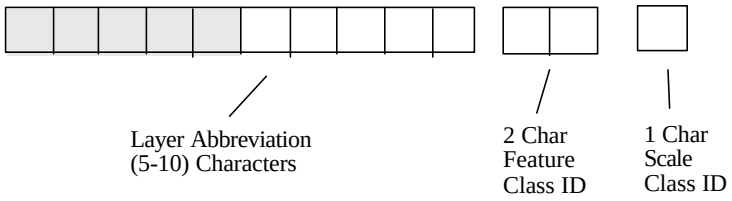
4.1.1.1 Layer Naming Constraints

Constraint	*Specification*	*Discussion*
Layer name length	Total length of the name will not exceed 13 characters	Name length is constrained by ARC/INFO coverage name limits. It should be noted that 13 character specification causes layer names to be unsupported in DOS and ISO9660 file systems.
Reserved characters	Non-alphabetic and non-numeric characters (including spaces) are disallowed in layer names with the exception of the underscore ("_") character.	DNR's diverse storage and application environment will create many potential opportunities for conflict with non-alphabetic characters (e.g. <,>,&,@,%,* all have known conflicts).
Case	All characters are lower case.	Lower case specification assists in standardizing application programming environments which may exhibit some degree of case sensitivity.

4.1.1.2 Layer Naming Structural Rules

The revised naming standard is a variable length string of eight to thirteen characters with three components, including: 1) layer abbreviation, 2) feature class, and 3) scale specification. Layer abbreviations can be from 5 to 10 characters in length, and form the thematic identity of the layer. They exist so that users can identify the general content and/or source of the data. Feature class refers to the type of cartographic primitive used to represent the feature (e.g. point, polygon, pixel). This part of the specification communicates an important data characteristic to users, and instructs applications as to feature class, an important criteria in controlling machine-based application processes. Scale specification refers to the scale range within which the cartographic and thematic components are expected to function, and can be appropriately applied. This part of the name will be used to help control scale-based display properties and provide users with a visual indication of a layer's scale properties. The specification is described graphically in Figure 1.

Figure 1: Layer Name Composition



Veteran GIS Core Database users will note the similarities between this specification and the one used in the current Core Database Storage Standard. The primary difference between the two is the expansion of the abbreviation component from 5 to (up to) 10 characters, and the new standard's variable length string property. In revising the specification, each of the three structural components was thought to be important to retain in the new environment.

4.1.1.2.1 Layer Abbreviation

The layer abbreviation component of the layer name is from five to ten characters in length. The first part of the string is reserved for a thematic subclass keyword that suggests the content. The second part of the string is used to further describe the specific data set. Thematic subclass keywords have not been fully defined at this time, but a partial list of candidate values might include:

keyword	***Meaning***	***Description***	***Example***
adm	administration	ownership units at some level of geography	adm_tawlndpy3
bdry	boundary	area and line features used to demarcate managed areas and DNR administrative areas	bdry_stprkpy3

keyword	*Meaning*	*Description*	*Example*
cent	centroid	centers of features that are generally represented as areas	cent_stprkpy3
ecs	ecological classification system	area and line features used to demarcate ECS mapping units	ecs_landtpy2
elev	elevation	features that describe terrestrial height of land	elev_dem30ra3
fldwy	floodway	area features that characterize the extent of actual or hypothetical flood events	fldwy_femapy3
forst	forest cover	features that describe elements of the forested landscape	forst_csapy3
img	imagery	registered images rendered through various remote sensing methods in raster format	img_doq03im4
indx	index	area features that are used almost exclusively to reference other features	indx_q024kpy3
lake	lake	area and line features used to demarcate basins principally with some deepwater habitats; may include palaustrine and riverine habitats (Glenn: I realize this is problematic, any suggestions?)	lake_usgspy2
lfrm	landforms	area or line features used to depict geomorphology characteristics	lfrm_mngeopy2
lulc	landuse-land cover	areas of generalized land cover and use characterization	lukc_gap94py3
map	map image	scanned and regstered map images	map_drg024im3
plce	place	features describing place names	plce_gnispt3
pls	public land survey	multiple feature classes used to reference public land survey elements	pls_fortypy3
prcp	precipitation	features that depict precipitation events and statistics	prcp_jan30py1
pveg	presettlement vegetation	features that describe presettlement vegetation characteristics	pveg_btreept3
rail	railroad	railroad facilities	rail_linesln3
rip	riparian	data related specifically to riparian habitats	rip_bf200py3
road	road	roadway features, either as lines or areas	road_majorln3
shdr	shaded relief	shaded relief images of terrestrial, submerged, or subsurface phenomena	shdr_elevim3

keyword	***Meaning***	***Description***	***Example***
shor	shoreline	shoreline facilities	shor_waspt3
slpe	slope	features that depict slope gradients or area aggregations of terrestrial or submerged lands	slpe_met25py2
soil	soil	features that describe soil characteristics	soil_surgopy3
strm	watercourses	line and area features used to describe overland streams, artificial flow paths, and area-based riverine habitats	strm_troutln3
temp	temperature	features that depict temperature events and statistics	temp_jan30py1
trl	trail	trail features, typically as lines	trl_ohvln3
well	well	features that depict well facilities	well_cwipt3
wetl	wetland	area and line features used to demarcate basins with principally palaustrine habitats; may include deepwater and riverine habitats	wetl_nwipy3
wsh	watershed	line and area features used to demarcate drainage areas on the landscape	wsh_maj81py3

4.1.1.2.2 Feature Class

The feature class ID component of the layer name is a 2 character substring which describes the method of spatial data representation (points, lines, polygons, etc.). It occupies the eleventh and twelfth positions in the layer name. Data on Data Resource SItes are single feature class entities. As such, composite Core Database feature class types are not part of the name domain. The domain of the feature class component is as follows:

Feature Class Abbreviation	*Feature Class Name*	*Description*
py	polygon	Area features
ln	line	Line features
pt	point	Single point locations
ra	raster	Thematic raster data sets, such as a raster land cover data set
ms	measures	lines with measures
im	image	Non-thematic raster data set, such as a scanned map, or satellite image

4.1.1.2.3 Scale Class

The scale class identifier is a number from 1 to 4 which generally indicates the effective scale of the data. This may indicate the actual scale of the cartographic source, the effective resolution of the data, or be a commentary on the positional and/or thematic accuracy of the data. It occupies the thirteenth position in the layer name. The domain of the scale class identifier is as follows:

Scale Range	*Name*	*Scale Identifier*	*Application Type/Extent*	*Notes*
400K-2000K	State	1	Statewide mapping of summary data	The most generalized and least (positionally) accurate data
80K-400K	Planning	2	County Level Mapping County Level Planning Landscape Modeling	Very generalized. Not appropriate for locational measurement. Often includes simplified attribute schemes
10K-80K	Resource	3	Watershed Planning Landscape Modeling Timber Planning	Moderately (positionally) accurate data. Locational measurement can be performed with caution.
>10K	Site	4	Site-specific resource and facility management	The most (positionally) accurate data available

4.1.2 Layer Description

Layer descriptions are the so-called "long data names" assigned to layers. These are used in applications and narrative metadata as a more descriptive means of identifying layers. Like layer names, layer descriptions have certain constraints associated with them.

Constraint	*Specification*	*Discussion*
Reserved characters	Non-alphabetic characters are disallowed in layer names with the exception of the underscore "_", "(", ")", and "space" characters.	DNR's diverse storage and application environment will create many potential opportunities for conflict with non-alphabetic characters (e.g. <,>,&,@,%,* all have known conflicts). Parentheses have been commonly used in long names for several years without incident.
String length	60 wide characters	Historically, long names of up to 48 characters have been used without difficulty. This somewhat arbitrary limit is intended to improve readability and formatting on web-based documents.

Layer descriptions should be prepared with title style capitalization, but are otherwise unconstrained with regard to physical formatting outside of the constraints outlined above.

4.1.3 Coordinate Storage Specification

Application sites will conform to the existing DNR coordinate specifications. The standard DNR coordinate system is Extended UTM Zone 15 Unshifted Coordinates expressed in Meters. The standard datum is the North American Datum of 1983 (NAD83). The standard spheroid is GRS1980. All database elements should be stored in double precision coordinates.

4.1.4 Thematic Classification

Data are referenced in application environments and cataloged using discrete thematic categories as follows:

New Thematic Group	*Thematic Index Number*	*Example Layer Groups*
Control	21	Section Corner Locations (Control Point Inventory) USGS Quadrangle Indexes
Administrative Features	22	State Border County Borders Administrative Boundaries (region-area-district) State Legislative Districts Recreational Sites State Parks Wildlife Management Areas Populated Places County Seats
Public Land Survey	23	Public Land Survey
Ownership	24	State Land Ownership GAP Stewardship
Land Cover	25	Forest Inventory Remaining Natural Vegetation Land Use/Land Cover Presettlement Vegetation
Hydrography	26	Lakes/Basins/Wetlands Streams Watersheds
Topography	27	DEMs Shaded Relief Products Spot Elevations (not used specifically for control)

New Thematic Group	*Thematic Index Number*	*Example Layer Groups*
Geology and Soils	28	Soils ECS Units Geomorphology of Minnesota Bedrock Geology County Well Index Groundwater
Climate	29	Precipitation and Temperature
Transportation	30	Roads Railroad Lines
Facilities	31	Trails State Park Infrastructure DNR-Managed Roads Water and Sewer Field Service Resources (vehicles, tools, etc.)
Events and Sitings	32	Rare Natural Features Wildlife Sitings Tornado Paths Floods Wildfires Pollutant spills Fish Stockings Arrests Burn Histories Road Kills Depredation Reports Rare Natural Features
Plant and Animal Ranges	33	Range extents for specific plant and animal species
Cultural Resources	34	Cultural Sites (such as Presettlement sites, Homesteads, Trading Posts)
Environmental Quality	35	Environmental quality indicators expressed as isolines, or aggregated by some unit of geography
Demographics	36	TIGER-based data sets (Block Groups, Tracts, MCDs.)
Land Imagery	37	DOQs LandSat TM
Scanned Maps	38	DRGs Scanned Lake Maps
Metadata	39	Spatial index to data availability and quality characteristics

4.2 File Server Specifications for Data Storage

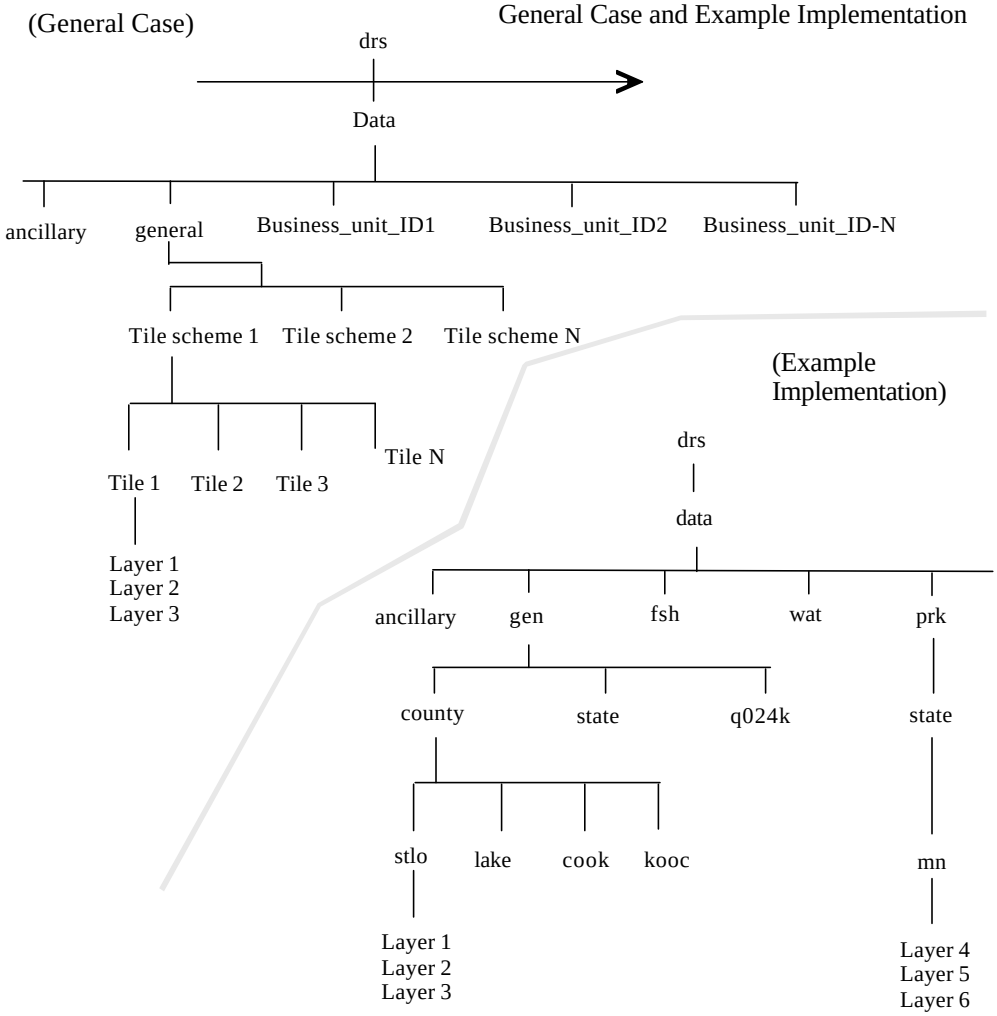
This section outlines the standard methods for organizing layer data on file servers (or local data

stores for stand-alone application environments). Structural directory hierarchies and naming standards are very critical in these types of environments, in that they form the basis for machine-based data access processes. These include: 1) Data file server directory hierarchy and associated naming standards, and 2) ARC/INFO librarian implementations and associated standards. Each is treated in separate subsections below.

4.2.1 File Server Directory Hierarchy

Data stored on resource sites will exist within a strict directory hierarchy. All resources sites will have a single root location on the file system. Subdirectories may be physically present or linked locations. An overview of the structure is provided in Figure 2.

Figure 2: File Server Directory Standard:
General Case and Example Implementation



The directory structure reflects a geographic organization scheme where data sets are organized into defined spatial extents. This basic approach is much the same as the current Core Database implementation. There are three principal reasons to retain this configuration: 1) ARC/INFO Librarian data structures require it, 2) the DNR user base has grown accustomed to it, and 3) DNR staff continue to operate within a geographic, or "place' context.

The only real structural difference between this approach and the existing Core Database approach is the introduction of a business unit tier between the tiling scheme level and the data "type" level . The reason for this addition is two-fold. First, the derived product paradigm existing at the heart of the Revised Core Database Architecture will result in significant expansion in the number of layers administered on the various sites. This extra tier will help to reduce the total number of files in some directories, preventing the performance impacts associated with large numbers of files in single directories. The second reason for the addition is the need to isolate subtly different data layers from each other, a condition that will likely arise as data are tailored to business-specific applications. These business products will be isolated within specific portions of the file system. All applications will draw upon a general source for some base level data (counties, regions, roads, etc.).

Each of the various levels in the scheme have their own standards associated with them. These are described in their own subsections. The reader may note that the first two levels of the hierarchy were describe in Section 3.1.

4.2.1.1 Business Unit ID Specifications

Business Unit ID's are text strings representing DNR administrative units that are used to segregate portions of file systems below the Resource Type level in the Data Resource Site hierarchy. Assigning specific portions of the file system to business units will help to encapsulate data for use in applications.

Business Unit ID's are assigned to the Section level within DNR Divisions, and a level for DNR Bureaus. The domain of the ID's is as follows:

Business Unit ID	*Business Unit Name*
com	Commissioner's Office
wld	Division of Wildlife
fsh	Division of Fisheries
eco	Division of Ecological Services
enf	Enforcement

fos	Field Operations and Support
for	Forestry
hrs	Human Resource Bureau
iel	Bureau of Information, Education and Licensing
omb	Office of Management and Budget
lam	Lands and Minerals
mis	Management Information Services Bureau
prk	Parks and Recreation
wat	Waters
taw	Trails and Waterways
gen	General data storage area for common data elements (i.e. base data, and general application data)

4.2.1.2 Tiling Scheme ID

The Tiling Scheme ID level in the data storage hierarchy describes the means by which data are physically partitioned. This level has a fixed domain of tiling scheme types, as presented in the following table.

Tiling Scheme ID	*Tiling Scheme Name*	*Notes/Discussion*
county	County	Commonly used for county-based mapping. Source index may be subject to update as more accurate boundary information is obtained.
Township	Township	Used primarily with resource scale data (See Section 2).
q250k	250K Index	Tiling scheme based on a 1 degree by 2 degree grid defined by the U.S. Geological Survey. Basis for the USGS 1:250,000 scale map series.
q100k	100K Index	Tiling scheme based on a ½ degree by 1 degree grid defined by the U.S. Geological Survey. Basis for the USGS 1:100,000 scale map series.
q024k	24K Index	Tiling scheme based on a 7 ½ minute by 7 ½ minute grid defined by the U.S. Geological Survey. Basis for the USGS 1:24,000 scale map series.
majwshed	Major Watershed	Tiling scheme based on 1:24,000 scale watershed delineations maintained through the DNR Division of Waters.
Leg_dist	Legislative	Legislative Districts from 1990 Census (TIGER

Tiling Scheme ID	*Tiling Scheme Name*	*Notes/Discussion*
	District	data)
region	DNR Region	DNR Regions based on a combination of the standard County coverage and selected road features
ecs_sub	Ecological Classification System Subsection	ECS Subsections from master core version.
mndotms	MnDOT Highway Map Sheet	County and sub-county tiling scheme institutionalized by the Minnesota DOT Trunk Highway map series.
state	Statewide	Statewide data extent
forarea	Division of Forestry Area	Division of Forestry administrative areas
dowlknum	DOW Basin	Basin delineation extents developed by Ecological Services Lake Mapping program. May include multiple shapes per tile

4.2.1.3 Tile Ids

Each tiling scheme list above has a specific domain associated with it. These are describe in a series of subsections below.

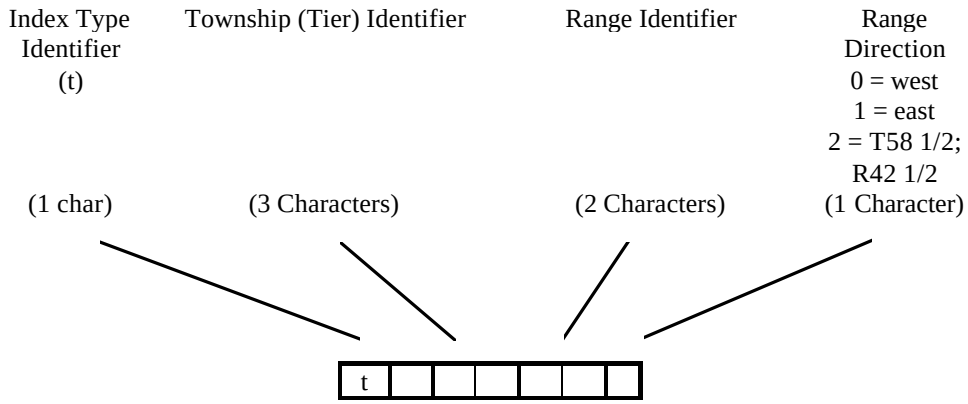
4.2.1.3.1 County Tile Identifiers

Tile Name	*Name of County*
aitk	Aitkin
anok	Anoka
beck	Becker
belt	Beltrami
bent	Benton
bigs	Big Stone
blue	Blue Earth
brow	Brown
carl	Carlton
carv	Carver
cass	Cass
chip	Chippewa
chis	Chisago
clay	Clay
clea	Clearwater
cook	Cook
cott	Cottonwood
crow	Crow Wing
dako	Dakota
dodg	Dodge
doug	Douglas
fari	Faribault
fill	Fillmore
free	Freeborn
good	Goodhue
gran	Grant
henn	Hennepin
hous	Houston
hubb	Hubbard
isan	Isanti
itas	Itasca
jack	Jackson
kana	Kanabec
kand	Kandiyohi
kitt	Kittson
kooc	Koochiching
lacq	Lac Qui Parle
lake	Lake
lotw	Lake of the Woods
lesu	Le Sueur
linc	Lincoln
lyon	Lyon
mcle	McLeod
mahn	Mahnomen
mars	Marshall

Tile Name	*Name of County*
mart	Martin
meek	Meeker
mill	Mille Lacs
morr	Morrison
mowe	Mower
murr	Murray
nico	Nicollet
nobl	Nobles
norm	Norman
olms	Olmsted
otte	Otter Tail
penn	Pennington
pine	Pine
pipe	Pipestone
polk	Polk
pope	Pope
rams	Ramsey
redl	Red Lake
redw	Redwood
renv	Renville
rice	Rice
rock	Rock
rose	Roseau
stlo	St. Louis
scot	Scott
sher	Sherburne
sibl	Sibley
stea	Stearns
stee	Steele
stev	Stevens
swif	Swift
todd	Todd
trav	Traverse
waba	Wabasha
wade	Wadena
wase	Waseca
wash	Washington
wato	Watonwan
wilk	Wilkin
wino	Winona
wrig	Wright
yell	Yellow Medicine

4.2.1.3.2 Township Tile Identifiers

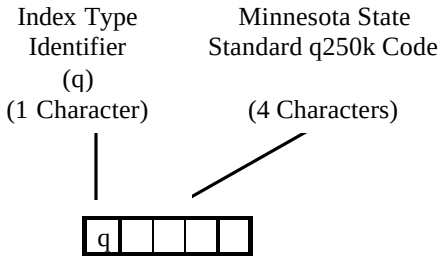
Township identifiers conform to the following standard:



Examples:

Township 102, Range 12 West would be specified as: t102120
Township 61, Range 1 East (Cook County) would be specified as: t061011

4.2.1.3.3 USGS 250K Tile Identifiers

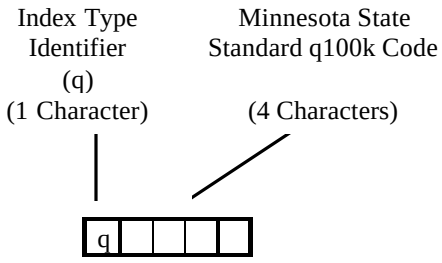


Examples:

A tile identifier for the quad with the Minnesota State q250k code of “1126” is: q1126
A tile identifier for the quad with the Minnesota State q250k code of “326” is: q0326

4.2.1.3.4 USGS 100K Tile Identifiers

100K tile identifiers conform to the following standard:

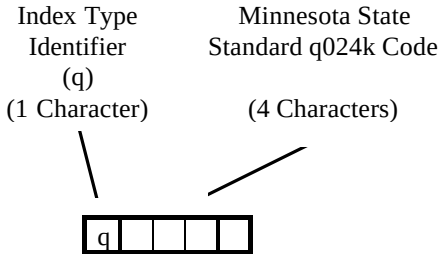


Examples:

A tile identifier for the quad with the Minnesota State q100k code of “2342” is: q2342
A tile identifier for the quad with the Minnesota State q100k code of “318” is: q0318

4.2.1.3.5 USGS 24K Tile Identifiers

24K tile identifiers conform to the following standard:

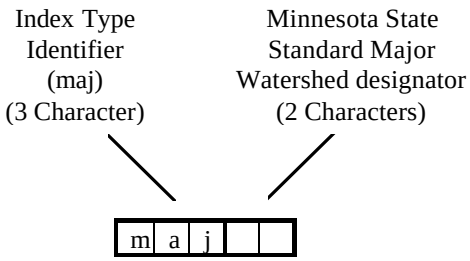


Examples:

A tile identifier for the quad with the Minnesota State q024k code of “1011” is: q1011
A tile identifier for the quad with the Minnesota State q024k code of “118” is: q0118

4.2.1.3.6 Major Watersheds

Major watershed tile identifiers conform to the following standard:

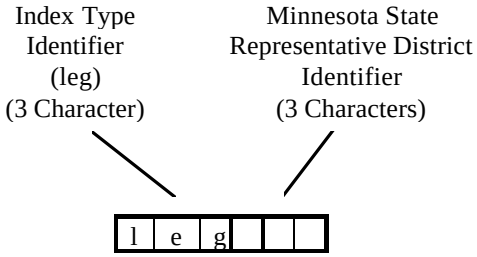


Examples:

A tile identifier for the Minnesota Major Watershed code of “38” is: maj38
A tile identifier for the Minnesota Major Watershed code of “1” is: maj01

4.2.1.3.7 Legislative District

Legislative district tile identifiers conform to the following standard:

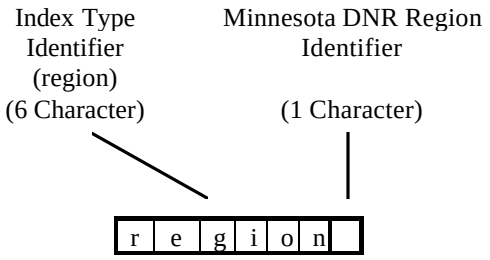


Examples:

A tile identifier for the Minnesota Representative District “1a” is: leg01a
A tile identifier for the Minnesota Representative District “10a” is: leg10a

4.2.1.3.8 Region

Region tile identifiers conform to the following standard:

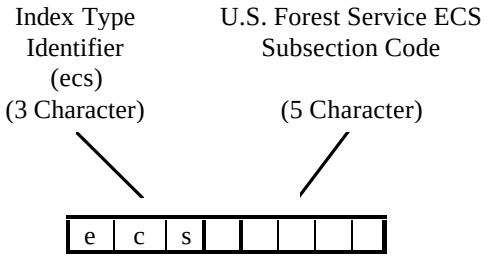


Examples:

A tile identifier for Minnesota Region 1 is: region1

4.2.1.3.9 Ecological Classification System Subsection

ECS Subsection tile identifiers conform to the following standard:



Examples:

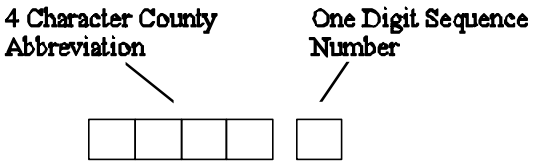
A tile identifier for the Red River Prairie (Code 251Aa) is: ecs251aa

Valid domains are defined in the Minnesota DNR Geocoding standard for ECS Subsections.

4.2.1.3.10 MnDOT Map Sheet Identifier (mndotms)

MnDOT map sheets are a statewide scheme which divides large counties into multiple tiles while retaining small counties as single tiles, in effect, creating a roughly normalized “county-like” scheme.

MnDOT Map Sheet identifiers consist of the four character county abbreviation, followed by the map sheet sequence number, as follows:



Example: Hennepin County Sheet 1 would be "henn1"

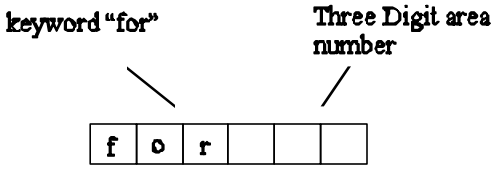
4.2.1.3.11 State

The state tiling scheme simply fits the 2 character postal code for each state in lower case format, and includes: mn, wi, nd, sd, ia

4.2.1.3.12 Forarea

This tiling scheme exists to support data tiled by Division of Forestry administrative area.

Forarea tile identifiers conform to the following standard:



Example: Aitkin Forestry Area would be "for334"

Domain values for this element are as follows:

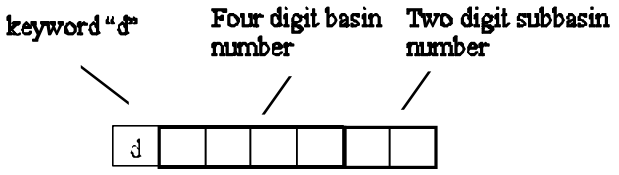
for111	Bemidji Area
for116	Bagley Area
for117	Blackduck Area
for121	Warroad Area
for123	Wannaska Area
for131	Baudette Area
for161	Park Rapids Area

for162	Alexandria Area
for163	Detroit Lakes Area
for221	Deer River Area
for222	Effie Area
for234	Hibbing Area
for241	Orr Area
for245	Tower Area
for251	Cloquet Area
for253	Two Harbors Area
for255	Grand Marais Area
for261	Littlefork Area
for311	Brainerd Area
for311	Brainerd Area
for312	Little Falls Area
for312	Little Falls Area
for312	Little Falls Area
for321	Backus Area
for323	Pequot Lakes Area
for331	Hill City Area
for334	Aitkin Area
for342	Moose Lake Area
for344	Hinckley Area
for351	Cambridge Area
for353	St. Cloud Area
for442	Mankato Area
for443	New Ulm Area
for444	Willmar Area
for531	Lewiston Area
for532	Caledonia Area
for533	Preston Area
for534	Lake City Area
for541	Rochester Area
for545	Faribault Area
for611	North Metro Area
for612	East Metro Area
for613	West Metro Area
for999	Red Lake Reservation

4.2.1.3.12 dowlknum

This tiling scheme exists to support lake survey developed by the Division of Ecological Services Lake Mapping program.

Dowlknum tile identifiers conform to the following standard:



Example: Aitkin Forestry Area would be "for334"

4.2.1.4 Layer Names

The Layer Names level in the data storage hierarchy uses the layer naming conventions outlined in Section 3 of this document. Shapefiles and image file types use the naming standard directly by adding the extensions appropriate for their file type (e.g., .shp, or .tif). ARC/INFO Librarian layers represent a special case, and are treated in the following section.

4.2.2 ARC/INFO Librarian Implementations

A key LIBRARIAN implementation issue relates to the placement of the library database directory on the file system. These will be placed in a standard, consolidated location under the metadata section of the Data Resource Site file system. The rationale for separating these metadata from the tile level of the hierarchy (where they reside in the old Core Database schema), is to maintain the purity of the data stores. At present, all applications that interact with the tile level of the directory hierarchy must account for, and ignore the LIBRARIAN "database" directory. It would be preferable to remove this exception from that environment. Consolidating these directories also allows system administration staff to reference all libraries with a single environment variable. The general LIBRARIAN schema is depicted graphically in Figure 3.

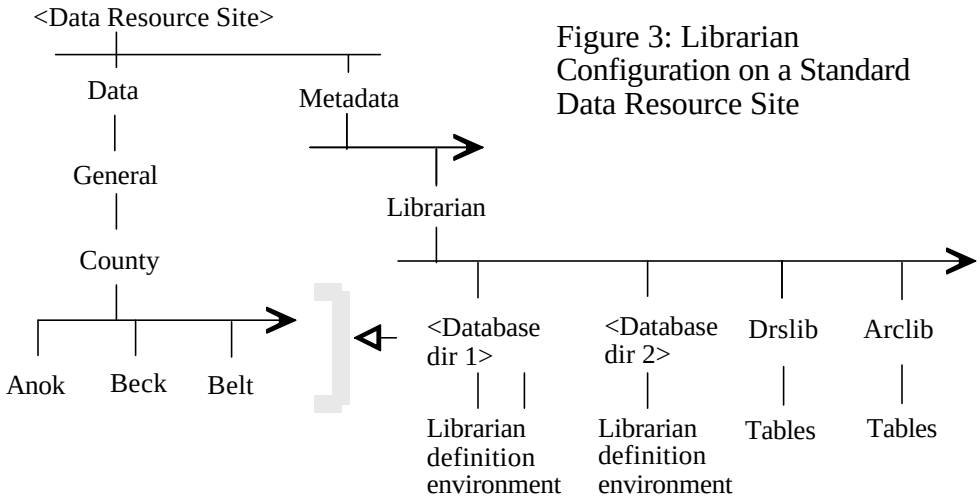


Figure 3: Librarian Configuration on a Standard Data Resource Site

Specifications are also required for library names and LIBRARIAN database directory names (at ../metadata/librarian locations in Figure 3). Library names will conform to the following specification:

<business unit id>_<tiling scheme id>

Examples:
fsh_county, gen_q024k

Librarian database directory names will follow the identical name specification.

The <Data Resource Site>/metadata/librarian location also includes instances of the libraries file registry (normally found under $ARCHOME/tables). The drslib directory contains an info directory named “tables” which stores a table named LIBRARIES which is a registry of librarian instances available on the Data Resource Site. This is a LIBRARIES registry instance which is built around the site. Desktop applications should reference this location as the user $ARCHOME location (ARCHOME=<Data Resource Site>/metadata/librarian/drslib). The arclib location is a symbolic link to the UNIX-based ARC/INFO location, and should not (or may not) be referenced by regular desktop applications.

4.2.3 Image Catalogs

Like librarian metafiles, image catalogs are stored within the metadata location. In this case, they

are all stored within a single directory location:

<Data Resource Site>/metadata/image_cats/info

Image catalogs are named for the layer they reference. Since these are INFO files, they are stored in ALL CAPS. Examples include:

MAP_DRG024IM3
IMG_DOQ03IM4

4.3 Database Server Specifications

Database servers within the Revised Core environment only exist at the Central Office. Therefore, no site-specific implementation standards are required, except in reference to the data provided by that environment. The primary distinction to be made is between data served by SDE database server technology and that provided by native Oracle. These are treated in separate subsections.

4.3.1 Data Retrieved from SDE Database Servers {NOTE: this section not deployed at this time}

Five information elements are needed to retrieve data from SDE. There are: 1) server name, 2) instance name, 3) layer name, 4) user name, and 5) password. The latter two elements are discussed in Appendix 2 (System Security Configuration) and Section 5.1.3 of this document. Server names are controlled through the system-wide data dictionary. Instance names will conform to the following specification:

sde_<server type>_<business unit ID>

Examples:
sde_ora_eco (An SDE for Oracle server instance supporting Ecological Services)
sde_cov_gen (An SDE for Coverages server instance providing general data access)

Layer names are controlled by the SDE server and will follow the specifications provided in Section 3.1 of this document, except that the server requires a second term be appended to the layer name stating the "spatial column"[3] of the feature class being presented. Therefore, SDE layer names will conform to the following format:

<dnr layer name>.<spatial column>

Examples:
lake_pwipy3.shape#
wet_circ39py3.cir39py3#

[3] A spatial column is a table column that uniquely identifies a feature. In tiled data sets, it provides the ability to "knit" together spatial objects that are split by tile boundaries

4.3.2 Data Retrieved from Oracle Database Servers {NOTE: this section not deployed at this time}

DNR's traditional tabular database environment is not supported by data structuring standards (for table names, view names, server instance names). However, tabular information is an important part of the application environment supported by data resource sites. Tabular data access will be managed through the use of Windows NT ODBC protocols, and will therefore require some level of client configuration and setup (driver installation, ODBC source registration). Tabular data access requires several information elements, including: 1) ODBC source, 2) User name, 3) Password, and 4) table name. <No specific standards have been developed at this time that would affect any of these elements. This part of the specification is still under development.>

<metadata structures needed to support table access are not currently developed. The Oracle data dictionary will likely provide this information>

5.0 System Level Metadata

System level metadata is the Data Resource Site center of activity. This section of the Resource Site file system contains a variety of descriptive information that instructs applications, and facilitates site administration. It provides a roadmap of the resources available through the site.

Three types of metadata elements are described in this section of the file system, including: 1) data definition, 2) narrative metadata, and 3) ARC/INFO LIBRARIAN references. The directory structure that supports these data elements is described in Figure 4 and discussed below.

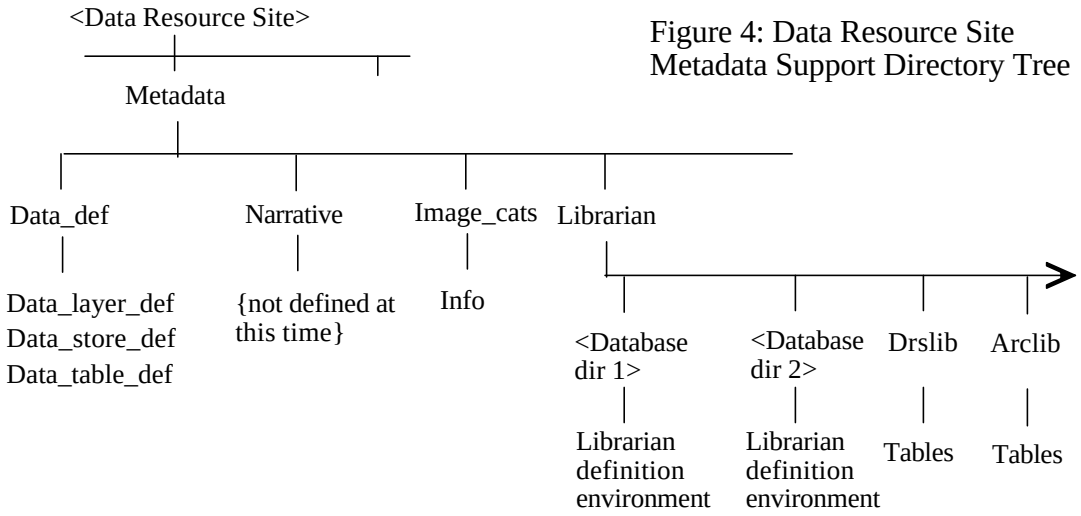


Figure 4: Data Resource Site Metadata Support Directory Tree

5.1 Data Definition

The Data Definition section of the Data Resource Site exists to define the suite of data that are available to users on a particular site and the methods needed to access the data in the desktop environment. Various registry types are described below which store information about particular data types.

5.1.1 Data Layer Registry

The Data Layer Registry is a comprehensive list of the GIS data layers available on a Data Resource Site. It's physical name is "data_layer_def". It is an ASCII format, comma delimited file that lists a standard set of elements that define basic data layer characteristics. It uses the keyword "null" as a placeholder for null elements, and single quote character " ' " as a comment line indicator in the first position of the record.

The formal record layout is as follows:

```
[layer_desc],[DNR_Unit],[tiling_scheme],[product_type],[resource_description],
[vendor_object_type],[server],[instance],[derived product id], [thematic class
id]
```

- *layer_desc* is the so-called "long name" of the layer describe in Section 4.1.2 of this document
- *DNR_unit* is the DNR Business Unit ID described in Section 4.2.11 of this document
- *tiling_scheme* is the standard tiling scheme ID described in Section 4.2.1.2 of this document
- *product_type* is Data Dictionary element that describes the physical format in which the data are made available. It is a numeric code with the following domain:

  1 - coverage
  2 - shapefile
  3 - sde
  4 - geodatabase
  5 - region
  6 - grid
  7 - tin
  8 - route
  9 - tif
  10 - jpeg
  11 - oracle
  12 - info
  13 - dbase
  14 - eppl

15 - db2
16 - text
17 - photograph
18 - mrsid
19 - erdas_imagine
20 - erdas_gis
21 - ARC/INFO librarian

- *resource_description* is the 13 character layer name described in Section 4.1.2 of this document
- *vendor_object_type* is the feature class of the data. It's domain as of this writing is:

polygon
point
line
raster

- *server* is the name of the database server used to serve SDE data types
- *instance* is either: 1) the name of the SDE instance used to serve the data, or 2) the name of the librarian database that the data are stored within. SDE instances conform to the rules described in Section 4.3.1 of this document. Librarian database instances conform to the rules described in Section 4.2.2.
- *derived product id* is the unique derived product identifier, composed in the data dictionary of the MetaProductID, ProductTypeID, and FeatureClassID columns. This is used to query the Data Dictionary using the JSP-based intranet metadata reports.
- *thematic class id* is the numeric theme class identifier (Section 4.1.4)

Data Layer Registry records example:

```
Hybrid County Boundaries,gen,state,2,bdry_counpy3,polygon,null,null,220000030301,22
24K Quadrangle Index,gen,state,2,indx_q024kpy4,polygon,null,null,2100000202011,21
```

5.1.2 Data Table Registry {NOTE: this section not deployed at this time}

The Data Table Registry functions much the same way as the Data Layer Registry, except that it provides a comprehensive listing of available tables. Within DNR's enterprise environment, all registered data tables come from centralized database servers.

The Data Table Registry's physical name is "data_table_def". It is an ASCII format, comma delimited file that lists a standard set of elements that define basic data table characteristics. It uses the keyword "null" as a placeholder for null elements, and single quote character " ' " as a comment line indicator in the first position of the record.

The formal record layout is as follows:

<not defined yet>

5.1.3 Database Source Registry

The Database Source Registry provides user name and password information for centralized data server access. It's physical name is "data_store_def". It is an ASCII format, comma delimited file that lists a standard set of elements that define basic data table characteristics. It uses the single quote character " ' " as a comment line indicator in the first position of the record. It is related to the Data Table and Data Layer Registries by the "instance" element in each table, which serves as the primary key in the table.

The formal record layout is as follows:

[instance],[server],[user name],[password]

Example record:

```
sde_ora_gen,pinchot,reg2_user,sde
```

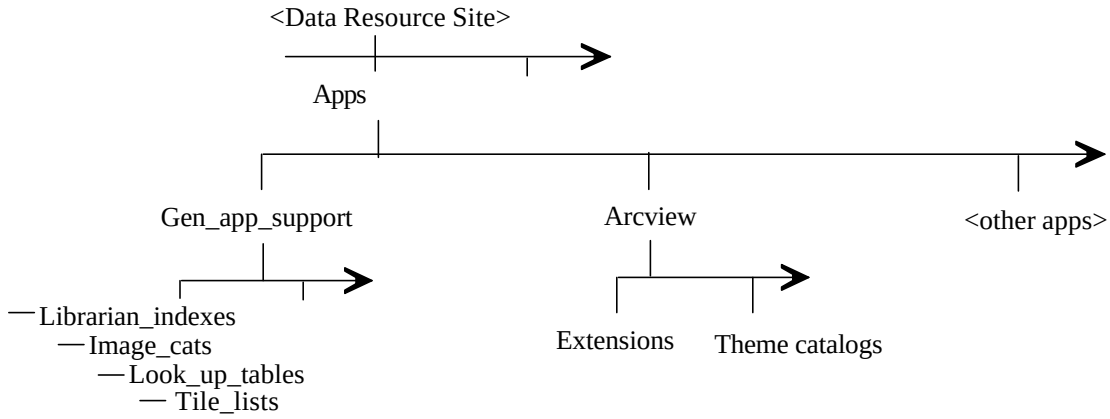
5.2 Narrative Metadata {This section has been withdrawn}

5.3 ARC/INFO Librarian References

The ARC/INFO Librarian Reference Environment is described in Section 4.2.2 and will not be repeated here.

6.0 Application Components

Data Resource Sites house a variety of application software, including programs and related supporting data, and software extensions (i.e. for ARCVIEW). This part of the filesystem supports both UNIX and NT client environments. The general layout of this part of the Data Resource Site filesystem is presented below.



The Application section of a Data Resource Site consolidates fileserver-based application software into a single location, facilitating distribution of updates, enhancements, and new additions. Common program support elements are also consolidated under a gen_app_support location.

Gen_app_support includes sets of generic, static metadata that support a variety of application functions. As such, they are part of the DRS specification, and require some level of description. The following elements are described in separate subsections below: 1) drs_theme_list, 2) lib_tmplt, 3) librarian_indexes, 4) look_up_tables, 5) search_indexes, and 6) tile_lists.

6.1 drs_theme_list

The drs_theme_list is a single comma-delimited ASCII-formatted text file that cross references theme categories with theme numbers. It is used by applications that need to associated theme numbers (see Section 4.1.4) from the DRS data_layer_def file with textual thematic class descriptions. A sample set of records is as follows:

```
21,Control
22,Administrative Features
23,Public Land Survey
24,Ownership
```

6.2 lib_tmplt

The lib_tmpt section includes a set of templates that are used to create librarian “database” directories. There is one template per tiling scheme. Each template is a system directory

containing a librarian index coverage and INFO workspaces and files to hold layer and access (permissions) information. These are used during the site update process to build new librarian data structures that reflect the content of each individual DRS.

6.3 librarian_indexes

The librarian_indexes are the master tile index coverages. These are the ARC/INFO index coverage sources for the lib_tmplt location and also are used for status file generation in QuickThemes.

6.4 look_up_tables

look_up_tables contains an INFO workspace that includes a set of files that cross-references common attribute codes with their meanings. These are added to support applications on an on-going basis. At the time of this writing, eight look-up tables are included here:

QQ024K.LUT
COUNTY.LUT
Q024K.LUT
Q100K.LUT
MAJWSHED.LUT
Q250K.LUT
REGION.LUT
dowlknum.lut

These tables follow the format:

```
COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE NAME     INDEXED?
    1  TILE-NAME            12    12     C      -                          -
   13  LONG-NAME            40    40     C      -                          -
```

6.5 search_indexes

search_indexes contains an INFO directory, which in turn contains a set of files that cross list tiles against the county tiling scheme. Stated another way, these are lists of tiles paired with county tiles that they overlap. These files are used to determine which tiles in a particular tiling scheme intersect particular county areas. They are essentially the tablular product of an intersection overlay between a particular tiling scheme and the county scheme. They are used in the site update process to determine which tiles of a scheme are needed to support a particular area of interest (AOI) expressed in county extents.

A search index is named for the tiling scheme with the string "_XRF" appended to it (e.g. Q100K_XRF). The conform to the following format:

```
COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE NAME     INDEXED?
    1  CASE#                 4     5     B      -                          -
    5  FREQUENCY             4     5     B      -                          -
    9  COUNTY               32    32     C      -                          -
   41 <tile Scheme Name>    32    32     C      -                          -
```

An example set of records are:

```
CASE#                    =   273
FREQUENCY                =     1
COUNTY                   = stea
Q100K                    = q3518
          274
CASE#                    =   274
FREQUENCY                =     1
COUNTY                   = stea
Q100K                    = q3526
          275
CASE#                    =   275
FREQUENCY                =     1
COUNTY                   = stee
Q100K                    = q4334
          276
CASE#                    =   276
FREQUENCY                =     1
COUNTY                   = stee
Q100K                    = q4734
          277
CASE#                    =   277
FREQUENCY                =     1
COUNTY                   = stev
Q100K                    = q3110
```

6.6 tile_lists

The tile_lists location contains the master lists of available tiles within each tiling scheme. These are ASCII formatted text files, one file per tiling scheme that list the available tiles in lower case strings. Tile list names follow the format:

```
mast_<tile scheme name>_list
```

For example:

```
mast_q100k_list
```

example contents (for q100k) are as follows:

```
q0318
q0326
q0702
q0710
q0718
q0726
q0734
q0742
q1102
q1110
q1118
q1126
```

...

7.0 Process

Data Resource Sites will be the subject of many automated processes to facilitate maintenance and access. The "process" subdirectory below the resource site root location provides a place to execute programs, write out log files, and create temporary data. The principle reason for creating this location is to keep that portion of the file system used to store data completely clean. There has been a tendency to introduce temporary files, programs, and backup directories into Core data locations. This potentially compromises the integrity of applications which move data between locations, report data availability and status, and determine data volumes. Consolidating temporary files in a single location also facilitates site clean-up. It should be noted that data that are accessed regularly by applications should be stored in the Application section of the site.

8.0 Wrap-up

The Data Resource Site concept reflects a philosophy of distributed data access which will likely be rendered obsolete within a four to six year period. The first step in this process will be the elimination of most of the metadata elements from the sites in favor of centralized access of these descriptive components. Data resources themselves will also migrate to centralized locations through time (from file servers to database server environments). With a solid distributed architecture, this migration from local to remote access (distributed to centralized) can be accomplished in a gradual, orderly fashion when improvements in network infrastructure (bandwidth) become available.

Appendix 1: Operating Environment Baselines

Minnesota DNR Revised Core Database Architecture

Access Site Specification Notes

1.11 Access Site Specification Elements

- Metastore specification (including data model and physical format)
- Layer naming
- File system organization
- Supported data types

2.Operating Environment Description
- 2.1 Operating systems
  - 2.1.1 Windows NT/95/98/2000
  - 2.1.2 Solaris 7.X
  - 2.1.3 AS/400
- 2.2 Hardware
  - 2.2.1 Sun Microsystems Workstations/Servers
  - 2.2.2 Intel workstations
    - 300 MHz Pentium
    - 64 Mb RAM
    - 4 Mb graphics card
    - Microsoft-compatible CD-ROM reader
- 2.3 Client application environment
  - 2.3.1 ARCVIEW (Windows/UNIX)
  - 2.3.2 VB-Map Objects
  - 2.3.3 Workstation ARC/INFO (UNIX/Windows)
  - 2.3.4 ARC/INFO 8 Desktop
- 2.4 Database administration environment
  - 2.4.1 Workstation ARC/INFO (UNIX)
  - 2.4.2 Oracle 8i
  - 2.4.3 UDB
  - 2.4.4 SDE
  - 2.4.5 pkzip-unzip compression (Windows/UNIX)
- 2.5 Application Server Environment
  - 2.5.1 MapServer
  - 2.5.2 ARC IMS

3.Operational Requirements
- 3.1 Integration with data dictionary
- 3.2 Optimize mirroring/site update processes
- 3.3 Optimize for facilitated desktop access
- 3.4 Suitability for ad hoc desktop access (file system browsing, appearance in add themes view window

Access Site Specification Notes (cont.)

4.Filesystem Storage Standards
- 4.1 Layer naming elements (potential)-filesystem, librarian, and SDE types
  - Layer abbreviation/description
  - Feature class
  - Scale class
  - Native tiling scheme

4.2 Directory organization (issues and potential approaches)
- Thematic/spatial hybrid
- Top level thematic organization (library) or (possibly) top-level discipline organization (fisheries, forestry, etc); supplemented with a single general access area
- Sub-levels oriented around geography (core-like tiling scheme ID) with tile sub-directories

4.3 Tile specifications (should remain intact from current core specs, except possibly township, which has always been anomalous).