USDA Service Center Agencies
Geospatial Data Management Team
Data Management Plan For
National Land Cover Dataset (NLCD) 2001 Data
Revised March 2009
Ken Becker; Elaine Ortiz

I. Purpose and Scope (business case)

A. Purpose

The National Land Cover Database 2001 land cover layer was produced through a cooperative project conducted by the Multi-Resolution Land Characteristics (MRLC) Consortium. The MRLC Consortium is a partnership of federal agencies (www.mrlc.gov), consisting of the U.S. Geological Survey (USGS), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Environmental Protection Agency (EPA), the U.S. Department of Agriculture (USDA), the U.S. Forest Service (USFS), the National Park Service (NPS), the U.S. Fish and Wildlife Service (FWS), the Bureau of Land Management (BLM) and the USDA Natural Resources Conservation Service (NRCS). One of the primary goals is to generate a current, consistent, seamless, and accurate National Land cover Database (NLCD) circa 2001 for the United States at medium spatial resolution. 2001 refers to the nominal year from which most of the Landsat 5 and Landsat 7 imagery was acquired. This landcover map and all documents pertaining to it are considered "provisional" until a formal accuracy assessment can be conducted. For a detailed definition and discussion on MRLC and the NLCD 2001 products, refer to Homer et al. (2004) - http://www.mrlc.gov/pdfs/July_PERS.pdf ; and http://www.mrlc.gov/mrlc2k.asp.

The NLCD 2001was created by partitioning the U.S. into mapping zones. A total of 66 mapping zones were delineated within the conterminous U.S. based on ecoregion and geographical characteristics, edge matching features and the size requirement of Landsat mosaics. The completed single pixel product was generalized to a 1 acre minimum mapping unit product.

B. Scope

The data extent is currently the 48 conterminous states plus Alaska coverage.

See the status map at: http://www.mrlc.gov/mrlc2k_nlcd_map.asp for data availability.

II. Acquisition

A. Data Source

1. Producer Information
   a. Name

      Multi-Resolution Land Characteristics (MRLC) Consortium - USGS, EPA, NOAA, USFS, BLM, NASA, NPS, NRCS, USFWS, OSM

   b. Location of Headquarters

      Consortium Lead: U.S. Geological Survey
      USGS/EROS

47914 252nd Street
Sioux Falls, SD 57198-0001

c. Internet Address

http://www.mrlc.gov/index.asp

2. Publisher Information

a. Name

Multi-Resolution Land Characteristics (MRLC) Consortium - USGS, EPA, NOAA, USFS, BLM, NASA, NPS, NRCS, USFWS, OSM

b. Location of Headquarters

Consortium Lead: U.S. Geological Survey
USGS/EROS
47914 252nd Street
Sioux Falls, SD 57198-0001

c. Internet Address

http://www.mrlc.gov/index.asp

3. Acquisition Information

d. Delivery Media

FTP download

e. Download URL

http://www.mrlc.gov/index.asp

f. Projected Data Availability Schedule

Currently available is for the 48 conterminous states and Alaska; for the remainder of Hawaii and Puerto Rico – projected availability is December 2008.

B. Standards Information

1. Geospatial Data Standard

a. Standard Name and Steward Information

NLCD 2001 was generated according to method protocols outlined in Homer et al. (2004) using 65 mapping zones for the conterminous United States. Production occurred across 12 mapping teams from both the government and private sector. In order to ensure product consistency among teams, products were generated using a standardized process spanning data preparation, classification and quality control. USGS EROS was responsible for oversight of product development, data preparation, classification training, quality control and product synthesis. A brief overview of standard methods:

Source Data Preparation

All NLCD 2001 products were generated from a standardized set of data layers mosaiced by mapping zone. Typical zonal layers included multi-season Landsat 5 and Landsat 7 imagery centered on a nominal collection year of 2001, and Digital Elevation Model based derivatives (Figure 1). This standard set of mosaiced zonal layer stacks

often consisted of 18 or more layers that provided the best available data resources to derive the desired products. Application of layers could vary slightly, as mapping protocols were shifted to meet unique regional conditions. All data were geo-registered to Albers equal area projection grid, and resampled to 30m grid cells.

Land Cover Classification

The land cover classification was accomplished using commercial decision tree (DT) software called See5* (Quinlan 1993) applied to zonal layer stacks prepared for each mapping zone. In addition to the See5 software, an interface was created for ERDAS IMAGINE to extrapolate derived DT models into classified pixels. DT is a supervised classification method that relies on large amounts of training data, which was initially collected from a variety of sources including high-resolution orthoimagery, local datasets, field collected points, and Forest Inventory Analysis (FIA) plot data. In many mapping zones, training data collection took advantage of existing regional land cover maps such as NLCD 1992, Gap Analysis Program (GAP), and National Agricultural Statistics Service (NASS) cropland data to improve classification efficiency. Predictions from these multiple products were compared to find areas of agreement, then spatially sampled to avoid transitional or edge pixels to generate training samples randomly across classes in proportion to their population. The result produced evenly distributed training data allowing for optimization of the DT models. Training data were used to map all land cover classes except for the four urban classes which were derived from thresholding of the imperviousness data product. Once an initial classification was completed, typically a number of subsequent DT iterations were necessary to improve the classification result. A series of scripts specifically written for this project were employed to gauge success and make adjustments to the See5 data file as required to generate an acceptable map. Once the product had evolved as far as DT methods could take it, additional localized modeling and hand-editing were typically required to produce the final product. However, localized modeling and hand-editing affected proportionately few pixels.

b. Standard Version

None

c. Standard URL

None

2. Metadata Standard

a. Standard Name and Steward Information

Metadata are compliant with:
Federal Geographic Data Committee (FGDC) Content Standard for Digital Geographic Metadata FGDC STD-001-1998.

The NLCD is documented using the Federal Geographic Data Committee (FGDC) content standard for geospatial metadata, and will be kept in conformance with a future ISO metadata standard. NLCD will be offered using the Spatial Data Transfer Standard and in particular, the raster profile, an FGDC endorsed (FGDC-STD-002.5) profile of the ANSI and FGDC endorsed SDTS (ANSI NCITS 320:1998, FGDC-STD-002, FIPS 173-1). NLCD is provided in formats that are fully compliant with Executive Order 12906, OMB Circular A130, and OMB Circular A-119.

b. Description of Metadata Captured

The Federal Geographic Data Committee's content standard for digital geospatial metadata (http://www.fgdc.gov/metadata/contstan.html) is used to document NLCD data.

c. Metadata Accuracy and Completeness Assessment

The metadata is complete.

C. Acquired Data Structure

1. Geospatial Data Format

a. Format (raster, vector, etc.)

Raster

b. Format Name

Raster IMG Imagine.

c. Data Extent

The data extent is currently the 48 conterminous states plus Alaska coverage.

See the status map at: http://www.mrlc.gov/mrlc2k_nlcd_map.asp for data availability.

d. Horizontal and Vertical Resolution

30 meters by 30 meters

e. Absolute Horizontal and Vertical Accuracy

Horizontal - 50 meters root mean squared error overall

f. Nominal Scale

1:100,000

g. Horizontal and Vertical Datum

Projection type: Albers conical equal area
Spheroid name: GRS 1980
Datum name: NAD83 – Conterminous US
Datum name: WGS_1984 – Alaska

h. Projection

Geographic

i. Coordinate Units

Meters

j. Average Data Set Size

The full dataset is

k. Symbology

Each image has Color Indexes that correspond to the particular land cover.

2. Attribute Data Format

a. Format Name

Raster

b. Database Size

N/A

3. Data Model

a. Geospatial Data Structure

IMG Imagine.

b. Attribute Data Structure

Raster data –N/A

Strata

11. Open Water - All areas of open water, generally with less than 25% cover of vegetation or soil.

12. Perennial Ice/Snow - All areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.

21. Developed, Open Space - Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes

22. Developed, Low Intensity - Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family housing units.

23. Developed, Medium Intensity - Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 percent of the total cover. These areas most commonly include single-family housing units.

24. Developed, High Intensity - Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80 to100 percent of the total cover.

31. Barren Land (Rock/Sand/Clay) - Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.

32. Unconsolidated Shore* - Unconsolidated material such as silt, sand, or gravel that is subject to inundation and redistribution due to the action of water. Characterized by substrates lacking vegetation except for pioneering plants that become established during brief periods when growing conditions are favorable. Erosion and deposition by waves and currents produce a number of landforms representing this class.

41. Deciduous Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.

42. Evergreen Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.

43. Mixed Forest - Areas dominated by trees generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.

51. Dwarf Scrub - Alaska only areas dominated by shrubs less than 20 centimeters tall with shrub canopy typically greater than 20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.

52. Shrub/Scrub - Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.

71. Grassland/Herbaceous - Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.

72. Sedge/Herbaceous - Alaska only areas dominated by sedges and forbs, generally greater than 80% of total vegetation. This type can occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.

73. Lichens - Alaska only areas dominated by fruticose or foliose lichens generally greater than 80% of total vegetation.

74. Moss - Alaska only areas dominated by mosses, generally greater than 80% of total vegetation.

81. Pasture/Hay - Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.

82. Cultivated Crops - Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20 percent of total vegetation. This class also includes all land being actively tilled.

90. Woody Wetlands - Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

91. Palustrine Forested Wetland* -Includes all tidal and non-tidal wetlands dominated by woody vegetation greater than or equal to 5 meters in height and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent.

92. Palustrine Scrub/Shrub Wetland* - Includes all tidal and non-tidal wetlands dominated by woody vegetation less than 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent. The species present could be true shrubs, young trees and shrubs or trees that are small or stunted due to environmental conditions.

93. Estuarine Forested Wetland* - Includes all tidal wetlands dominated by woody vegetation greater than or equal to 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent. Total vegetation coverage is greater than 20 percent.

94. Estuarine Scrub/Shrub Wetland* - Includes all tidal wetlands dominated by woody vegetation less than 5 meters in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent. Total vegetation coverage is greater than 20 percent.

95. Emergent Herbaceous Wetlands - Areas where perennial herbaceous vegetation accounts for greater than 80 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

96. Palustrine Emergent Wetland (Persistent)* - Includes all tidal and non-tidal wetlands dominated by persistent emergent vascular plants, emergent mosses or lichens, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Plants generally remain standing until the next growing season.

97. Estuarine Emergent Wetland* - Includes all tidal wetlands dominated by erect, rooted, herbaceous hydrophytes (excluding mosses and lichens) and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent and that are present for most of the growing season in most years. Perennial plants usually dominate these wetlands.

98. Palustrine Aquatic Bed* - The Palustrine Aquatic Bed class includes tidal and nontidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is below 0.5 percent and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, detached floating mats, and rooted vascular plant assemblages.

99. Estuarine Aquatic Bed* - Includes tidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is equal to or greater than 0.5 percent and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, kelp beds, and rooted vascular plant assemblages.

* Coastal NLCD class only

c. Database Table Definition

d. Data Relationship Definition

e. Data Dictionary

See http://www.mrlc.gov/nlcd_definitions.asp

D. Policies

1. Restrictions

a. Use Constraints
The data is in the public domain, and you are free to do with it as you choose. MRLC would appreciate acknowledgment or credit regarding the source of the categorized images in any uses that you may have.

b. Access Constraints
None

c. Certification Issues
None

2. Maintenance
a. Temporal Information

b. Average Update Cycle

No update cycle known.

E. Acquisition Cost

1. Cooperative Agreement
a. Description of Agreement

None

b. Status of Agreement

2. Cost to Acquire Data

None

III. Integration

A. Value Added Process

1. Benefit to the Service Center

The purpose of the NLCD is to provide a current, consistent, and accurate National Land cover Database (NLCD) circa 2001 for the United States, at medium spatial resolution These data are intended for geographic display and analysis at the state level.

2. Process Model
a. Flow Diagram

b. Process Description

For prepping and loading the data onto the Data Gateway:

Summary: The data was downloaded by region from the NLCD site: http://www.mrlc.gov/scripts/mapserv.exe?map=d%3A%5CInetpub%5Cwwwroot%5Clccp%5Cmrlc2k%5Czones%5Czones.map.

One dataset for each state was produced, except for Alaska, it had to be produced by 11 UTM zones, and projected to UTM NAD83, Hawaii produced by 2 UTM zones, and Puerto Rico produced by 2 UTM zones - one predominant UTM zone per state, again except for Alaska, Hawaii, and Puerto Rico. Some situations required that the downloaded data from different regions be mosaiced together before the state data could be extracted.

ArcGIS is used for processing the data.
Specific Steps:
1. Bring landcover .img's into ArcMap, and bring buffered state shapefile into ArcMap.

2. Spatial Analyst Tools > Extraction > Extract by Mask
3. If more than one .img for the state, extract each piece separately. This produces clipped nlcd GRID file(s).

4. Go to Data Management Tools > Raster > Mosaic to New Raster. Add each of the extracted GRID files. If there is only one GRID file, that's OK – STILL do this step. For the output Raster file, navigate to the same directory and name the mosaic WITH .tif EXTENSION. Coordinate System – Choose UTM NAD83. Cellsize = 30. Mosaic Method - MINIMUM Accept other Default values.
5. Right click on one of the original landcover .img files, Properties > Symbology, and change Colormap to Unique Values, click Apply and OK. Save the .img file as a layer file, and bring into ArcMap.
6. Remove any other .img files in the Table of Contents, leaving only the .lyr file.
7. Go to Data Management Tools > Fields > Add Fields. Add Red, Green, and Blue fields – Text; length 20.
8. Open Both Tables first, then close
9. Join the .tif table to the .lyr file table.
10. Turn off all layers.
11. In tif table, Options > Select All
12. Go to Data Management Tools > Fields > Calculate Field. Use "Number". Calc the .tif Red, Green, Blue fields to the lyr file Red, Green, and Blue fields.
13. Right click on the .tif file, Remove Join.
14. Open the attribute table for the .tif file, and clear selection.
15. Copy files to UTM Directory and rename
16. Create world file in ArcCatalog
17. Create Metadata in ArcCatalog

* Please Note: If 3 or more images are needed to complete a state use these steps in place of Steps 1 – 4 above

1. Copy original images to Grids
   (Use Data Management Tools – Raster – Copy Raster
   (* Ignore Background Value 255)
2. Create Raster Dataset (grid)
   (Use Data Management Tools) –2a. Create Raster Dataset
   (Select UTM projection and zone)
   2b. Mosaic (select images) (target will be Raster Dataset created – 2a)
3. Spatial Analyst Tools – Extraction – Extract by mask (create a .img file)
4. Export .img file to. tif file and then continue with steps 5 – 17

18. Remote Login in and connect to IMS2, Run CatalogFP_Maker and generate the catalog shape files.
    My computer
    D: Drive
    Gateway
    GatewayCatalogs
    Make.Catalogs.exe

19. Run MakePreviews for Images and Metadata.
    My computer
    D: Drive
    Gateway
    GatewayCatalogs
    MakePreview.exe (Click on Everything in Source Path)
20. Create the Status Maps
    (link from "Status Maps" page)
    D: Drive
    Status Maps

21. Notify gateway Fort Collins team to load the catalogs, status maps and news

3. Technical Issues
   a. Tiling

      None

   b. Compression

      None

   c. Scale

      1:100,000

   d. Tonal Matching

      None

   e. Edge-matching

      None

4. Quality Control
   a. Procedures

      None – accepted as is

   b. Acceptance Criteria

      None – accepted as is

5. Data Steward
   a. Name and Organization

      National Cartography and Geospatial Center
      Natural Resources Conservation Service
      US Department of Agriculture
      501 Felix Street, Building 23
      P. O. Box 6567
      Fort Worth, Texas 76115-0567 USA

   b. Responsibilities

      Storage and access of the data.

B. Integrated Data Structure

1. Geospatial Data Format
   a. Format (raster, vector, etc.)

      Raster

   b. Format Name

Raster format for the NLCD:
TIFF is a non-proprietary format. It is a 32-bit floating-point raster format.

c. Data Extent

Same as source data

d. Horizontal and Vertical Resolution

Same as source data

e. Absolute Horizontal and Vertical Accuracy

Same as source data

f. Nominal Scale

1:100,000

g. Horizontal and Vertical Datum

Same as source data

h. Projection

UTM NAD83

i. Coordinate Units

Meters

j. Symbology

None

3. Attribute Data Format

a. Format Name

b. Database Size

The full dataset is.

4. Data Model

a. Geospatial Data Structure

Raster – tif

b. Attribute Data Structure

N/A raster data

c. Database Table Definition

N/A raster data

d. Data Relationship Definition

N/A raster data

e. Data Dictionary

N/A raster data

C. Resource Requirements

1. Hardware and Software

This is unknown at this time.

2. Staffing

This is unknown at this time.

D. Integration Cost

1. Hardware and Software

This is unknown at this time.

2. Staffing

This is unknown at this time.

IV. Delivery

A. Specifications

1. Directory Structure
   a. Folder Theme Data is Stored In

2. File Naming Convention
   a. List of Theme Files and The File Naming Convention

File structure

NLCD

<st>_<fc>

nlcd_<st>_utm<xx>.aux
nlcd_<st>_utm<xx>.tfw
nlcd_<st>_utm<xx>.tif
nlcd_<st>_utm<xx>.tif.vat.dbf
nlcd_<st>_utm<xx>.html
nlcd_<st>_utm<xx>.xml

where <st> = state; <fc> = State Fips Code; <xx> = predominant UTM Zone

Example – Alabama –

al_01

nlcd_al_utm16.aux
nlcd_al_utm16.tfw
nlcd_al_utm16.tif
nlcd_al_utm16.tif.vat.dbf

nlcd_al_utm16.html
nlcd_al_utm16.xml

B. User Information

1. Accuracy Assessment

a. Alignment with Other Theme Geospatial Data

This data should be considered not sufficiently detailed for the purpose of analysis at large scales. Alignment with the other data layers will not be perfect due to the fact that the data is captured at different scales and at different dates from other data.

b. Content

No formal accuracy assessment of the NLCD 2001 products has yet been completed; however, one is planned in the near future. Accuracy estimates across mapping zones ranged from 70% to 98%, with an overall average accuracy across all mapping zones of 83.9%. For more information, refer to Homer et al. (2004) - http://www.mrlc.gov/pdfs/July_PERS.pdf

2. Appropriate Uses of the Geospatial Data

a. Display Scale

These data are intended for geographic display and analysis at the state level. The data should be displayed and analyzed at a scale of 1:100,000 or smaller (e.g., 1:250,000, etc.).

Plot Scale

b. Area Calculations

Area Calculations are as accurate as the source data and capture scale and the algorithm used by ESRI software.

c. Decision Making

The data is as accurate as the source data and capture scale and the algorithm used by ESRI software.

C. Maintenance and Updating

1. Recommendations and Guidelines

a. Original data location and structure

The integrated database is at NCGC and the data is delivered to the Service Center.

b. Update Cycle

No update cycle known.

c. Availability

When the source data is updated.

d. Change Control

This is to be determined.