

**USDA Service Center Agencies
Geospatial Data Management Team
Data Management Plan For**

Land Use -NASS Cropland Data Layer (CDL) Data

July 2008

Ken Becker and Elaine Ortiz

I. Purpose and Scope (business case)

A. Purpose

The Cropland Data Layer (CDL) Program provides the National Agricultural Statistics Service (NASS) with internal proprietary county and state level acreage indications of major crop commodities, and secondarily provides the public with "statewide" (where available) raster, geo-referenced, categorized land cover data products after the public release of county estimates.

A collection of scenes from the satellites Landsat5, Landsat7, or RESOURCESAT-1, corresponding to an entire state or a major portion of a state, are categorized based on ground truth information collected from farmers by USDA enumerators. However, no farmer reported data is revealed or derivable from the categorized Landsat or RESOURCESAT-1 scenes, due to confidentiality protections. The individual categorized Landsat or RESOURCESAT-1 scenes need to be geo-referenced and stitched together to a common ortho-rectified base in order to be released as a public use GIS file. EarthSat Inc.'s GeoCover stock mosaic was chosen as the ortho base, because the GeoCover product offered accuracy and vast coverage over all of our project areas. The registration of the GeoCover mosaicked scene and the individual raw input scenes are used to get an approximate correspondence.

A correlation procedure is used on the raw Landsat or RESOURCESAT-1 scenes and the mosaicked scene to get an exact mapping of each pixel from the input Landsat or RESOURCESAT-1 scenes to the mosaicked scene. The results of the correlation are used to remap the pixels from the individual input scenes into the coordinate system of the mosaicked scene. The image analyst then specifies the mosaic priorities for scene placement, and the mosaic process begins by using the polynomials from the correlation to place the categorized pixels. A classified ortho-rectified mosaicked image (tif) is then output for distribution into the public domain.

This is primarily a raster dataset but does include one polygon shapefile for each state that is called the Area Sampling Frame, or Land Use Strata. Unlike the raster datasets that change yearly, this data does not change from year to year. The area sampling frame is a delineation of all parcels of land for the purpose of later sampling the parcels. The area frame is constructed by visually interpreting satellite imagery to divide a state into homogenous land use areas (strata) based on percent cultivated. The strata are typically defined as low, medium or high percent cultivated, non-agricultural land, urban use, agri-urban, or water. The boundaries of the strata usually follow identifiable features such as roads, railroads and waterways. The strata boundaries do not coincide with any political boundaries, with the exception of state boundaries.

B. Scope

The data extent is dependent on the year desired, and currently includes all or part of several states. See the status map at: <http://www.nass.usda.gov/research/Cropland/cdl-map.jpg> for data availability.

II. Acquisition

A. Data Source

1. Producer Information

a. Name

USDA/National Agricultural Statistics Service

b. Location of Headquarters

1400 Independence Avenue, SW, Room 5829-S
Washington, D.C. 20250-9410

c. Internet Address

<http://www.nass.usda.gov>

2. Publisher Information

a. Name

USDA/National Agricultural Statistics Service

b. Location of Headquarters

Spatial Analysis Research Section
3251 Old Lee Hwy, Suite 305
Fairfax, VA 22030

c. Internet Address

<http://www.nass.usda.gov>

3. Acquisition Information

a. Delivery Media

CD/DVD

b. Download URL

Available on CD-Rom through the official website
<<http://www.nass.usda.gov/research/Cropland/SARS1a.htm>>

c. Projected Data Availability Schedule

Currently available

B. Standards Information

1. Geospatial Data Standard

a. Standard Name and Steward Information

The NASS CDL is built upon the most current applicable standards for geospatial data and metadata. In particular, the CDL foundation is a raster, geo-referenced, categorized land cover data layer produced using satellite imagery from 1) the Thematic Mapper (TM) instrument on Landsat 5; 2) Landsat 7 ETM gap-filled data; or 3) Indiana Remote Sensing (IRS) Advanced Wide Field Sensor (AWiFS) on RESOURCESAT-1.

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 Version 2 revised June 1998.

The Cropland Data Layer (CDL) 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. CDL 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). CDL 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 CDL 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
Vector

b. Format Name

Raster
TIFF is a non-proprietary format. It is a 32-bit floating-point raster format.

Vector – ESRI Shapefile

c. Data Extent

The data extent is dependent on the year desired, and currently includes all or part of several states. See the status map at: <http://www.nass.usda.gov/research/Cropland/cdl-map.jpg> for data availability.

d. Horizontal and Vertical Resolution

TM product - 30 meters by 30 meters
AWiFS Product – 56 meters by 56 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

World Geodetic System 1984 for all appropriate areas (UTM zone 3 through 22).

h. Projection

UTM

i. Coordinate Units

Meters

j. Average Data Set Size

The full dataset is over 10.6GB and will grow with each update of new data.

k. Symbology

Each image has a color coded legend that corresponds to the particular crop or land cover.

2. Attribute Data Format

a. Format Name

Raster

Attribute Domain Values and Definitions: ROW CROPS 1-20

Categorization Code Land Cover

"1"	Corn
"2"	Cotton
"3"	Rice
"4"	Sorghum
"5"	Soybeans
"6"	Sunflowers
"10"	Peanuts
"11"	Tobacco

Raster

Attribute Domain Values and Definitions: SMALL GRAINS & HAY 21-40

Categorization Code	Land Cover
"21"	Barley
"22"	Durum Wheat
"23"	Spring Wheat
"24"	Winter Wheat (AR, IL, MS, NM)
"25"	Other Small Grains & Hay (Oats, Millet, Rye & Winter Wheat, Alfalfa & Other Hay)
"26"	Winter Wheat/Soybeans Double Cropped
"27"	Rye
"28"	Oats
"29"	Millet
"30"	Speltz
"31"	Canola
"32"	Flaxseed
"33"	Safflower
"34"	Rape seed
"35"	Mustard
"36"	Alfalfa

Raster

Attribute Domain Values and Definitions: OTHER CROPS 41-60

Categorization Code	Land Cover
"41"	Beets
"42"	Dry Edible Beans
"43"	Potatoes
"44"	Other Crops (Canola, Flaxseed, Safflower & very small acreage crops)
"45"	Sugar Cane
"46"	Sweet Potatoes
"47"	Misc. Fruit and Veg.
"48"	Watermelon
"50"	State 560 (State Specific Crop)
"51"	State 561 (State Specific Crop)
"52"	State 562 (State Specific Crop)
"53"	State 563 (State Specific Crop)
"54"	State 564 (State Specific Crop)
"55"	State 565 (State Specific Crop)
"56"	State 566 (State Specific Crop)
"57"	State 567 (State Specific Crop)
"58"	State 568 (State Specific Crop)
"59"	State 569 (State Specific Crop)

Raster

Attribute Domain Values and Definitions: OTHER LAND 61-80

Categorization Code	Land Cover
"61"	Fallow/Idle Cropland
"62"	Pasture/Range/CRP/Non Ag (Permanent & Cropland Pasture, Waste & Farmstead)
"63"	Woods

Raster

Attribute Domain Values and Definitions: TREE CROPS 66-80

"67"	Peaches
"68"	Apples

"69"	Grapes
"70"	Christmas Trees
"71"	Orchards, State 721-729 (State-specific)
"72"	Citrus
"73"	Managed Forest
"80"	Other Fruit

Raster

Attribute Domain Values and Definitions: OTHER 81-99

Categorization Code Land Cover

"81"	Clouds
"82"	Urban
"83"	Water
"84"	Roads/Railroads
"85"	Ditches/Waterways
"86"	Buildings/Homes/Subdivisions
"87"	Wetlands
"88"	Grassland
"90"	Mixed Water/Crops
"91"	Mixed Water/Clouds
"92"	Aquaculture Raster

Attribute Domain Values and Definitions: OTHER CROPS 100-119

Categorization Code Land Cover

"100"	Pickles
"101"	Chick Peas
"102"	Lentils
"103"	Peas
"104"	Fallow Sugarcane

Raster

Attribute Domain Values and Definitions: NLCD OTHER CATEGORIES 120-149

Categorization Code Land Cover

"120"	Developed, Open Space
"121"	Developed, Low Intensity
"122"	Developed, High Intensity
"123"	Commercial/Industrial/Transportation
"127"	Bare Rock/Sand/Clay
"128"	Quarries/Strip Mines/Gravel Pits
"129"	Transitional
"130"	Barren
"131"	Deciduous Forest
"132"	Evergreen Forest
"133"	Mixed Forest
"136"	Shrubland
"140"	Grasslands/Herbaceous
"142"	Urban/Recreational Grasses
"143"	Woody Wetlands
"144"	Emergent Herbaceous Wetlands

Vector – Poly Shapefile

Strata

Land use category based on percent cultivated (Source: U.S. Department of Agriculture, National Agricultural Statistics Service)

Value	Definition
11	> 75% cultivated, (cultivated categories consist of land used for crops, the remaining portion of land in the cultivated strata may include pasture, woods, waste, rangeland, etc.)
21	25%-75% cultivated, (cultivated categories consist of land used for crops, the remaining portion of land in the cultivated strata may include pasture, woods, waste, rangeland, etc.)
31	Agri-urban with > 100 homes per square mile (less than 15% cultivated)
32	Commercial/dense urban with > 100 homes per square mile (city center with 0% cultivated)
42	< 25% cultivated, (cultivated categories consist of land used for crops, the remaining portion of land in the cultivated strata may include pasture, woods, waste, rangeland, etc.)
50	Non-agricultural (state or national parks, military bases, etc.)
62	Water

b. Database Size

N/A

3. Data Model

a. Geospatial Data Structure

TIFF is a non-proprietary format. It is a 32 bit floating point grid format.

<http://partners.adobe.com/asn/developer/graphics/graphics.html>

ESRI Shapefile

b. Attribute Data Structure

Raster data –N/A

Vector Data –

Land Use Strata

Homogeneous land use area based on percent cultivated (Source: U.S. Department of Agriculture, National Agricultural Statistics Service)

Strata

Land use category based on percent cultivated (Source: U.S. Department of Agriculture, National Agricultural Statistics Service)

Value	Definition
11	> 75% cultivated, (cultivated categories consist of land used for crops, the remaining portion of land in the cultivated strata may include pasture, woods, waste, rangeland, etc.)
21	25%-75% cultivated, (cultivated categories consist of land used for crops, the remaining portion of land in the cultivated strata may include pasture, woods, waste, rangeland, etc.)
31	Agri-urban with > 100 homes per square mile (less than 15% cultivated)
32	Commercial/dense urban with > 100 homes per square mile (city center with 0% cultivated)
42	< 25% cultivated, (cultivated categories consist of land used for crops, the remaining

	portion of land in the cultivated strata may include pasture, woods, waste, rangeland, etc.)
50	Non-agricultural (state or national parks, military bases, etc.)
62	Water

- c. Database Table Definition
- d. Data Relationship Definition
- e. Data Dictionary

USDA - NATIONAL AGRICULTURE STATISTICS SERVICE'S 1:100,000-SCALE CROPLAND DATA LAYER - Specific for each state. See <http://www.nass.usda.gov/>

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. NASS 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
The CDL is updated yearly. Previous yearly data will be retained.

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 Cropland Data Layer Program is to use satellite imagery on an annual basis to (1) provide supplemental acreage estimates for the state's major commodities, (2) produce digital, crop specific, categorized geo-referenced output products, and (3) provide for data sharing among partners.

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:

1. Product Description – link to

<http://www.nass.usda.gov/research/Cropland/SARS1a.htm>

2. "<st><yr>nass" directories (<st>=state and <yr>=two digit year) sent from NASS are loaded in to state directory containing data from other years "<st>_nass" (working directory is located at <\\v480d\qfs6\nass>). Parent directory "<st>_nass" is then zipped into a single deliverable file "nass_<st>.zip" and moved to its final location (<\\v480d\qfs2\nass>) to be served via the Gateway. Remove zip files from <\\v480d\qfs6\nass> directory.

NASS_CD_L map and metadata are in a .zip file. The map and metadata must be extracted from the .zip and put somewhere that MakePreviews can use.

3. Run MakeCatalogs and generate the catalog shape files. (MakeCatalogs and MakePreviews are typically located in D:/GatewayCatalogs.)

4. Make a tmp directory under where NASS is stored and set permissions to 1755.

5. Extract the most recent .tif map and the metadata file for each state into that directory that are listed in the new and modified maps in fpDest/fp_NASS_CD_L.log

6. Rename the files to the standard name nass_<ST> where <ST> is the lower case postal abbreviation.

7. Run MakePreviews for Images and Metadata. First select the NASS_CD_L product then append "\tmp" to the Source Path field. The everything option must be used. Review the log for any problems.

8. Remove the /tmp folder created above.

9. In ArcMap, Create/Revise the Status Map (link from "Status Maps" page)

10. In ArcMap, export the status map to .jpg format (96dpi)

11. Create News Item in this directory: <\\Gateway2\ftp\GatewayNews>

12. 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

USDA/NASS Customer Service
c/o USDA/NASS Customer Service
1400 Independence Avenue, SW, Room 5038-S
Washington DC 20250-9410 USA

b. Responsibilities

Storage and access of the data.

B. Integrated Data Structure

1. Geospatial Data Format

a. Format (raster, vector, etc.)

Raster
Vector

b. Format Name

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

Vector format – ESRI shapefile (for state land use stratum; 1 per state – does not change from year to year)

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 over 10.6GB and will grow with each update of new data.

4. Data Model

a. Geospatial Data Structure

Raster – tif

Vector – ESRI Shapefile

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

Zip file containing all files for each state:

nass_<st>.zip, where st = 2-letter state

In the zip file are the following files:

Land use strata file for each state (graphic, metadata, & shapefile) - (does not change from yr to yr):

Graphic: strata_a_<st>.pdf

Metadata: strata_a_<st>.txt

Shapefile and associated files: strata_a_<st>.shp

where <st> = 2-letter state; where _a = is area feature for the strata shape file

Cropland Data layer analysis areas. These are not maps. These are metadata and accuracy assessments. One is metadata and the other is the percent correct. The info*.jpg files show the extents and dates of satellite coverage for that year.

Statinfo files (metadata and accuracy assessments) for each state, by year:

Graphic: cdlinfo_<procode>_i_<st>_<yyyy>.jpg

Metadata: cdldata_<procode>_i_<st>.htm

cdlstats_r_<st>_<yyyy>.htm

where <st> = 2-letter state; where <yyyy> is the calendar year; where _i = JPEG for the analysis areas

procode = either awifs (advanced wide field sensor) or tm (thematic mapper)

Cropland Data layer Classified Image (tif or jpg files) for each state, by year:

Raster: cdl_<procode>_r_<st>.tif

cdl_<procode>_r_<st>.tfw

Legend: cdl_legend_r_<st>_<yyyy>.jpg

Metadata: cdl_meta_r_<st>_<yyyy>.htm

where <st> = 2-letter state; where <yyyy> is the calendar year; where _r = raster for the classified. tif image

procode = either awifs (advanced wide field sensor) or tm (thematic mapper)

B. User Information

1. Accuracy Assessment

a. Alignment with Other Theme Geospatial Data

The data is captured at a scale of 1:100,000. 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

NASS publishes all available accuracy statistics for end-user viewing. The Percent Correct is calculated for each cover type in the ground truth, it shows how many of the total pixels were correctly classified (i.e. across all cover types). 'Commission Error' is the calculated percentage of all pixels categorized to a specific cover type that were not of that cover type in the ground truth (i.e. incorrectly categorized). CAUTION: a quoted Percent Correct for a specific cover type is worthless unless accompanied by its respective Commission Error. Example: if you classify every pixel in a scene to 'wheat', then you have a 100% correct wheat classifier (however its Commission Error is also almost 100%). The 'Kappa Statistic' is an attempt to adjust the Percent Correct using information gained from the confusion matrix for that cover type. Many remote sensing groups use the Percent Correct and/or Kappa statistics as their final measure of classification accuracy. Classification accuracy is generally between 85% to 95% correct for agricultural-related land cover categories

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 database is at NASS, and a copy of the data is at NCGC, on CD/DVD.

b. Update Cycle

The CDL is created yearly. Prior years are kept and are available for download.

c. Availability

When the source data is updated.

d. Change Control

This is to be determined.