



United States  
Department of  
Agriculture

Service Center  
Implementation  
(SCI)

# Geospatial Data Management Requirements Document

*Prepared by*  
*Science Applications International Corporation (SAIC)*  
**for**  
**The Data Management Team #5 on Geospatial Data Standards**

**Abstract:** This document represents a framework of geospatial dataset requirements and system requirements that can be used to implement the management of geospatial datasets for the USDA Service Center Initiative.

**Keywords:** geospatial, data management, dataset, requirements

**Published and distributed in April 28, 2000**

## Introduction

(This introduction is not part of the Geospatial Data Requirements Document)

The Service Center Initiative (SCI) Data Management Team #5: Geospatial Data Standards developed the April 2000 *Geospatial Data Requirements Document* to illustrate the requirements for geospatial datasets and the system requirements that are necessary to support the management and dissemination of those datasets.

The individuals who contributed to the development of this guide are:

David Anderson, (NRCS) Service Center Data Team Leader		
Carol Ernst, (FSA) Co-Leader		
Emil Horvath, (NRCS) Co-Leader		
Dennis Lytle (NRCS)	Larry L. Davis (NRCS)	Nicole Soltyka (SAIC)
Kent Williams (FSA)	Dwain Daniels (NRCS)	Tom McCarty (SAIC)
Steve Nechero (NRCS)	Elaine Ortiz (NRCS)	Randy Frosh (Unisys)
Kenny Legleither (NRCS)	Elizabeth Young(NRCS)	

Figure 1-1 Working group list

1. Overview .....	1
1.1. Scope.....	1
1.2. Purpose .....	1
1.3. Acronyms.....	2
1.4. Assumptions .....	3
2. Dataset descriptions .....	4
2.1. Geospatial datasets.....	4
2.2. USDA datasets .....	15
2.2.1. Soils.....	15
2.2.2. Common Land Unit (CLU).....	16
2.2.3. Land use land cover .....	16
2.2.4. Ortho Imagery.....	17
2.2.5. Climate .....	18
2.2.6. Environmental easements .....	20
2.2.7. Hydrologic unit (11/14-digit).....	20
2.2.8. Plants .....	21
2.2.9. Administrative areas .....	21
2.2.10. Digital Map Fininshing (DMF).....	24
2.2.11. Rivers and Streams (DMF) .....	24
2.2.12. Railroads (DMF).....	24
2.2.13. Roads (DMF) .....	24
2.2.14. Census .....	24
2.3. Non-USDA datasets.....	25
2.3.1. Air quality .....	25
2.3.2. Cadastral.....	26
2.3.3. Census .....	27
2.3.4. Elevation .....	34
2.3.5. Endangered habitat.....	34
2.3.6. Geographic Names Information System (GNIS) .....	34
2.3.7. Government units .....	34
2.3.8. Hydrography .....	36
2.3.9. Hydrologic unit (8-digit).....	37
2.3.10. Land Use Land Cover (LULC).....	37
2.3.11. Map indexes .....	38
2.3.12. Topographic images (Digital Raster Graphs - DRG) .....	38
2.3.13. Transportation.....	39
2.3.14. Wetlands – National Wetlands Inventory (NWI) .....	40
3. Dataset Physical characteristics .....	40
3.1. Source .....	40
3.1.1. Alternative sources.....	41
3.2. FGDC – NSDI .....	41
3.2.1. USGS datasets.....	42
3.2.2. Census Bureau datasets.....	42
3.2.3. State.....	43
3.2.4. Regional, county, and local agencies.....	43

3.2.5. Commercial vendors .....	43
3.3. Size .....	44
3.4. Evolution of Database Size.....	45
3.5. Format.....	46
3.5.1. Vector datasets .....	46
3.5.2. Raster datasets.....	46
3.5.3. Spatially enabled databases.....	46
3.6. Tiling.....	46
3.7. Temporality.....	47
3.8. Change control.....	47
3.9. Other physical characteristics .....	47
4. Current data management.....	49
4.1. Ownership .....	50
4.2. Storage .....	50
4.3. Distribution.....	51
5. Dataset management requirements.....	51
5.1. User requirements (general use case scenarios) .....	51
5.1.1. Definition.....	51
5.1.2. Geospatial Data Acquisition .....	52
5.1.3. Geospatial Data Integration .....	56
5.1.4. Geospatial Data Delivery.....	59
5.2. Data management system requirements .....	62
5.2.1. Browsing.....	62
5.2.2. Storage .....	63
5.2.3. Management.....	63
5.2.4. Delivery.....	63
5.3. Standards and specifications .....	64
5.4. Metadata .....	64
5.5. Transactional datasets .....	64
5.6. Availability .....	64
5.7. Security.....	65
5.8. Privacy.....	65
5.9. Disaster recovery .....	65
5.10. Performance measures.....	65
6. Summary .....	66
Appendix A – Bibliography.....	67
Table 2-1 USDA datasets.....	5
Table 2-2 Non-USDA datasets .....	10
Table 3-1 Other physical characteristics .....	48
Table 5-1 Geospatial Data Acquisition Use Case Description.....	52
Table 5-2 Geospatial Data Integration Use Case Description .....	56
Figure 1-1 Working group list .....	ii
Figure 3-1 Geospatial Data Distribution Timeline .....	45
Figure 4-1 Current Geospatial Data Management Operations .....	50

Figure 5-1 Geospatial Data Acquisition Use Case Diagram ..... 52  
Figure 5-2 Geospatial Data Integration Use Case Diagram..... 56

---

## Draft Geospatial Data Requirements Document

---

### 1. Overview

#### 1.1. Scope

The scope of the *Geospatial Data Requirements Document* encompasses the realm of data management activities currently in practice within the Service Center Initiative (SCI) and those proposed for future application. This document represents the first part of a two-part overall *Geospatial Data Implementation Plan*.

Part one of this plan encompasses descriptions of the geospatial datasets identified in the *Service Center Geographic Information System Strategy* [A1] that address the business needs of United States Department of Agriculture (USDA) programs served from the Service Centers. It also describes the physical characteristics of those datasets, including size, format and change control practices for the datasets as well as their attributes and associated metadata. In addition, the dataset management system requirements, including storage, access and distribution and format requirements are described.

The second part of the overall *Geospatial Data Implementation Plan*, a separate document, will begin with an inventory of the existing components of the technical infrastructure and how that translates into the overall implementation of geospatial dataset management. From there, the plan will identify the technical architecture required to implement the data management requirements outlined in this document. It will include recommendations for the physical location of servers; operations; metadata management; interaction between production units, state offices and Service Centers. In addition, the document will detail the hardware, software, telecommunications infrastructure, human resources, and their roles and responsibilities.

#### 1.2. Purpose

The purpose of this document is to provide a framework for data management based on a set of geospatial datasets and data management requirements identified by the SCI team. This document provides a set of coordinated processes and procedures required for managing geospatial datasets among the three USDA partner agencies involved in Business Process Reengineering (BPR) of program delivery at the Service Centers. The partner agencies involved in this effort are Farm Services Agency (FSA), Natural Resources Conservation Service (NRCS) and Rural Development (RD). This document will form the basis for a technical architecture. This document may also be used as a reference in a comparative cost study that examines the cost of managing geospatial data in-house versus outsourcing some or the entire management infrastructure.

### 1.3. Acronyms

AID	Acquisition, Integration and Delivery
AIRS	Aerometric Information Retrieval System
AQS	Air Quality System
ASP	Application Service Provider
BIA	Bureau of Indian Affairs
BPR	Business Process Reengineering
CCE	Common Computing Environment
CD-ROM	Compact Disc Read Only Memory
CLU	Common Land Unit
COE	Corps of Engineers
CONUS	Conterminous United States
COTS	Commercial Off-the-Shelf
CRP	Conservation Reserve Program
CST	Customer Service Toolkit
CTTP	Census Transportation Planning Package
DEM	Digital Elevation Model
DLG	Digital Line Graphs
DMF	Digital Map Finishing
DMT	Data Management Team
DRG	Digital Raster Graphs
EROS	Earth Resources Observation System
EDC	EROS Data Center
EPA	Environmental Protection Agency
ESRI	Environmental Systems Research Institute
EQIP	Environmental Quality Incentives Program
EWP-FP	Emergency Watershed Protection Program – Flood Plain
FOSA	Field Office Service Area
FPP	Farm Protection Program
FTP	File Transfer Protocol
FEMA	Federal Emergency Management Agency
FGDC	Federal Geographic Data Committee
FIPS	Federal Information Processing Standard
FIRM	Flood Insurance Rate Map
FSA	Farm Service Agency
FWS	Fish and Wildlife Service
GIRAS	Geographic Information Retrieval Analysis System
GNIS	Geographic Names Information System
GUI	Graphical User Interface
HU	Hydrologic Unit
HUC	Hydrologic Unit Code
IBM	International Business Machines
IT	Information Technology
I/O	Inter-Operability
LAN	Local Area Network
MLRA	Major Land Resource Agency

MrSID	Multi-Resolution Seamless Image Database
MUIR	Map Unit Interpretations Record
NAD	North American Datum
NAPP	National Aerial Photography Program
NASIS	National Soil Information System
NASA	National Air and Space Administration
NCSS	National Cooperative Soil Survey
NDOP	National Digital Orthophoto Program
NMD	National Mapping Division
NPS	National Park Service
NRCS	Natural Resources Conservation Service
NRI	Natural Resource Inventory
NSDI	National Spatial Data Infrastructure
NWCC	National Weather and Climate Center
NWI	National Wetlands Inventory
PLSS	Public Land Survey System
PRISM	Parameter-elevation Regressions on Independent Slopes Model
RCD	Resource Conservation and Development
RD	Rural Development
RUSLE	Revised Universal Soil Loss Equation
SCI	Service Center Initiative
SCDM	Service Center Data Management
SDE	Spatial Database Engine
SFWMD	South Florida Water Management District
SSA	Soil Survey Area
SSURGO	Soil Survey Geographic Data
STF	Summary Tape Files
SWCD	Soil and Water Conservation District
TAZ	Traffic Analysis Zones
TIGER	Topologically Integrated Geographic Encoding and Referencing
URL	Uniform Resource Locator
USDA	United States Department of Agriculture
USGS	United States Geological Survey
USFS	United States Forest Service
USPS	United States Postal Service
UTM	Universal Transverse Mercator
WAN	Wide Area Network
WHO	World Health Organization
WRP	Wetlands Reserve Program

#### **1.4. Assumptions**

To determine the data management requirements for this document, the following assumptions have been made:

1. USDA telecommunications are supported by a backbone with a high bandwidth line between major Information Technology (IT) centers but maintains low bandwidth lines (56-128kb) between Service Centers, state offices and major IT centers.
2. Data formats are specific for currently identified USDA business needs and may change as new technology becomes available and business needs change.
3. The Geospatial Technical Architecture referred to in this document will initially support the Customer Service Toolkit (CST) application, Wetland Reserve Program (WRP) application, Resource Data Gateway, all NRCS programs and production centers and FSA production centers.

## 2. Dataset descriptions

The *Standard for Geospatial Data* [A2] defines dataset categories as a logical group or division of a geospatial dataset collection that includes at least one geospatial dataset complete with all dataset metadata and feature metadata. Geospatial datasets are simply a group of similar spatial phenomena in a geospatial dataset category and are often referred to as themes. In 1998, the *Service Center Geographic Information System Strategy* identified more than 20 separate geospatial dataset categories that were important to the business processes carried out by the three partner agencies that conduct business in Service Centers. Since that time the list has been refined to include all geospatial dataset categories that have been deemed necessary by the BPR activities, the Data Management Team (DMT) and the Geospatial Data Acquisition, Integration and Delivery (AID) Team, each working jointly under the SCI. Of those geospatial dataset categories, four have been identified as especially important and have been flagged as high priority. The high priority dataset categories are soils, Common Land Unit (CLU), orthophotography and demographics.

The Data AID Team completed their work in September 1999 and presented their findings in the *Geospatial Data Acquisition, Implementation and Delivery National Implementation Strategy Plan* [A3]. Their mandate, during their tenure, was to deliver comprehensive geospatial datasets, available on a nationally consistent basis, to the nine BPR pilot sites. This document expands the scope of the Data AID Team and presents the requirements for a technical architecture required to manage all geospatial datasets used by the SCI.

### 2.1. Geospatial datasets

The geospatial datasets identified as important to USDA business have been grouped into “USDA datasets” and “non-USDA datasets” for the purposes of documenting the geospatial dataset management requirements. USDA datasets are produced and maintained by USDA and identified as especially important and required to support the SCI BPR. These include soils, ortho imagery, CLU, climate (precipitation and temperature), environmental easements, plants, 11/14-digit hydrologic unit boundaries, and administrative areas that include servicing areas, national program areas and Office Information Profile (OIP) locations. These USDA datasets are maintained through

established maintenance programs. Both soils and CLU datasets are products of on-going USDA programs. First generation ortho imagery is being acquired through the National Digital Ortho Photography Program (NDOP). Both FSA and NRCS cost share in the NDOP program. Neither agency is currently cost sharing in second generation ortho through NDOP.

Non-USDA geospatial datasets include mostly static vector and raster themes and are commonly used for reference purposes during geographic display. However, these non-USDA reference datasets are often used during analysis to reveal spatial relationships and answer spatial business queries. Tables 2-1 and 2-2 detail the dataset category name, producer, USDA distributor, tiling scheme, format and on-line source. It should be noted that the data available on-line is not always provided in a format ready to be used in the USDA Service Center environment. Post processing, format conversion and projection transformations may be necessary. In some cases production years of the dataset are provided, this information is provided for reference purposes only and does not indicate that the dataset will be maintained for historical purposes.

**Table 2-1 USDA datasets**

Dataset category	Producer	USDA distributor	Tiling scheme <sup>1</sup>	Format	On-line source and alternate source if applicable
<b>Soils</b>					
SSURGO – areas	USDA - NRCS	USDA - NRCS	survey area	vector polygons	<a href="http://www.ftw.nrcs.usda.gov/ssur_ftp.html">http://www.ftw.nrcs.usda.gov/ssur_ftp.html</a>
SSURGO – lines	USDA - NRCS	USDA - NRCS	survey area	vector lines	<a href="http://www.ftw.nrcs.usda.gov/ssur_ftp.html">http://www.ftw.nrcs.usda.gov/ssur_ftp.html</a>
SSURGO – points	USDA - NRCS	USDA - NRCS	survey area	vector points	<a href="http://www.ftw.nrcs.usda.gov/ssur_ftp.html">http://www.ftw.nrcs.usda.gov/ssur_ftp.html</a>
SSURGO – linear features	USDA - NRCS	USDA - NRCS	survey area	vector lines	<a href="http://www.ftw.nrcs.usda.gov/ssur_ftp.html">http://www.ftw.nrcs.usda.gov/ssur_ftp.html</a>
Frozen soils (1990)	USDA - NRCS	USDA - NRCS	survey area	tabular database	At this time, data is delivered by State Soil Scientists on a case by case basis.
NASIS soil attributes	USDA - NRCS	USDA - NRCS	county	tabular database	<a href="http://nasis.nrcs.usda.gov/downloads/">http://nasis.nrcs.usda.gov/downloads/</a>
SSA boundary	USDA - NRCS	USDA - NRCS	survey area	vector polygons	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/">ftp://ftp.ftw.nrcs.usda.gov/pub/</a>
<b>Common Land Unit</b>					
Common Land Unit areas	USDA - FSA	USDA - FSA	county	vector polygons	<a href="ftp://apfo.usda.gov/pub/incoming/clu">ftp://apfo.usda.gov/pub/incoming/clu</a>
<b>Land Use Land Cover</b>					
Land Cover	USDA - FSA	USDA - FSA	county	vector polygons	Currently not defined.
Land Use	USDA - FSA	USDA - FSA	county	vector polygons, points	Currently not defined.

<sup>1</sup> Tiling indicates current tiling scheme

Dataset category	Producer	USDA distributor	Tiling scheme <sup>1</sup>	Format	On-line source and alternate source if applicable
<b>Ortho Imagery</b>					
Mosaicked Digital Orthophoto Quadrangle (MDOQ)	USDA-FSA	USDA-FSA	quadrangle	imagery in MrSID format	APFO not currently distributing on-line.
County Orthomosaic	USDA-FSA	USDA-FSA	county	imagery in TIF format	APFO not currently distributing on-line.
Digital Orthophoto Quarter Quadrangle (DOQQ)	USGS	USDA - APFO	quarter quadrangle	imagery in TIF format	<a href="http://edcwww.cr.usgs.gov/webglis/glisbin/searich.pl?DOQ_QQUAD">http://edcwww.cr.usgs.gov/webglis/glisbin/searich.pl?DOQ_QQUAD</a>
Ortho image catalog	USDA - FSA, NRCS	USDA - APFO	county	tabular database	APFO not currently distributing on-line.
<b>Climate - Precipitation</b>					
Annual mean precipitation	Oregon Climate Service	USDA - NRCS	state, regional, CONUS (conterminous US)	vector polygons	<a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Annual mean precipitation	Oregon Climate Service	USDA - NRCS	state, regional, CONUS (conterminous US)	raster	<a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean precipitation	Oregon Climate Service	USDA - NRCS	state, regional, CONUS (conterminous US)	vector polygons	<a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean precipitation	Oregon Climate Service	USDA - NRCS	state, regional, CONUS (conterminous US)	raster	<a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
<b>Climate - Temperature</b>					
Annual mean maximum temperature	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean maximum temperature	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Annual mean minimum	Oregon Climate	USDA - NRCS	CONUS (conterminous	raster	No files delivered to date.

Dataset category	Producer	USDA distributor	Tiling scheme <sup>1</sup>	Format	On-line source and alternate source if applicable
temperature	Service		US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)		<a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean minimum temperature	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Annual mean temperature	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean temperature	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
<b>Climate - Dewpoint</b>					
Annual mean maximum dewpoint	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean maximum dewpoint	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>

Dataset category	Producer	USDA distributor	Tiling scheme <sup>1</sup>	Format	On-line source and alternate source if applicable
			Rico)		
Annual mean minimum dewpoint	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean minimum dewpoint	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US), Alaska and other US regions (Hawaii, Pacific Basin Islands, Puerto Rico)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
<b>Climate - Other, Derived Climate Variables</b>					
R-factor for the US	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Annual mean growing degree days at base	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Monthly mean growing degree days at base	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Average last spring frost date (32F)	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Average last spring frost date (28F)	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Average first fall frost date (32F)	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
Average first fall frost date (28F)	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
January one in ten minimum temperature	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>

Dataset category	Producer	USDA distributor	Tiling scheme <sup>1</sup>	Format	On-line source and alternate source if applicable
Monthly mean snowfall	Oregon Climate Service	USDA - NRCS	CONUS (conterminous US)	raster	No files delivered to date. <a href="http://www.ocs.orst.edu/prism/prism_new.html">http://www.ocs.orst.edu/prism/prism_new.html</a>
<b>Environmental Easements</b>					
Wetland Reserve Program (WRP)	USDA - NRCS	USDA - NRCS	county	vector polygons	No files delivered to date.
<b>Hydrologic units</b>					
11/14-digit hydrologic units	USDA - NRCS	USDA - NRCS	county	vector polygons	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/hu/">ftp://ftp.ftw.nrcs.usda.gov/pub/hu/</a>
<b>Plants</b>					
Plants database	USDA - NRCS	USDA - NRCS	state	ASCII database, comma delimited	<a href="http://plants.usda.gov/plants/cgi_bin/topics.cgi?earl=dl_state.html">http://plants.usda.gov/plants/cgi_bin/topics.cgi?earl=dl_state.html</a>
<b>Administrative areas (USDA National program datasets)</b>					
Environmental Quality Improvement Program Areas (EQIP)	USDA - NRCS	USDA - NCGC	national, state	vector polygons	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/">ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/</a>
Emergency Watershed Protection Program - Flood Plain (EWP-EP)	USDA - NRCS	N/A	national, state	vector polygons	Program data has not been developed
Farmland Protection Program (FPP)	USDA - NRCS	N/A	national, state	vector polygons	Program data has not been developed
Field Office Service Areas (FOSA)	USDA - NCGC	USDA - NCGC	national	vector polygons	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/">ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/</a>
Soil and Water Conservation Districts (SWCD)	USDA - NRCS	USDA - NCGC	national	vector polygons	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/">ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/</a>
Office Information Profile (OIP)	USDA - NRCS	USDA - NRCS	national	tabular database, vector points	<a href="http://offices.usda.gov/scripts/ndISAPI.dll/oip_public/USA_map">http://offices.usda.gov/scripts/ndISAPI.dll/oip_public/USA_map</a>
Resource Conservation and District (RC&D) boundaries	USDA - NRCS	USDA - NCGC	national	vector polygons	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/">ftp://ftp.ftw.nrcs.usda.gov/pub/ams/gisdata/</a>
Natural Resources Inventory (NRI)	USDA - NRCS	USDA - NRCS	county	tabular database	<a href="http://www.nhq.nrcs.usda.gov/NRI/1997/OASys97.html">http://www.nhq.nrcs.usda.gov/NRI/1997/OASys97.html</a> <b>or</b> <a href="ftp://ftp.ftw.nrcs.usda.gov/pub/nri/">ftp://ftp.ftw.nrcs.usda.gov/pub/nri/</a>
NRI cross boundaries	USDA - NRCS	USDA - NRCS	national	vector polygons	<a href="ftp://ftp.nhq.nrcs.usda.gov/pub/outgoing/land/arc_export/us48cross.e00.zip">ftp://ftp.nhq.nrcs.usda.gov/pub/outgoing/land/arc_export/us48cross.e00.zip</a>
<b>Digital Map Finishing (DMF)</b>					

Dataset category	Producer	USDA distributor	Tiling scheme <sup>1</sup>	Format	On-line source and alternate source if applicable
Rivers and streams (DMF – soil survey)	USDA – NRCS	USDA – NCGC	county	vector lines, annotation	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/mapfinishing/procedures/states/">ftp://ftp.ftw.nrcs.usda.gov/pub/mapfinishing/procedures/states/</a>
Railroads (DMF – soil survey)	USDA – NRCS	USDA – NCGC	county	vector lines, annotation	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/mapfinishing/procedures/states/">ftp://ftp.ftw.nrcs.usda.gov/pub/mapfinishing/procedures/states/</a>
Roads (DMF – soil survey)	USDA – NRCS	USDA – NCGC	county	vector lines, annotation	<a href="ftp://ftp.ftw.nrcs.usda.gov/pub/mapfinishing/procedures/states/">ftp://ftp.ftw.nrcs.usda.gov/pub/mapfinishing/procedures/states/</a>
Census					
Census of Agriculture	NASS, US Census	Undecided	county	tabular database	<a href="http://www.nass.usda.gov/census">http://www.nass.usda.gov/census</a>

Table 2-2 Non-USDA datasets

Dataset category	Producer	USDA distributor	Tiling scheme <sup>2</sup>	Format	On-line source and alternate source if applicable
Air quality					
Air Quality System (AQS)	EPA	N/A	state	tabular database	<a href="http://www.epa.gov/aqspubl1/air_quality_tables.html">http://www.epa.gov/aqspubl1/air_quality_tables.html</a> or <a href="http://www.epa.gov/airs/aewin/">http://www.epa.gov/airs/aewin/</a>
Cadastral					
National Park Service (NPS) - areas	NPS	USDA – NCGC	national	vector polygons	<a href="http://www.nps.gov/gis/national_data.html">http://www.nps.gov/gis/national_data.html</a>
NPS – lines	NPS	USDA – NCGC	county	vector lines	<a href="http://www.nps.gov/gis/national_data.html">http://www.nps.gov/gis/national_data.html</a> (lines created from polygon file located at this site)
Native American Lands	US Census, EPA, BIA	N/A	county	vector polygons	<a href="http://es.epa.gov/oeca/ai/lesp/rmailesp.html">http://es.epa.gov/oeca/ai/lesp/rmailesp.html</a> or <a href="http://www.wes.army.mil/el/ccspt/natamap/usapg.html">http://www.wes.army.mil/el/ccspt/natamap/usapg.html</a> or <a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Public Land Survey System (PLSS) – areas	USGS, BLM	USDA – NCGC	county	vector polygons	<a href="ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/">ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/</a> or <a href="http://www.blm.gov/gcdb/">http://www.blm.gov/gcdb/</a>

<sup>2</sup> Tiling indicates current tiling scheme

<b>Dataset category</b>	<b>Producer</b>	<b>USDA distributor</b>	<b>Tiling scheme<sup>2</sup></b>	<b>Format</b>	<b>On-line source and alternate source if applicable</b>
PLSS - lines	USGS, BLM	USDA – NCGC	county	vector lines	<a href="ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/">ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/</a>
<b>Census</b>					
Census block demographics	US Census	USDA – NCGC	county	tabular database	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Census blocks	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Coastal water	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Census transportation planning package areas	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Census block group demographics	US Census	USDA – NCGC	county	tabular database	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Census block groups	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Key geographic locations	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Landmark areas	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Railroads	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Utility lines	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Landmark lines	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Ridge and fence lines	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Streams	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Unclassified lines	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Landmark points	US Census	USDA – NCGC	county	vector points	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Minor civil divisions	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>

<b>Dataset category</b>	<b>Producer</b>	<b>USDA distributor</b>	<b>Tiling scheme<sup>2</sup></b>	<b>Format</b>	<b>On-line source and alternate source if applicable</b>
Minor civil divisions	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Places – block	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Places – tract	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Elementary school districts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Middle school districts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Secondary school districts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Unified school districts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Coastal shoreline	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Alternate street names	US Census	USDA – NCGC	county	tabular database	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Streets	US Census	USDA – NCGC	county	vector lines	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Census tract demographics	US Census	USDA – NCGC	county	tabular database	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Traffic analysis zones	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Census tracts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Urban areas	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Urban and rural outlines	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a> <a href="http://data.esri.com/data/online/tiger/index.html">http://data.esri.com/data/online/tiger/index.html</a>
Voting districts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Land and water outlines	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>

Dataset category	Producer	USDA distributor	Tiling scheme <sup>2</sup>	Format	On-line source and alternate source if applicable
					<a href="#">l.html</a>
Water bodies	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
<b>Elevation</b>					
Contours – lines	USGS	USDA – NCGC	county	vector lines	<a href="http://edcwww.cr.usgs.gov/webglis">http://edcwww.cr.usgs.gov/webglis</a>
Geodetic survey monuments – points	NOAA – NGS	USDA – NCGC	county	vector points	<a href="http://www.ngs.noaa.gov/datasheet.html">http://www.ngs.noaa.gov/datasheet.html</a>
<b>Endangered habitat</b>					
Endangered habitat	USGS	USDA – NCGC	N/A	vector polygons, tabular database	<a href="http://orion.cr.usgs.gov/">http://orion.cr.usgs.gov/</a> or <a href="http://orion.cr.usgs.gov/ecos/ecos_index.html">http://orion.cr.usgs.gov/ecos/ecos_index.html</a>
<b>Geographic Names Information System (GNIS)</b>					
Geographic names	USGS	USDA – NCGC	county	vector points	<a href="http://mapping.usgs.gov/www/gnis/gnisftp.html">http://mapping.usgs.gov/www/gnis/gnisftp.html</a>
<b>Government Units</b>					
Administrative boundaries	USGS	USDA – NCGC	county	vector lines	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
Administrative areas	USGS	USDA – NCGC	county	vector polygons	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
Congressional districts	US Census	USDA – NCGC	county	vector polygons	<a href="http://www.census.gov/mp/www/rom/msrom121.html">http://www.census.gov/mp/www/rom/msrom121.html</a>
Cities	USGS	USDA – NCGC	county	vector points	<a href="http://mapping.usgs.gov/www/gnis/gnisftp.html">http://mapping.usgs.gov/www/gnis/gnisftp.html</a>
County boundaries	USGS	USDA – NCGC	county	vector polygons	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
County lines	USGS	USDA – NCGC	county	vector lines	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
Man made area features	USGS	USDA – NCGC	county	vector polygons	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
Man made line features	USGS	USDA – NCGC	county	vector lines	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
State areas	USGS	USDA – NCGC	state	vector polygons	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
State boundaries	USGS	USDA – NCGC	state	vector lines	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html">http://edcwww.cr.usgs.gov/glis/hyper/guide/100kdlgfig/states.html</a>
ZIP code centroids	US Postal Service, US Census	USDA – NCGC	national	vector points	Dataset available with ArcView 3.2 software purchase.
<b>Hydrography</b>					

<b>Dataset category</b>	<b>Producer</b>	<b>USDA distributor</b>	<b>Tiling scheme<sup>2</sup></b>	<b>Format</b>	<b>On-line source and alternate source if applicable</b>
Damsites	Army COE, EPA	USDA – NCGC	county	vector, points	<a href="http://www.epa.gov/ost/basins/metadata.htm">http://www.epa.gov/ost/basins/metadata.htm</a>
Flood hazard areas	FEMA	USDA – NCGC	county	vector polygons	<a href="http://www.fema.gov/MSG/states.htm">http://www.fema.gov/MSG/states.htm</a>
Rivers and streams	USGS	USDA – NCGC	county	vector lines	<a href="http://edcwww.cr.usgs.gov/doc/edchome/ndcd/b/dlg_large/states.html">http://edcwww.cr.usgs.gov/doc/edchome/ndcd/b/dlg_large/states.html</a>
Rivers and streams annotation	USDA	USDA – NCGC	county	vector lines	Not currently available on the Web.
Rivers and streams (EPA Reach 3)	EPA	USDA – NCGC	county	vector lines	<a href="http://www.epa.gov/ost/basins/metadata.htm">http://www.epa.gov/ost/basins/metadata.htm</a>
<b>Hydrologic Units</b>					
8-digit hydrologic units	USGS	USDA – NCGC	county	vector polygons	<a href="http://nsdi.usgs.gov/products/huc.html">http://nsdi.usgs.gov/products/huc.html</a>
<b>Land Use Land Cover</b>					
Multi-Resolution Land Cover (MRLC)	USGS	USDA – NCGC	county	imagery	<a href="http://www.epa.gov/mrlc/Databases.html">http://www.epa.gov/mrlc/Databases.html</a>
Land use land cover	USGS	USDA – NCGC	county	vector polygons	<a href="http://www.epa.gov/enviro/html/esdls/data_sets.html">http://www.epa.gov/enviro/html/esdls/data_sets.html</a>
Barren land	USGS	USDA – NCGC	county	vector polygons	<a href="http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dlgqmap/8dqm0496.pdf">http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/dlgqmap/8dqm0496.pdf</a>
Surface cover	USGS	USDA – NCGC	county	vector polygons	<a href="http://bard.wr.usgs.gov/html/dir/dlg_html/dlginfo.html">http://bard.wr.usgs.gov/html/dir/dlg_html/dlginfo.html</a>
<b>Map Indexes</b>					
Photo index	USGS	USDA – NCGC	county	vector points	<a href="http://edcwww.cr.usgs.gov/glis/hyper/guide/napp#napp15">http://edcwww.cr.usgs.gov/glis/hyper/guide/napp#napp15</a>
Quadrangle areas	USGS	USDA – NCGC	quadrangle	vector polygons	<a href="http://greenwood.cr.usgs.gov/pub/quad-indexes/">http://greenwood.cr.usgs.gov/pub/quad-indexes/</a>
Quadrangle lines	USGS	USDA – NCGC	quadrangle	vector lines	<a href="http://greenwood.cr.usgs.gov/pub/quad-indexes/">http://greenwood.cr.usgs.gov/pub/quad-indexes/</a>
<b>Topographic Images (Digital Raster Graphics – DRGs)</b>					
Topographic image catalog	USDA – NCGC	USDA – NCGC	county	index table	Produced at NCGC on county by county basis.
DRG – compressed	USDA-NCGC	USDA – NCGC	county	raster MrSID	Will be available from NCGC FTP site.
DRG – standard	USGS	USDA – NCGC	quadrangle	raster	<a href="http://edcwww.cr.usgs.gov/Webglis/glisbin/search.pl?DRG_INDIV">http://edcwww.cr.usgs.gov/Webglis/glisbin/search.pl?DRG_INDIV</a>
DRG – enhanced	USDA-NCGC	USDA – NCGC	quadrangle	raster	Will be available from NCGC FTP site.
<b>Transportation</b>					
Utility lines	USGS	USDA – NCGC	county	vector lines	<a href="ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/">ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/</a>

Dataset category	Producer	USDA distributor	Tiling scheme <sup>2</sup>	Format	On-line source and alternate source if applicable
Railroad lines	USGS	USDA – NCGC	county	vector lines	<a href="ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/">ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/</a>
Roads	USGS	USDA – NCGC	county	vector lines	<a href="ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/">ftp://edcftp.cr.usgs.gov/pub/data/DLG/24k/</a>
<b>Wetlands -NWI</b>					
National Wetland Inventory areas	F&WS	USDA – NCGC	county	vector polygons	<a href="http://www.nwi.fws.gov/download.htm">http://www.nwi.fws.gov/download.htm</a>
National Wetland Inventory linear features	F&WS	USDA – NCGC	county	vector lines	<a href="http://www.nwi.fws.gov/download.htm">http://www.nwi.fws.gov/download.htm</a>
National Wetland Inventory lines	F&WS	USDA – NCGC	county	vector lines	<a href="http://www.nwi.fws.gov/download.htm">http://www.nwi.fws.gov/download.htm</a>
National Wetland Inventory points	F&WS	USDA – NCGC	county	vector points	<a href="http://www.nwi.fws.gov/download.htm">http://www.nwi.fws.gov/download.htm</a>

## 2.2. USDA datasets

The USDA datasets described in this document are those datasets that are produced and maintained by USDA that support the SCI.

### 2.2.1. Soils

The soil geospatial dataset category contains several separate soil-based datasets produced and maintained by USDA-NRCS. This dataset category contains both spatial and tabular soil datasets. Tabular datasets provide additional soil attributes and are linked to the features in the spatial datasets through a common field.

#### 2.2.1.1. SSURGO

The Soil Survey Geographic Data (SSURGO) dataset is the most detailed level of soil geographic data available. This dataset is produced by NRCS according to National Cooperative Soil Survey (NCSS) standards. NCSS is a collaborative effort between federal and non-federal agencies charged with the documentation of soil resources of private land holdings. Production of this dataset is performed through a combination of field surveys and aerial photograph interpretation. Soil scientists record soils along delineation boundaries to determine map unit composition by field traverses and transects. Aerial photographs comprise the field base map.

SSURGO data is distributed as four different datasets each representing different feature type geometry. The four feature types are areas (polygons), lines, points and linear features, as indicated in Table 1.

#### **2.2.1.2. Frozen soils**

The Conservation Reserve Program (CRP) 1990 Frozen Soils list is a county level database that contains fields extracted from the Map Unit Interpretations Record (MUIR) attribute database by a state soil scientist and verified against SSURGO map units. This database is an important part of the Conservation Reserve Program (CRP) eligibility process. At this time only a few counties located within the BPR pilot sites have received this database.

#### **2.2.1.3. National Soils Information System (NASIS)**

NASIS is a relational database that captures attribute information about soils. NASIS exports are often used to link additional soil attributes to the SSURGO product. This link is performed through the map unit symbol, a field common to both datasets. This field is actually a concatenation of two NASIS data elements called ALTMUKEY.

#### **2.2.1.4. Soil Survey Areas (SSA)**

A Soil Survey Area (SSA) is the boundary of a certified soil survey. The SSA is archived at NCGC. These boundaries are generated by NCGC in cooperation with states and are updated on a regular basis. These areas are represented as polygons and lines.

### **2.2.2. Common Land Unit (CLU)**

The CLU represents the smallest land unit delineation that has a permanent, contiguous boundary, common land cover management, a common owner and a common producer association. This dataset ties tabular program/producer information to a unique spatial location. The CLU defines the relationship between the customers (operators, tenants, land owners/managers, borrowers, etc.) and the land (farm, tracts, fields, pastures, etc.) and are often referred to as “fields”. This dataset represents CLU as polygons.

#### **2.2.2.1. CLU areas**

These boundaries represent the actual farm field boundaries that delineate the owner’s holdings. CLU boundaries are updated on a daily basis at Service Centers. CLU boundaries are represented as polygons.

### **2.2.3. Land use land cover**

BPR projects are currently underway to define and generate land use and land cover datasets for the partner agencies. Each agency has a slightly different definition for these two datasets and will coordinate on a common definition and appropriate land cover and land use designations. Land use and land cover designations will be added as an attribute to the CLU dataset.

### **2.2.3.1. Land cover**

FSA land cover classifications are defined in the 8-CM handbook as delineations according to agriculture, range land, water body, forest, built-up, barren and tundra. At this time it is unclear if FSA will continue to use these definitions, adopt the land cover definitions from another source or develop new land cover classifications. No decision has been made as whether the land cover will be an attribute of CLU or a separate dataset in itself.

### **2.2.3.2. Land use**

The detail agricultural use of land for a stated period of time and is associated with customer involved in this use. This dataset ties tabular program/producer information to a unique spatial location. The dataset represents these spatial locations as either a polygon or point. The polygons represent CLU's (fields) or subdivisions of CLU's, and points represent locations of commodities such as beehives, aquaculture (fish), container nurseries, etc. The tabular data includes information as to the commodity (crop) name and type, intend use, irrigation practice, production, production units, planting (beginning) date, harvest (ending) date, and customer and his share. This information will be reported several times per year (may be as many as seven revisions). Historical versions of this dataset will need to be maintained for temporal analysis and modeling.

### **2.2.4. Ortho Imagery**

Digital orthophotos combine the geometric qualities of a map with the image qualities of a photograph. Unlike a standard aerial photograph, relief displacement in orthophotos has been removed so that ground features are displayed in their true ground position. This allows for the direct measurement of distance, areas, angles, and positions. DOQs are produced through the National Digital Orthophoto Program (NDOP). USGS is the lead agency in this program and includes collaboration from FSA, NRCS and the United States Forest Service (USFS).

USDA currently uses several different formats of digital orthophoto imagery including a County Orthomosaic that is compressed and stored in MrSID format, Mosaicked DOQs (MDOQ) and DOQ Quarter Quadrangles (DOQQ). The compressed county and MDOQ formats are produced exclusively by USDA whereas the DOQQ is produced by USGS.

#### **2.2.4.1. Mosaicked Digital Orthophoto Quadrangle (MDOQ)**

Mosaicked DOQ datasets are produced by combining all four individual DOQQs that make up one 7.5-minute DOQ into one comprehensive coverage. During this process all visible seam-lines that exist between individual DOQQs are removed and overlapping areas are aligned and removed. Additional processing ensures that each MDOQ, when displayed with its adjacent MDOQ, is properly aligned. This is accomplished through edgematching techniques. Edgematching involves adjusting the image tone of as many as 16 edges to one common value to produce one MDOQ.

#### **2.2.4.2. County Orthomosaic**

Combining all the MDOQs that comprise a county into one single map produces a County Orthomosaic. The dataset is compressed using MrSID (Multi-resolution Seamless Image Database) compression techniques to reduce the file size. File sizes are often reduced as much as 80% without compromising the integrity of the original file. Refer to previous section

#### **2.2.4.3. Digital Orthophoto Quarter Quadrangles (DOQQ)**

The photography used to produce the 1-meter DOQs for the NDOP program is obtained primarily through the National Aerial Photography Program (NAPP). These black and white and color infrared aerial photography is quarter-quadrangle centered (3.75-minutes of latitude by 3.75-minutes of longitude in geographic extent) and taken at an aircraft altitude of approximately 20,000 feet above mean terrain. The scale of the NAPP photography is approximately 1:40,000. DOQs are produced by ortho-rectifying NAPP aerial photography with digital elevation models (DEM), ground control and camera calibration information.

#### **2.2.4.4. Ortho imagery catalog**

The ortho imagery catalog is a dBase format file that is used to coalesce all the individual DOQs and MDOQs for a particular geographic area, in this case a county. The catalog allows users to access a series of image themes as one logical image theme.

This dBase file is created by one of the distribution centers prior to shipping the datasets to a Service Center or State Office. The file consists of one record for each image. Each image has five attributes associated with it, the image name and location and four fields that contain the minimum and maximum extents for both the x and the y-axes. When loaded in a GIS, an image catalog serves as an index for the image tiles in a given view. When the display is refreshed or the view is moved, the image tiles in a view are automatically displayed. This catalog provides quick and easy access to the full-resolution DOQs.

#### **2.2.5. Climate**

Climate data is made available through a partnership between USDA-NRCS National Water and Climate Center (NWCC) and the Oregon Climate Service at Oregon State University. Climate datasets are derived using a system called PRISM (Parameter-elevation Regressions on Independent Slopes Model). PRISM is an expert system that employs point data and a digital elevation model (DEM) to generate gridded estimates of climate parameters. Since PRISM incorporates elevation data into its algorithms it can more accurately model the effects of terrain on climate. PRISM is used to map climate, precipitation and other indicators.

### **2.2.5.1. Climate - Precipitation**

Precipitation estimates indicate both the amount of precipitation as well as the spatial distribution of precipitation.

#### **2.2.5.1.1. Annual and monthly mean precipitation**

Annual and monthly mean precipitation datasets are available for the state, region (western US, central US and eastern US) and CONUS (Conterminous United States) levels of geography. These datasets are represented in vector polygon format and raster format. In the future only raster datasets will be available from NWCC for the entire US.

### **2.2.5.2. Climate – Temperature**

Originally developed for precipitation measures, PRISM has been generalized to calculate temperature estimates. Temperature datasets will be delivered in raster format.

#### **2.2.5.2.1. Annual and monthly mean temperatures**

Temperature estimates are available for mean maximum temperature, mean minimum temperature and mean temperature. Annual and monthly temperature datasets are available for CONUS, Alaska and other US regions (Hawaii, Pacific Basin Islands, and Puerto Rico). These datasets are represented raster format and the units are represented in degrees Celsius times 100.

### **2.2.5.3. Climate - Dewpoint**

Dewpoint is the temperature at which dew begins to form and varies with the humidity and temperature of the atmosphere.

#### **2.2.5.3.1. Annual and monthly mean dewpoint**

Annual and monthly mean dewpoint datasets are available for CONUS, Alaska and other US regions (Hawaii, Pacific Basin Islands, and Puerto Rico). Datasets represent mean maximum dewpoint and mean minimum dewpoint. These datasets are represented in raster format.

### **2.2.5.4. Climate – Other, derived climate variables**

In addition to temperature, precipitation and dewpoint, other climate variables have been derived and produced as datasets in a raster format. Not all of these datasets are immediately available.

#### **2.2.5.4.1. R-factor**

R-factor values from the RUSLE (Revised Universal Soil Loss Equation) exist for the CONUS and are represented in raster format.

#### **2.2.5.4.2. Growing degree days**

Annual and monthly mean growing degree-days at base temperature will be available for the CONUS in raster format.

#### **2.2.5.4.3. Last spring frost date**

The average last spring frost date for both 32 degrees Fahrenheit and 28 degrees Fahrenheit for the CONUS in raster format.

#### **2.2.5.4.4. First fall frost date**

Average first fall frost date for both 32 degrees Fahrenheit and 28 degrees Fahrenheit for the CONUS in raster format.

#### **2.2.5.4.5. January one in ten minimum temperature**

One in ten-year occurrence of extreme minimum temperature in January for the CONUS in raster format.

#### **2.2.5.4.6. Monthly mean snowfall**

Monthly mean snowfall datasets for the CONUS in raster format.

### **2.2.6. Environmental easements**

Landowners that volunteer to participate in conservation easement programs enter into an agreement with USDA to restore and protect that portion of land that falls within the boundary of a designated protective easement.

#### **2.2.6.1. Wetland Reserve Program (WRP)**

The WRP dataset is produced by USDA-NRCS Service Center offices according to the requirements outlined in the Food Security Act of 1985 and amended by the 1990 and 1996 Farm Bills. NRCS administers the program in conjunction with FSA. USDA WRP areas are part of a program designed to restore and protect wetlands on private property. Landowners that volunteer to participate in the WRP conservation easement program enter into an agreement with USDA to restore and protect wetlands. This dataset delineates the area bounding that portion of land under a protective easement. There are three different types of easements that can be arranged, permanent easements, 30-year easements and restoration cost-share agreements with a 10-year minimum duration. The emergency WRP dataset has not been developed at this time.

### **2.2.7. Hydrologic unit (11/14-digit)**

Hydrologic units (HU) are defined by NRCS as an area of land upstream from a specific point on a stream (designated as the mouth or outlet), which is defined by a hydrologic

boundary that includes all of the source area that could contribute surface water runoff directly and indirectly to the designated outlet point. The HU may also contain associated surface areas such as related contributing drainage, non-contributing, and transbasin diverted areas to form a landform associated with the certain designated outlet point. It may also have one or more of the same level hydrologic units completely contained within an upstream drainage area.

The Hydrologic Unit Code (HUC) is the specific identifier used to represent the specific subdivisions of the nations drainage at different levels of details or according to size. The code is a composite of four standard USGS levels of delineation (region, sub-region, accounting unit, and cataloging unit) and two NRCS levels of delineation (watershed and sub-watershed). The watershed level typically contains 40,000 to 250,000 acres while the sub-watershed level contains 10,000 to 40,000 acres, with a minimum of 3,000 acres. The USGS accounting unit maps to NRCS basin and the USGS cataloging unit maps to NRCS sub-basin). The entire code results in an 11/14-digit number identifying the hydrologic unit. The system of reporting hydrologic units in the 11/14-digit coding system will ultimately be replaced by a 10/12-digit system once the standards for the 10/12 system have been completed. This dataset is represented as polygons.

#### **2.2.8. Plants**

The PLANTS Database is a single source of standardized information about plants. This database focuses on vascular plants, mosses, liverworts, hornworts and lichens of the US and its territories. The PLANTS Database includes names, checklists, automated tools, identification information, species abstracts, distributional data, crop information, plant symbols, plant growth data, plant materials information, plant links, references, and other plant information. PLANTS reduces costs by minimizing duplication and making information exchange possible across agencies and disciplines. Currently, this data set only records the presence or absence of a plant in a county.

#### **2.2.9. Administrative areas**

Administrative areas are those datasets produced and maintained by USDA that delineate the coverage or servicing area of national level programs.

##### **2.2.9.1. Environmental Quality Improvement Program (EQIP)**

The Environmental Quality Incentives Program (EQIP) provides technical, educational, and financial assistance to eligible farmers and ranchers to address soil, water, and related natural resource concerns on their lands in an environmentally beneficial and cost-effective manner. The program provides assistance to farmers and ranchers in complying with Federal, State, and Tribal environmental laws, and encourages environmental enhancement. The purposes of the program are achieved through the implementation of a conservation plan that includes structural, vegetative, and land management practices on eligible land. Five- to ten-year contracts are made with eligible producers. Cost-share payments may be made to implement one or more eligible structural or vegetative practices, such as animal waste management facilities, terraces, filter strips, tree planting,

and permanent wildlife habitat. Incentive payments can be made to implement one or more land management practices, such as nutrient management, pest management, and grazing land management.

Currently, this dataset is developed at NCGC in conjunction with input received from state offices and is represented as polygons.

#### **2.2.9.2. Emergency Watershed Protection Program – Flood Plain (EWP-FP)**

The Flood Control Act of December 22, 1944 authorized the Secretary of Agriculture to install watershed improvement measures to reduce flood, sedimentation, and erosion damages; further the conservation, development, utilization, and disposal of water; and the conservation and proper utilization of land. This Act authorized eleven flood prevention watersheds. The NRCS and USFS carry out this responsibility with assistance from other bureaus and agencies within and outside USDA. Watershed protection and flood prevention work currently underway in small upstream watersheds all over the United States sprang from the exploratory flood prevention work authorized by the Flood Control Act of 1944, Public Law 534 and from the intervening 54 pilot watershed projects authorized by the Agriculture Appropriation Act of 1953. This dataset has not been developed.

#### **2.2.9.3. Farmland Protection Program (FPP)**

The Farmland Protection Program (FPP) provides funds to help purchase development rights to maintain productive farmland for agricultural use. Working through existing programs, USDA joins with State, Tribal, or local governments to acquire conservation easements or other interests from landowners. USDA provides up to 50 percent of the fair market easement value. To qualify, farmland must be part of a pending offer from a State, Tribe, or local farmland protection program; be privately owned; have a conservation plan; be large enough to sustain agricultural production; be accessible to markets for what the land produces; have adequate infrastructure and agricultural support services; and have surrounding parcels of land that can support long-term agricultural production. Depending on funding availability, proposals must be submitted by the government entities to the appropriate NRCS State Office during the application window. This dataset has not been developed.

#### **2.2.9.4. Field Office Service Areas (FOSA)**

Field Office Service Areas (FOSA) can be derived from the OIP dataset (see Section 2.2.9.6.) by aggregating the ZIP codes to the county level of geography and “dissolving” the boundary between the counties that are served by one Field Office.

#### **2.2.9.5. Soil and Water Conservation Districts (SWCD)**

Soil and Water Conservation District (SWCD) boundaries are created to assist with administrative mapping and planning. This is a national dataset. This dataset is currently developed at NCGC with input from the States and is represented as polygons.

#### **2.2.9.6. Office Information Profile (OIP)**

At this time, this dataset has not been distributed to the Service Centers. The OIP database contains over 4000 records for sites with over 10,000 organizational units (agency offices). This is a one-to-many relationship between OIP sites and the organizational units represented at that site. To create the geospatial component of the OIP database, the ZIP code for each county office was associated with a dataset of ZIP code centroids (see Section 2.3.7.7 ZIP code centroids). The resultant dataset is a point location dataset of OIP locations.

#### **2.2.9.7. Resource Conservation and Development (RC&D)**

The purpose of the Resource Conservation and Development (RCD) program is to accelerate the conservation, development and utilization of natural resources, improve the general level of economic activity, and to enhance the environment and standard of living in authorized RCD areas. It improves the capability of State, Tribal and local units of government and local nonprofit organizations in rural areas to plan, develop and carry out programs for resource conservation and development. The program also establishes or improves coordination systems in rural areas. Current program objectives focus on improvement of quality of life achieved through natural resources conservation and community development which leads to sustainable communities, prudent use (development), and the management and conservation of natural resources. Authorized RCD areas are locally sponsored areas designated by the Secretary of Agriculture for RCD technical and financial assistance program funds. NRCS can provide grants for land conservation, water management, community development, and environmental needs in authorized RCD areas. Currently, this dataset is under development at NCGC with input from State Offices.

#### **2.2.9.8. National Resources Inventory (NRI) database**

The National Resources Inventory (NRI) is a statistically based survey that has been designed and implemented using scientific principles to assess conditions and trends of soil, water, and related resources on non-Federal lands in the United States. The NRI is conducted by the USDA-NRCS, in cooperation with the Iowa State University Statistical Laboratory. The NRI provides scientifically valid, timely, and relevant information that is used to formulate effective agricultural and environmental policies and legislation, implement resource conservation programs, and enhance the public's understanding of natural resources and environmental conditions.

The NRI consists of over 800,000 randomly located monitoring sites on land throughout the United States. Data are collected annually and historical datasets can be used to analyze trends. This database can be used to produce NRI summary reports according to state, county, Major Land Resource Areas (MLRA), or hydrologic units (HU). There are no current plans to distribute this database to Service Centers.

### **2.2.9.9. NRI cross boundaries**

The NRI cross boundary layer contains the NRI mapping polygon boundaries for the conterminous United States. This is the smallest size and the densest distribution by which NRI data can be mapped due to security concerns surrounding the location of the monitoring sites located on private land. NRI point location data is aggregated and tabulated according to the cross boundary that it falls within. The NRI cross boundaries form at the coincidence of county boundaries, MLRA, eight digit hydrologic unit and Federal land datasets. This dataset is represented as polygons. There are no current plans to distribute this dataset to the Service Centers.

### **2.2.10. Digital Map Finishing (DMF)**

The Digital Map Finishing (DMF) is the final transfer of line work and text from which press ready negatives are prepared for soil survey publications. These vectors and annotation can be used in a GIS along with other datasets for cartographic output, display and analysis.

### **2.2.11. Rivers and Streams (DMF)**

This version of rivers and streams is a product of the Digital Map Finishing (DMF) division of NCGC. These datasets contain vector lines generated through line following techniques using digital ortho imagery as a base map. The linework has been symbolized for use in high quality cartographic output. In addition to the symbolization, the dataset includes annotation. This dataset represents rivers and streams as lines.

### **2.2.12. Railroads (DMF)**

This version of railroads is a product of the Digital Map Finishing (DMF) division of NCGC. These datasets contain vector lines generated through line following techniques using digital ortho imagery as a base map. The linework has been symbolized for use in high quality cartographic output. In addition to the symbolization, the dataset includes annotation. This dataset represents railroads as lines.

### **2.2.13. Roads (DMF)**

This version of roads is a product of the Digital Map Finishing (DMF) division of NCGC. These datasets contain vector lines generated using line following techniques using digital ortho imagery as a base map. The linework has been symbolized for use in high quality cartographic output. In addition to the symbolization, the dataset includes annotation. This dataset represents roads as lines.

### **2.2.14. Census**

The responsibility of conducting the Census of Agriculture was transferred to the Department of Agriculture, National Agricultural Statistics Service (NASS) in 1997. In

prior years, the Bureau of the Census conducted the agricultural census. This dataset has not been delivered to Service Centers for the BPR

#### **2.2.14.1. Census of agriculture**

An enumeration of the nations farms, farm population, and agricultural production, conducted by the Census Bureau every five years, in years ending in two and seven. The census of agriculture is a complete accounting of United States agricultural production and is the only source of uniform, comprehensive agricultural data for every county in the Nation. For census purposes, farms are defined as a place from which \$1,000 or more of agricultural products were produced and sold or normally would have been sold during the census year.

### **2.3. Non-USDA datasets**

The datasets listed in this section are not generally produced or maintained by USDA. However, the SCI has identified these datasets as important to the business at the Service Centers. These datasets are owned by other federal agencies and state and local governments and are generally obtained through a partnership arrangement or directly from an on-line distribution site or in some cases delivered to USDA production centers on CD-ROM.

#### **2.3.1. Air quality**

The EPA, Office of Air Quality Planning and Standards Information Transfer and Program Integration Division, located in Research Triangle Park, North Carolina maintains a program called AIRS (Aerometric Information Retrieval System). AIRS is as a computer-based repository of information about airborne pollution in the United States and various World Health Organization (WHO) member countries.

One of the subsystems of AIRS is the Air Quality System (AQS). Currently AQS consists of an interim database that contains air quality data that has been extracted from the production AQS database that resides on the EPA IBM mainframe. This interim database contains air quality data for the sites and monitors that existed in the mainframe database as of 01/11/2000. The raw and summary data will be refreshed approximately every three months.

AQS contains measurement of ambient concentrations of air pollutants and associated meteorological data. The data is collected by thousands of monitoring stations operated by EPA, national, state and local agencies. EPA uses this data to assess the overall status of the nation's air quality and to prepare reports to Congress as mandated by the Clean Air Act. EPA also uses the data to identify areas where improvements in air quality are needed. At this time, this database has not been acquired or delivered to USDA Service Centers.

### **2.3.2. Cadastral**

The cadastral geospatial dataset category contains geospatial data that pertains to land-based information or information that delineates the administrative areas owned by land management agencies.

#### **2.3.2.1. National Park Service boundaries**

The National Park Service (NPS) is the managing authority for all National Park Service units. National Parks include natural resource-based parks, cultural resource-based parks, recreation areas, trails, parkways, and historic battlefields. This dataset contains approximately 340 unit boundaries located throughout the US and its territories. This dataset was compiled through a variety of source material including park-based GIS databases, USGS 7.5-minute quadrangle maps, NPS legal descriptions and NPS land status maps. These boundaries represent the designated administrative boundary for each park and are depicted separately in a polygon dataset and a line dataset that simply represents the boundaries as linear features.

#### **2.3.2.2. Native American lands**

The Indian reservation boundary file used for the lower 48 states contains boundary data and location information that encompasses American Indian reservations, American Indian trust lands, Tribal designated statistical areas, and Tribal Jurisdiction statistical areas. The original dataset was derived from the Census Bureau. The EPA Office of Pollution Prevention and Toxics combined this dataset with an auxiliary dataset received from the Bureau of Indian Affairs (BIA) containing Alaskan Indian reservation boundaries. Although the information received from the BIA contained the same reservation boundaries for the lower 48 states as the data extracted from the Census Bureau, BIA provided no documentation or accompanying metadata. Therefore, the dataset distributed by EPA contains value-added information making it a more useful dataset for USDA business needs. This dataset represents Indian reservation areas and associated features as polygons.

#### **2.3.2.3. Public Land Survey System (PLSS)**

Digital Line Graph (DLG) data are digital representations of cartographic information. DLG's of map features are converted to digital form from maps and related sources. Intermediate-scale DLG data are derived from USGS 1:24,000-scale 7.5- by 15-minute quadrangle maps. Intermediate-scale DLG's are available in five categories: (1) Public Land Survey System; (2) boundaries (3) transportation; (4) hydrography; and (5) hypsography. All DLG data distributed by the USGS are DLG - Level 3 (DLG-3), which means the data contain a full range of attribute codes, have full topological structuring, and have passed certain quality-control checks.

The Public Land Survey System (PLSS) is a grid system by which units of land in the United States are described in relation to established north-south (township) and east-west (range) base lines. The resultant areas are referred to as congressional townships,

survey townships, or townships, and generally are six-square-mile units. Townships, in turn, may be subdivided into one square-mile units, called sections, which also may be subdivided further. Because of the spherical shape of the earth and irregularities in the original surveys, occasional corrections, which appear on maps as jogs or offsets, are introduced along specific township and range lines. This dataset represents PLSS boundaries as polygons.

### **2.3.3. Census**

The census dataset category contains both spatial datasets and attribute databases that are collected as part of the US Census Bureau decennial census. Spatial datasets are extracted from a product called TIGER (Topologically Integrated Geographic Encoding and Referencing). TIGER is a digital database of geographic features such as roads, railroads, political boundaries and streams that are packaged according to county and exist for the entire United States and territories. The database describes these geographic features according to their location in latitude and longitude, the name, the type of feature, address ranges for most streets, the geographic relationship to other features, and other related information. This product is held in the public domain and can be processed for use with GIS.

Certain TIGER datasets, such as county boundaries, can be linked to Census Summary Tables to reveal the demographic characteristics of that area. Summary tables contain detailed social, economic and housing information including population characteristics, housing statistics and information concerning industry and business. Once the database is linked to a geographic entity, the attributes found in the Summary Tables can be used for spatial analysis, query and display.

Census geography and attributes play a large role in the SCI. This information is used by RD to locate that part of the population that is potentially under-served by RD programs, to geocode loan applicants and recipients according to their address and to summarize loan request and distribution information for Congressional reports. FSA and NRCS often use Census geography datasets, such as streets and hydrography, to provide a spatial reference when used in conjunction with orthophotos, soils and CLU data for visualization and map production.

#### **2.3.3.1. Census block demographics**

Census demographics at the block level include a summary of the basic composition of each housing unit and its occupants, the household relationship, sex, race, age, marital status, and Hispanic origin as well as housing questions concerning the number of units in each structure, the number of rooms per unit, tenure, and value. Census demographics are gathered during the decennial census through the completion of a questionnaire sent to all households. This information is maintained in a tabular database format.

### **2.3.3.2. Census blocks**

The smallest entity for which the Census Bureau collects and tabulates decennial census information is census blocks. Census blocks are bounded on all sides by visible features such as a river, street, railroad track, power line, shoreline, fence, ridge or cliff and non-visible features such as a city or county boundary or a property line as depicted on Census Bureau maps. This dataset is represented as polygons.

### **2.3.3.3. Coastal water**

This dataset contains coastal waters that bound the Atlantic and Pacific Oceans and the Great Lakes. Only counties that have a coastal boarder will have an associated coastal water dataset. This dataset depicts coastal waters as areas (polygons).

### **2.3.3.4. Census transportation planning package areas**

Census Transportation Planning Package (CTPP) areas are special purpose geographic entities designed to assist in transportation planning. Traffic Analysis Zones can be aggregated to the CTPP level of geography. There were approximately 300 CTPPs in the 1990 Census and approximately 200,000 TAZs. These areas are represented as polygons.

### **2.3.3.5. Census block group demographics**

Demographics aggregated to the block group level of geography are extracted from the US Census Bureau Summary Tape Files (STF) 1A and 3A, reflecting the most commonly used demographic characteristics. The data captured at this level is not available at the Census block level of geography due to concerns about the privacy of the public.

Demographics captured at this level include detailed information concerning the number of persons by population, family and household; their sex, ethnicity, age, marital status, family type, citizenship, education, employment, commute, income and the physical characteristics of their housing unit, such as plumbing and number of rooms. This dataset is maintained in a tabular database format.

### **2.3.3.6. Census block groups**

A combination of census blocks that is a subdivision of a census tract or block numbering area is a census block group. A block group consists of all blocks whose numbers begin with the same digit in a given census tract or block numbering area. The block group is the lowest level of geography for which the Census Bureau has tabulated sample data in the 1990 census. Sample data is detailed information provided by the one in six households that receive a "long form" during the decennial census. Block groups generally contain between 250 and 550 housing units and usually cover a contiguous area. This dataset is represented as polygons.

### **2.3.3.7. Key geographic locations**

A point feature that represents a particular type of landmark that possesses residential or economic significance, such as apartment buildings, shopping centers, airports and factories is a key geographic location. These locations are generally more widely known by their name than their address, which is often used in conjunction with the address to help geocode its location. This dataset is represented as polygons.

### **2.3.3.8. Landmark areas**

Cartographic landmarks are special features used to help enumerators during field operations and include airports, cemeteries, parks and educational facilities. Landmark areas are represented as polygons. Landmarks that have addresses and ZIP codes are called key geographic locations.

### **2.3.3.9. Railroads**

A railroad is a linear feature that can represent a railroad, a street car line or other car line depending upon the source material that was available and used during the update of the feature layer. These features are represented as lines.

### **2.3.3.10. Utility lines**

This dataset contains linear features including power transmission lines, pipelines and miscellaneous ground transportation lines, such as aerial tramways, monorails and ski lifts. This dataset is represented as lines.

### **2.3.3.11. Landmark lines**

Cartographic landmarks are special features used to help enumerators during field operations and include airports, cemeteries, parks and educational facilities. Linear landmark features are represented as lines if they do not possess enough thickness to be represented as an area, an example may be an airport runway. Landmarks that have addresses and ZIP codes are called key geographic locations.

### **2.3.3.12. Ridge and fence lines**

Ridge and fence lines are man made and physical features that are used as boundaries to separate property or reference a particular area. For example fence lines (man made features) surrounding a military installation or prison separates the reservation from civilian land. A ridge line (physical feature) is the highest elevation of a mountain. These topographic features are often used to delineate boundaries where no other visible features can be identified. Both of these feature types are represented as lines in this dataset.

### **2.3.3.13. Streams**

These are a category of hydrologic features that represent naturally flowing water, and are often referred to as streams and/or rivers. These features are represented as lines in this dataset.

### **2.3.3.14. Unclassified lines**

Unclassified lines are those non-visible boundaries (features) that typically delimit tabulation entities, property areas and legal and administrative entities. This type of feature is defined when it does not follow a visible feature such as a road, stream or ridge line. Unclassified boundaries have an unknown classification or have not been classified elsewhere and include tabulation boundaries reserved for future census delineation, map edges no longer in use, lines representing a line of sight between two points and offset, corridor and superceded legal and administrative boundaries. This dataset is represented as lines.

### **2.3.3.15. Landmark points**

Cartographic landmarks are special features used to help enumerators during field operations and include airports, cemeteries, parks and educational facilities. Landmark point features may actually represent the centroid of a landmark area or an area too small to depict a bounded area. Landmarks that have addresses and ZIP codes are called key geographic locations. This dataset is represented as points.

### **2.3.3.16. Minor civil divisions**

A MCD is a governmental unit that is the primary legal subdivision of a county in 28 states, created to govern or administer an area rather than a specific population. There are several types of MCDs that are identified by a variety of terms, such as town, township, and district, and include both functioning and nonfunctioning governmental units. Some MCDs also represent local, general purpose governmental units. This dataset depicts the delineation of MCDs for the 1990 Census and represents the areas as polygons.

### **2.3.3.17. Minor civil divisions**

See definition of MCDs in Section 2.3.3.17. An update to the 1990 version of the MCD boundary dataset. This dataset is represented as polygons.

### **2.3.3.18. Places – blocks**

A place is a concentration of population either legally bounded as an incorporated place, or identified by the Census Bureau as a census designated place. Incorporated places have political/statistical descriptions of borough (except in Alaska and New York), city, town (except in New England, New York, and Wisconsin), or village. These places are

aggregated to the census block level of geography. This dataset is represented as polygons.

#### **2.3.3.19. Places - tracts**

See definition of places in Section 2.3.3.19. This dataset represents places aggregated to the census tract level of geography. This dataset is represented as polygons.

#### **2.3.3.20. Elementary school districts**

A school district inclusive of either kindergarten or the first grade through the eighth or the ninth grade is an elementary school district. For data tabulations from the 1980 and 1990 decennial censuses, this term includes both elementary and intermediate/middle districts. This dataset represents districts as polygons.

#### **2.3.3.21. Middle school districts**

A middle school district is a school district inclusive of the fifth through eighth grade, the sixth through the ninth grade, the seventh and eighth grades, or the seventh through ninth grade. The Census Bureau did not provide a separate identification of intermediate/middle school districts in its 1990 census tabulations. This dataset represents districts as polygons.

#### **2.3.3.22. Secondary school districts**

Secondary school districts are school district inclusive of only high school, either the ninth through the twelfth grades or the tenth through the twelfth grades. This dataset represents districts as polygons.

#### **2.3.3.23. Unified school districts**

A school district inclusive of kindergarten through twelfth grade is a unified school district. This dataset represents districts as polygons.

#### **2.3.3.24. Coastal shorelines**

This dataset contains the outlines for coastal waters that bound the Atlantic and Pacific Oceans and the Great Lakes. Only counties that have a coastal boarder will have an associated coastal water dataset. This dataset depicts coastal shoreline features as lines.

#### **2.3.3.25. Alternate street names**

This dataset is a database containing additional street names that can be linked to the streets dataset to provide more information about the feature in addition to the it's primary name. Alternate names may be local names or route numbers. This dataset is maintained in a tabular database.

### **2.3.3.26. Streets**

The streets dataset includes primary roads (includes both divided or undivided US highways and some state and county highways), secondary (includes mostly state highways that are undivided), local, neighborhood and rural roads (designated for local traffic and usually consists of one lane in each direction, rural roads connecting small towns or parkways designated as scenic), as well as trails designated for vehicular or pedestrian use. Most of the nations roads are classified within the category of local, neighborhood and rural. This file contains attributes such as road type and feature name, address ranges for the left and right side of the roadway and ZIP code information that can be used to geocode features to the address level of geography and represent them as point locations. Geocoding is based on interpolating an address within the address ranges assigned to the streets dataset. The streets dataset represents streets as lines.

### **2.3.3.27. Census tract demographics**

Demographics aggregated to the Census Tract level of geography are extracted from the US Census Bureau Summary Tape Files (STF) 1A and 3A, reflecting the most commonly used demographic characteristics. The data captured at this level is not available at the Census Block level of geography.

Demographics captured at this level include detailed information concerning the number of persons by population, family and household; their sex, ethnicity, age, marital status, family type, citizenship, education, employment, commute, income and the physical characteristics of their housing unit, such as plumbing and number of rooms. This dataset is maintained as a tabular database.

### **2.3.3.28. Traffic analysis zones**

Traffic analysis zones are a special geographic entity delineated by a metropolitan planning organization for tabulating transportation statistics from the decennial census. The Census Bureau collected and tabulated data for approximately 200,000 Traffic Analysis Zones (TAZs) within approximately 300 CTPP areas for the 1990 decennial census.

### **2.3.3.29. Census tracts**

A census tracts is a small, relatively permanent statistical subdivision of a county in a metropolitan area or a selected non-metropolitan county, delineated by a local committee of census data users for the purpose of presenting decennial census data. Census tract boundaries normally follow visible features, but may follow governmental unit boundaries and other non-visible features in some instances, however they always nest within county boundaries.

Census Tracts are designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time they are established and usually contain between 2,500 and 8,000 inhabitants. They may be split by any sub-county geographic entity. This dataset is represented as polygons.

#### **2.3.3.30. Urban areas**

An Urban Area consists of at least one central place and together with the adjacent densely settled surrounding has a minimum population of 50,000 people. The surrounding territory generally consists of an overall population density of at least 1,000 people per square mile on a continuous basis. Urban areas are represented as polygons.

#### **2.3.3.31. Urban and rural outlines**

Urban, as defined by the Census Bureau consists of all territory and population in an Urban Area and in the urban portion of places with 2,500 or more people located outside of the Urban Area. The Census Bureau distinguishes the urban and rural population within incorporated places whose boundaries contain large, sparsely populated, or even unpopulated area. These extended cities have either 25 percent of their land area, or at least 25 square miles, classified as sparsely settled. The sparsely settled area must consist of at least one group of one or more contiguous census blocks. Each group must be at least five square miles in area and have an overall population density of less than 100 people per square mile. Urban and rural outlines are represented as polygons, each polygon is flagged with an "U" or an "R" depending on its classification.

#### **2.3.3.32. Voting districts**

Voting districts are any of a variety of areas, such as election districts, precincts, legislative districts, or wards established by states and local governments for voting purposes. The 1990 census term voting district replaces the 1980 term election precinct. This dataset represents voting districts as polygons.

#### **2.3.3.33. Land and water outlines**

This layer consists of two features covering the entire area of a county. One feature is of all land areas, including mainland and islands when present. The other feature is all water areas, which may include coastal waters, lakes, reservoirs, and any other water bodies. Both of these features are represented as polygons.

#### **2.3.3.34. Water bodies**

This dataset contains only those features that have been classified as water areas and includes coastal waters, lakes, reservoirs, and any other water bodies. These features are represented as polygons in this dataset.

### **2.3.4. Elevation**

Elevation data is obtained from DEM which are digital records of terrain elevations for ground positions at regularly spaced horizontal intervals. DEM's are developed from stereo models or digital contour line files derived from USGS topographic quadrangle maps. This dataset depicts continuous elevation as lines.

#### **2.3.4.1. Contours**

This dataset contains topographic contours that represent the physical relief of the earth's surface in the form of lines. Each line follows a continuous level of the surface of the earth. Contour lines are derived from USGS topographic quadrangle maps. This dataset is extracted from the USGS Digital Line Graph (DLG) product and is represented as lines.

#### **2.3.4.2. Geodetic survey monuments**

National Geodetic Survey Monuments show control stations established throughout the state by government as a source of primary horizontal survey control. A total of 22,277 monuments are captured in this coverage, with a NAD83 Datum. These primary control monuments are used to establish the location of federal, state, and private surveys. This dataset represents survey monuments as points.

### **2.3.5. Endangered habitat**

Endangered habitat areas are designated as specific geographic areas essential for the conservation of threatened or endangered species. These areas may require special management considerations and are not necessarily occupied by a threatened or endangered species at the time the area is designated. At this time habitat conservation data is available for the Pacific Northwest only and has not been delivered to any Service Centers in the BPR pilot.

### **2.3.6. Geographic Names Information System (GNIS)**

The Geographic Names Information System (GNIS) database contains information about physical and cultural geographic features in the United States, including the Federally recognized name and location. This database contains the latitude and longitude position of the feature as extracted from a topographic quadrangle. These coordinates are used to generate a point location for that particular feature. The GNIS is the Nation's official repository of domestic geographic name information and is often used for reference purposes.

### **2.3.7. Government units**

Government units are areas that have local, state or federal jurisdictional significance. These datasets are obtained from a variety of sources including the Census Bureau and USGS. This dataset is represented as polygons.

### **2.3.7.1. Administrative boundaries**

Administrative boundaries are extracted from the USGS 1:24,000 DLG product and include delineations for both national and state owned lands such as forests and parks. There are two datasets that represent administrative boundaries, boundary outlines represented as lines and boundary areas represented as polygons and points.

### **2.3.7.2. Congressional districts**

A congressional district is an area established by state officials or the courts for the purpose of electing a person to the US House of Representatives. Each state must contain, as nearly as possible, an equal number of inhabitants in each Congressional District. Congressional districts are represented as polygons.

### **2.3.7.3. Cities**

This is a dataset containing the locations of populated places or areas with clustered or scattered buildings and a permanent human population. Cities are represented as points in this dataset.

### **2.3.7.4. County boundaries**

This dataset contains a governmental unit delineating the primary legal subdivision of every state except Alaska and Louisiana. There are two county boundary datasets, one representing county boundary outlines as lines and the other as polygons.

### **2.3.7.5. Man made features**

Man made features are those features that were added to the landscape, such as fence lines. There are two man made feature datasets, one to represent features as a line and one to represent man made areas as polygons.

### **2.3.7.6. State boundaries**

This dataset is the governmental unit that is the primary legal subdivision of the United States. There are two state boundary datasets, one representing the boundary outlines as lines and other representing states as polygons.

### **2.3.7.7. ZIP code centroids**

A five-, seven-, nine-, or eleven-digit code assigned by the US Postal Service to a section of a street, a collection of streets, an establishment, structure, or group of post office boxes for the delivery of mail. This dataset represents each ZIP code area as a point using the geographic center of the area to represent the ZIP code.

### **2.3.8. Hydrography**

The hydrography dataset category contains datasets that depict naturally occurring and constructed bodies of water, paths through which water flows, and related entities. These datasets also include feature classification and other characteristics, delineation, geographic name, position and related measures.

#### **2.3.8.1. Damsites**

This dataset provides locations for 75,187 dams in the conterminous United States. The National Inventory of Dams was originally developed by the U.S. Army Corps of Engineers (COE) and the Federal Emergency Management Agency (FEMA) to help track dam related problem areas. This database includes the age of the dam, number of people living downstream, and some inspection information including nearest town, a description, size, reservoir capacity, owner and the regulatory oversight agency. This dataset represents dams as point locations.

#### **2.3.8.2. Flood hazard areas**

The Federal Emergency Management Agency (FEMA) publishes official maps where boundaries of flood areas, mudslides (i.e., mudflow) and related erosion areas having special hazards have been designated. The most widely distributed flood map product is the Flood Insurance Rate Map (FIRM). This information is based on historic, meteorologic, hydrologic, and hydraulic data, as well as open-space conditions, flood control works, and development. FEMA uses this information in conjunction with field studies to delineate areas that are prone to flooding and that have a one percent or greater chance of being equaled or exceeded during any given year. This type of flood is referred to as a base flood. A base flood has a 26-percent chance of occurring during a 30-year period. The base flood is a regulatory standard used by Federal agencies, and most states, to administer floodplain management programs, and is also used by the National Flood Insurance Program as the basis for insurance requirements nationwide. This dataset is not available digitally yet for all counties. At this time approximately 1300 counties are available. Flood hazard areas are represented as polygons.

#### **2.3.8.3. Rivers and streams**

The rivers and streams dataset contains EPA Reach Files which is a series of hydrographic databases of the surface waters of the continental United States and Hawaii. A key characteristic of the Reach files is attributes that define the connected stream network including flow direction. This attribute-based connectivity enables the Reach Files to provide hydrologic ordering of stream locations using reach codes (what is upstream and downstream of a given point in the stream network) as well as network navigation proceeding in either the upstream or downstream direction. This dataset is represented as lines.

### **2.3.9. Hydrologic unit (8-digit)**

This dataset depicts the hydrographic boundaries of major river basins and the numeric codes assigned to each river basin. The dataset was prepared through a cooperative project between the USGS and the US Water Resources Council. River basins are delineated areas that have drainage area greater than 700 square miles. The HUC is the specific identifier used to represent the specific subdivisions of the nations drainage at different levels of details or according to size. The code consists of a composite of four standard USGS levels of delineation. These two digit delineations are region, sub-region, accounting unit and cataloging unit. This dataset is represented as polygons.

### **2.3.10. Land Use Land Cover (LULC)**

LULC data are derived from thematic overlays registered to 1:250,000-scale base maps and a limited number of 1:100,000-scale base maps. LULC data provide information on urban or built-up land, agricultural land, range land, forest land, water, wetlands, barren land, tundra, and perennial snow or ice. Associated maps display information in five data categories: (1) political units, (2) hydrologic units, (3) census county subdivisions, (4) Federal land ownership, and (5) State land ownership.

#### **2.3.10.1. Multi-Resolution Land Cover (MRLC)**

The Multi-Resolution Land Characteristics (MRLC) is a land cover characteristics data base using Landsat Thematic Mapper (TM) data. The project is the joint effort of several US Government agencies whose goal is to provide multi-resolution land cover data of the conterminous United States for monitoring and assessing environmental change at the regional level in a very cost effective manner. The project consists of three separate inventories of associated data: MRLC Landsat TM Basic, MRLC Landsat TM Derivative, and MRLC Landsat TM. At this time, this dataset has not been delivered but has been identified as an important USDA dataset.

#### **2.3.10.2. Land use land cover**

Initial land use land cover maps were compiled through the manual interpretation of aerial photographs acquired from NASA (National Air and Space Administration) high-altitude missions and other aerial sources. Secondary sources from earlier land use maps and field surveys were also incorporated into the LULC maps as needed. At a later time, the LULC maps were digitized to create a national digital LULC database. The evolution of this process resulted in the creation of the Geographic Information Retrieval Analysis System (GIRAS). Initial source preparation involves the transfer of field survey information, photo classification detail and associated line work to a base map for digitization. This may be done through stereo compilation or monoscopic transfer techniques. Adjacent maps are also checked to ensure continuity. The maps are digitized and the appropriate classification codes are assigned for processing through GIRAS and checked for accuracy.

This dataset was mapped and coded using the Anderson classification system, which is a hierarchical system of general (level 1) to more specific (level 2) characterization. Some agencies have taken this to a level 3 classification, but this has not been done in the GIRAS series. This dataset represents land use land cover areas as polygon feature types.

### **2.3.10.3. Barren land**

Barren land classifications are extracted from the DLG dataset and represent places composed of bare rock, sand, silt, gravel or other earthen material with little or no vegetation and having limited ability to support life. This dataset represents barren land areas as polygons.

### **2.3.10.4. Surface cover**

Surface cover consists of both vegetative and non-vegetative surface cover. Vegetative surface cover this includes woods, scrub, orchards and vineyards. Non-vegetative surface cover includes the natural surface of the Earth such as lava, sand and gravel features. This dataset represents surface cover as polygons and is extracted from the DLG data product.

### **2.3.11. Map indexes**

This dataset category is reserved for map indexes that are used for reference purposes or cartographic display, and can be used to link a geographic area to an image or map associated with that location.

#### **2.3.11.1. Photo index**

NAPP photographs are acquired over pre-determined exposure points within a 7.5 minute quadrangle map. The principal points of exposure divide the quad into quarter sections. Each section involves a 3.75 minute sub-quadrangle. The NAPP flight lines run north-south down the middle of these quarter-quads. The photo index is a series of point locations representing the approximate centroid of each DOQQ produced by USGS through the NDOP program. This dataset represents each location as a point feature type.

#### **2.3.11.2. Quadrangle boundaries**

Quadrangle boundaries represent the outlines of the 7.5-minute (1:24,000) and 3.75-minute (1:12,000) topographic quadrangles. Two datasets are delivered in this category, a dataset of boundaries represented as areas (polygons) and a dataset of the boundaries represented as lines.

### **2.3.12. Topographic images (Digital Raster Graphs - DRG)**

Digital Raster Graphs (DRG) are scanned images of a USGS 7.5-minute quadrangle paper maps. The scanned image consists of the complete map area and map collar containing all the map parameters and metadata information. Maps are scanned at a

minimum resolution of 250 dots per inch. Once the map has been scanned, it is georeferenced to the UTM (Universe Transverse Mercator) projection system. The horizontal positional accuracy and datum of the DRG match that of the original paper product. These maps lend themselves well to the GIS environment since users that are already familiar with the paper map form are comfortable making the transition from paper format to digital format using the geographical reference grids, place names and features found on the maps

#### **2.3.12.1. Topographic image catalog**

The topographic image catalog is a dBase format file that is used to organize all the individual DRGs within a particular geographic area, in this case a county. The catalog allows users to access a series of image themes as one logical image theme.

This dBase file is created by one of the distribution centers prior to shipping the datasets to a Service Center or State Office. The file consists of one record for each image. Each image has five attributes associated with it, the image name and location and four fields that contain the minimum and maximum extents for both the x and the y-axes. This catalog provides quick and easy access to the DRGs.

#### **2.3.12.2. DRG compressed**

The compressed DRGs dataset consists of all the enhanced DRGs for a particular county mosaicked together into one comprehensive image theme. The dataset contains all the DRGs that fall within that county and is compressed using the MrSID compression technique to reduce the file size while maintaining the integrity of the image.

#### **2.3.12.3. DRG standard**

The standard version of DRGs are simply those datasets delivered in the format as received directly from USGS and described in 2.3.12.

#### **2.3.12.4. DRG enhanced**

Enhanced DRGs are created from the original DRGs obtained from USGS. Each DRG map collar is removed so only the topographic map remains. This allows the user to display adjacent DRGs without the map collar overlapping part of the adjacent map. These datasets are delivered in TIF format.

### **2.3.13. Transportation**

Transportation datasets includes any dataset that can be considered part of the infrastructure responsible for moving vehicles, trains or power. Transportation datasets are extracted from DLG's. DLG's contain vector line data, such as roads and streams, digitized from USGS topographic maps. DLG's offer a full range of attribute codes, are highly accurate, and are topologically structured, which makes them ideal for use in geographic information systems (GIS).

### **2.3.13.1. Utility lines**

This dataset contains linear features including power transmission lines, pipelines and miscellaneous ground transportation lines, such as aerial tramways, monorails and ski lifts.

### **2.3.13.2. Railroads**

This dataset contains linear features that can represent a railroad, a street carline or other carline depending upon the source material that was available and used during the update of the feature layer. These features are represented as lines.

### **2.3.13.3. Roads**

The roads dataset includes primary roads, includes both divided or undivided US highways and some state and county highways. This dataset is extracted from DLGs and represents roads as lines.

### **2.3.14. Wetlands – National Wetlands Inventory (NWI)**

The NWI is produced and maintained by the US FWS. This dataset documents the characteristics, extent, and status of the nation's wetlands and deepwater habitats. Federal, State, and local agencies, academic institutions, US Congress, and the private sector use this information. The Emergency Wetland Resources Act of 1986 directs the Service to map the wetlands of the United States. The first comprehensive estimate of the nation's wetlands was produced in 1982 and updated in 1990. National level updates are scheduled every 10-year intervals and the next update is scheduled to occur in 2000. This dataset represents wetlands as polygons, lines, linear features as lines and points.

## **3. Dataset Physical characteristics**

To accurately determine the management requirements of each dataset presented in Table 2-1 and Table 2-2, a complete description of the physical characteristics for each dataset must be documented including some datasets that have not been delivered but are anticipated to be delivered shortly. The subsequent sections provide a brief discussion on the physical characteristics to be measured and what impact that attribute has on the management of geospatial datasets. Some physical characteristics are very generic and do not need to be detailed on a dataset by dataset basis. Those characteristics that tend to vary according to each dataset or possess a quantitative measurement, such as size, are detailed on a dataset by dataset basis in Table 2-1 and Table 2-2.

### **3.1. Source**

The source of a geospatial dataset as described in this document includes the data producer, data publisher and the actual URL from where the dataset can be downloaded if

available. The producer is the party responsible for creation of the original data source product. Producers include USDA, other federal agencies, state and local agencies and other organizations that are charged with geospatial dataset production. Distributors are those agencies that are responsible for getting the geospatial datasets to the users. At this time USDA agencies are responsible for the distribution of all geospatial datasets. The URLs listed in Table 2-1 and Table 2-2 document, when available, the location from which the dataset can be downloaded. In all cases, the URL points to a location hosted by the producer or designated as a distribution site by the producer. However, not all sources conform to the specific requirements outlined by the SCI. These datasets may require post processing techniques to convert the dataset into a different projection or add additional attributes.

### **3.1.1. Alternative sources**

Some datasets can be obtained from more than one source including other federal agencies, state and local governments, data resellers and commercial firms that distribute data at no or low cost. Some examples of alternate sources for geospatial datasets are presented below.

### **3.2. FGDC – NSDI**

The Federal Geographic Data Committee (FGDC) is an interagency committee, organized in 1990 under OMB Circular A-16 to promote the coordinated use, sharing, and dissemination of geospatial data on a national basis. FGDC is responsible for the development of the National Spatial Data Infrastructure (NSDI), an organization composed of 17 federal agencies, state, local and tribal governments, the academic community and the private sector. NSDI is charged with the coordination of policies, standards and procedures that help organizations to cooperatively produce and share geospatial data. Designed to reduce the cost of geospatial data collection and minimize duplication among federal agencies, the NSDI serves an important role in the management of geospatial datasets important to USDA business areas. Circular A-16 can be viewed on-line at the following URL:

<http://www.whitehouse.gov/OMB/circulars/a016/a016.html>

FGDC sponsors a geospatial data discovery system that can be used to search for specific geospatial dataset requirements. The NSDI Clearinghouse is designed to provide access to digital spatial data through metadata stored in a searchable catalog. The Clearinghouse supports user requests to locate specific datasets covering a particular geographic area. Once the user has identified the dataset of interest through the metadata catalog, they can opt to view a browse graphic of the dataset, if available, or download it from the sponsoring agency data server. The Clearinghouse strongly encourages registered sites to incorporate URLs into the metadata records that allow users to view the browse graphics and download the dataset if desired.

NCGC is currently registered as a gateway and node on the NSDI Clearinghouse. This enables users to search and discover data assets maintained at NCGC and at other nodes. Where digital data are too large to be made available through the Internet or the data

products are made available for sale, linkage to an order form can be provided in lieu of a data set. Through this model, metadata provides low-cost advertising for providers of spatial data, both non-commercial and commercial, to potential customers via the Internet.

### **3.2.1. USGS datasets**

Some federal agencies, such as the USGS National Mapping Division (NMD) actually solicit private organizations to become distribution nodes for their products. At this time, USGS maintains over 2,000 Business Partners authorized to resell USGS topographic maps, digital lines of cartographic data products and aerial photographs. The most current list of USGS Business Partners can be viewed at the following URL:

<http://edcwww.cr.usgs.gov/buspartners/digital/currentpart.html>

Additionally, USGS provides links to non-USGS web sites that provide data on-line at no cost to the public. This information is organized by state and should be referred to as USDA begins to create data sharing partnerships for the distribution of low-cost geospatial datasets. The list of state-based organizations can be viewed at the following URL: [http://mcmcweb.er.usgs.gov/drg/free\\_drg.html](http://mcmcweb.er.usgs.gov/drg/free_drg.html)

### **3.2.2. Census Bureau datasets**

TIGER-line files are available from the Census Bureau for a small fee. States are grouped into regions and each individual CD ROM is sold for \$70. The complete 7-disc set is \$490 and includes complete technical documentation. The data consists of a series of line segments that represent legal and statistical boundaries and contain 17 different record types that are compressed and stored in fixed length ASCII format.

Most commercial off-the-shelf (COTS) GIS include routines that process the uncompressed files into their own native structure. Alternatively, users may opt to save time, money and effort on the processing and purchase the desired datasets from one of several vendors. The Census Bureau's main geospatial product, TIGER, requires a substantial amount of processing for it to be used with COTS GIS. As a result, many vendors have gone through the effort to process the raw TIGER line files into a variety of GIS formats and resell the processed products at a profit. The cost to purchase these products from vendors varies depending on the vendor and the level of post processing clean-up and "value-adding" that was performed.

Value adding includes the addition of address ranges to streets files to improve geocoding matches, correcting lines that do not connect to improve the accuracy of jurisdictional boundaries and features, moving lines that are misplaced to improve precision and adding detail to generalized boundaries such as coastlines to improve the cartographic quality. An example of a commercial vendor that resells TIGER datasets is GDT (Geographic Data Technology). GDT's product line can be viewed at the following URL:

<http://www.geographic.com/index.cfm>

Due to the relatively low cost, wide availability and widespread application of TIGER datasets for a variety of applications at all levels of government from national to local

levels, many jurisdictions have processed these files and in many cases make them available on-line at no cost or low-cost. View the following URLs to locate no and low-cost Census datasets: <http://www.census.gov/geo/www/tiger/vendors.html> or [http://factfinder.census.gov/java\\_prod/dads.ui.homePage.HomePage](http://factfinder.census.gov/java_prod/dads.ui.homePage.HomePage)

### **3.2.3. State**

As state-level GIS organizations get more organized and begin to disseminate GIS datasets to the public, there will be less emphasis on the need for USDA to purchase datasets and more of an emphasis to enter into partnership agreements with agencies to share each others data assets at no cost to either party.

Examples of state GIS clearinghouses that provide data at no cost include the Georgia Spatial Data Infrastructure (<http://www.gis.state.ga.us/>), New York State GIS Clearinghouse (<http://www.nysl.nysed.gov/gis/>), Oregon's State Service Center for Geographic Information Systems (<http://www.sscgis.state.or.us/>) and Pennsylvania GIS Consortium (<http://www.pagis.org/>). Additionally, some states such as New York have enacted GIS Data Sharing Cooperatives where users enter into an agreement with for the purpose of improving access to GIS datasets among its members.

### **3.2.4. Regional, county, and local agencies**

As more regions, counties and local agencies begin to digitize their own tax parcel datasets, transportation networks and school district boundaries, datasets will become more accurate and provide a higher level of detail than is currently available from national organizations, state agencies and commercial data vendors.

Examples of regional GIS distribution centers include the South Florida Water Management District (<http://www.sfwmd.gov>). The SFWMD is a special taxing district that consists of 16 counties and has the authority to collect property taxes from the landowners within that region. The agency provides flood control protection and water management supply protection to residents that live and work within that region. The SFWMD can be located on-line at the following URL: <http://www.sfwmd.gov/curre/requests/requests.htm>

An example of a county government agency that lists geospatial datasets on-line is Loudoun County, VA. The URL to the county GIS Web site is: [http://www.co.loudoun.va.us/omagi/docs/metadata\\_/default.htm](http://www.co.loudoun.va.us/omagi/docs/metadata_/default.htm)

Also the City of Sacramento has an on-line GIS located at the following URL: <http://citymaps.cityofsacramento.org/GISAPPS2/GenMap/>

### **3.2.5. Commercial vendors**

As GIS technology has advanced over the past ten years and entered into more mainstream business areas such as commercial and residential real estate, business geographics, health care, insurance and telecommunications, more and more commercial

vendors have realized that a lucrative market for geospatial data exists in non-traditional GIS areas. Business applications are ill-equipped to produce, update and manage large amounts of geospatial data, therefore they tend to purchase data from commercial vendors that is ready for use “out-of-the-box” in any of the common COTS GIS formats. In some cases it is cheaper, easier and faster for government agencies to pursue this method of data acquisition where the resources are not available or missions do not include a geospatial data production organization.

ESRI®<sup>3</sup> (Environmental Systems Research Institute, Inc.), a popular GIS software vendor and data reseller, maintains a Web site of private and commercial resources that make geospatial data available for profit (<http://www.esri.com/data/online/index.html>). ESRI also provides links to those sites that offer geospatial datasets at no cost. However, commercially available datasets, both for cost and no cost, often require license agreements and impose restrictions of the display and use of their products to the party that downloaded the data. Examples of commercial vendors include Geographic Data Technology (GDT) (<http://www.gdt1.com/products/index.cfm>) and Horizons Technology (<http://tango.horizons.com/>).

### 3.3. Size

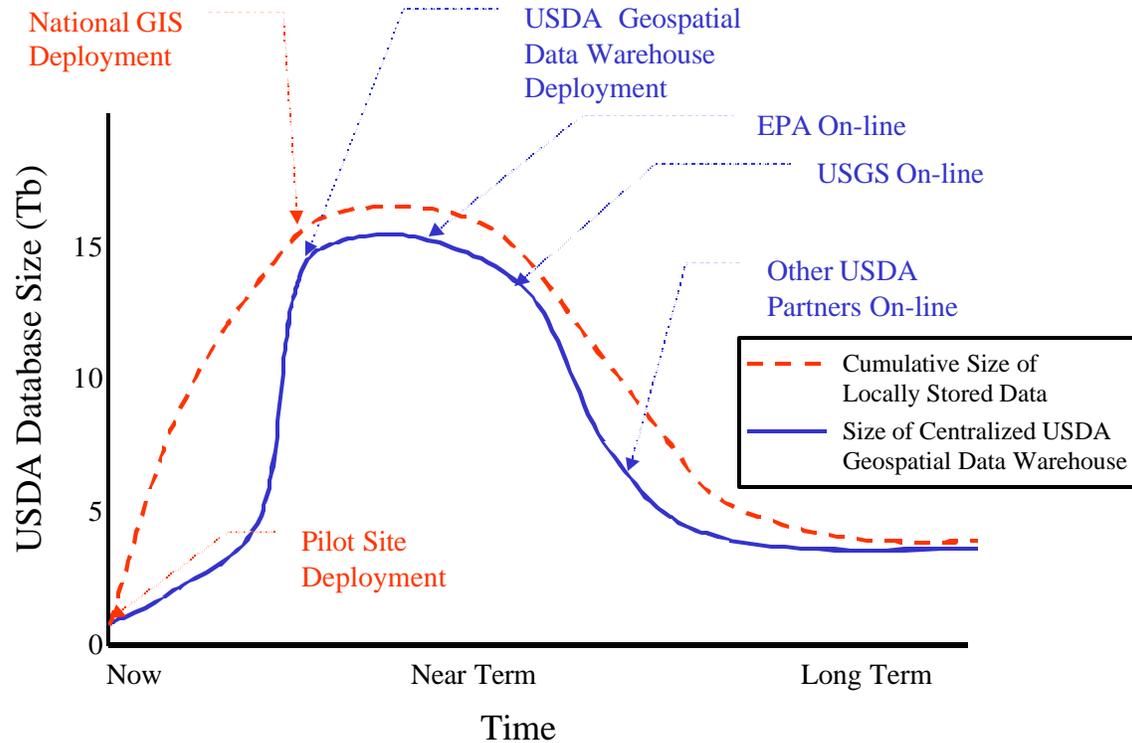
Each dataset category described in Table 2-1 and Table 2-2 contains multiple geospatial datasets that fall within that category. Preliminary estimates of the sizes for all datasets projected to be delivered to the Service Centers approaches 11 Terabytes. This information is contained in a report published in June 1999, in the *NRCS/ITC Evaluation of Service Center Data Themes Disk Space Requirements* document [A4]. These estimates include coverage for the 3111 counties of the CONUS and were based on the data sizing estimates for the 11 counties located within the nine BPR pilot sites. This study did not separate out the publicly owned land from the size estimates, which accounts for a significant increase in the estimated storage space requirement. It should be noted that of the 11 Terabytes of storage space that was estimated for national deployment, uncompressed ortho imagery consumed the bulk of this storage space at 10.3 Terabytes. The geographic component of the soils dataset (not including the NASIS attributes) consumes approximately 225,000 Megabytes and CLU consumes 22,000 Megabytes.

As with any database sizing efforts, certain assumptions are made about the rate of growth, replication of data, indexing, etc. Size estimates are subject to change, as the enterprise becomes fully operational, additional datasets are acquired and partnerships are formed among other agencies that produce the datasets. However, since the cost of storage continues to decrease, more significant factors such as the distribution of storage, data replication, data mirroring, data migration, etc. should be considered. Figure 3-1 is a notional representation of the database size requirements for USDA geospatial data.

---

<sup>3</sup> ESRI is a registered trademark of Environmental Systems Research Institute, Inc.

## Data Distribution Timeline



**Figure 3-1 Geospatial Data Distribution Timeline**

### 3.4. Evolution of Database Size

Currently, digital geospatial data is stored on a local server for the 11 counties in the nine BPR pilot sites. The size of the locally stored data ranges from five to 15 Gigabytes. As the GIS deployment continues in the near-term, it is anticipated that the cumulative size of the locally stored geospatial data will approach 11 terabytes. Of course, this number could vary widely depending on the method of storage, the duplication of data between Service Centers and amount needed for historical purposes. When the components of the USDA Geospatial Data Warehouse come to fruition, the data that is currently distributed among the Service Centers will be loaded into the Geospatial Data Warehouse. It is anticipated that the size of the cumulative locally stored data will initially be larger than that of the Centralized Geospatial Data Warehouse due to data duplication and less efficient means of storage. As the warehouse becomes populated and end-user applications begin to take advantage of the data distribution capabilities, less and less data will be distributed for storage locally. In the same time frame (due to similar initiatives) USDA's data partners will begin to implement interoperable geospatial data services that will enable end-users to access data directly from external data providers,

rather than from a redundant UDSA data store. This assumes that external data providers will provide data according to USDA standards, or the ability to convert to USDA standard format. The implementation of this capability at USGS is predicted to have the greatest impact on USDA database size. As illustrated by the chart, the amount of data stored centrally levels off in the long term. This represents the data owned and maintained by USDA. The amount of data stored locally also levels off at a smaller quantity. This represents the data that is acquired, used and never distributed by local entities as well as data cached for stand-alone applications.

### **3.5. Format**

Format determines the method of encoding that is used to portray the geospatial dataset in digital form. At this time datasets are either stored in vector or raster format and are maintained as ArcView<sup>®4</sup> shapefile, Arc/Info<sup>®5</sup> coverage, TIFF format, MrSID format and Spatial Database Engine<sup>®6</sup> (SDE) format.

#### **3.5.1. Vector datasets**

Currently, all vector datasets are processed in Arc/Info and subsequently converted to ArcView shapefile format for delivery to the pilot sites on CD-ROM. All datasets are being stored in UTM coordinates according to zone and are based on the North American Datum 1983 (NAD 83). These requirements may change as the scope of the enterprise GIS expands or customers require datasets be delivered in State Plane or other standard coordinate systems.

#### **3.5.2. Raster datasets**

Raster datasets are currently delivered in UTM coordinates in full-resolution TIFF format and MrSID compressed format. These needs may change, as new compression techniques become available. Datasets are copied to CD-ROM for delivery to the Service Centers.

#### **3.5.3. Spatially enabled databases**

Traditional tabular corporate databases can be “spatially enabled” by adding geographic location and feature type information to the database. The result is a continuous non-tiled spatial model that can be used for advanced spatial search functions, spatial geometry verification and projection functions. At this time none of the datasets described in this document are spatially enabled.

### **3.6. Tiling**

Currently, most vector datasets are tiled at the county level of geography. However some datasets are tiled at the state or national level. Ortho imagery (raster) is tiled according to

---

<sup>4</sup> ArcView is a registered trademarks of Environmental Systems Research Institute, Inc.

<sup>5</sup> Arc/Info is a registered trademarks of Environmental Systems Research Institute, Inc.

<sup>6</sup> Spatial Database Engine is a registered trademarks of Environmental Systems Research Institute, Inc.

both the county level of geography and to the quadrangle level. The county images are compressed using Lizard Tech's MrSID compression technique. Imagery tiled according to quadrangle is full-resolution, uncompressed. Some datasets, like the climate datasets are tiled and delivered at the state level of geography. Tiling is detailed dataset by dataset in Tables 2-1 and 2-2.

### **3.7. Temporality**

Temporality refers to the time period of the dataset maintained either on-line or in an historical archive. The data management system must maintain the most current version of the dataset on-line. Certain datasets also maintain an historical component and will require access to the most recent version as well as historical versions of USDA datasets that are deemed necessary according to the business requirements.

As datasets are updated or as new datasets are produced superceding an earlier version, an historical set of time series data is generated. Changes that have been made through time are captured in this series and can be used to display the temporal changes in the dataset. All temporal versions of USDA datasets will be maintained in the data warehouse designated for their storage. Historical versions of non-USDA datasets will not be maintained in USDA data warehouses. These datasets will be available directly from the agency charged with its production and maintenance.

### **3.8. Change control**

Any changes to the nature of the datasets must be handled through the data steward assigned to that dataset. All USDA datasets have a designated data steward charged with the management of the production, maintenance and update schedule of that particular dataset. The update schedule will vary on a dataset by dataset basis. Although non-USDA datasets are produced by their respective agency, a USDA data steward will be assigned for the management of that dataset within USDA's data management structure. In addition to data management, the data steward for the non-USDA datasets will have the authority to make changes to the dataset file names according to the change control policies established by the Inter-Operability (I/O) Lab in the *USDA Service Center Initiative Directory Structure and File Naming Convention Change Control Policy* [A5] and agreed to by the DMT on Geospatial Standards. Change control procedures will vary on a dataset by dataset basis.

### **3.9. Other physical characteristics**

Physical characteristics that have been standardized among all datasets and do not tend to vary on a dataset by dataset basis include projection, datum, scale, accuracy, horizontal topological integration, topology, horizontal integration, vertical integration, attributes and metadata are described in more detail in the *Standard for Geospatial Data*.

These physical characteristics defined in Table 3-1 pertain to the datasets that are delivered to the users and do not reflect the method of on-line storage and/or archival in a data warehouse. Geospatial datasets should be stored in the most economical and

functional format possible, as long as the delivery of the data involves translating the data into the physical parameters below.

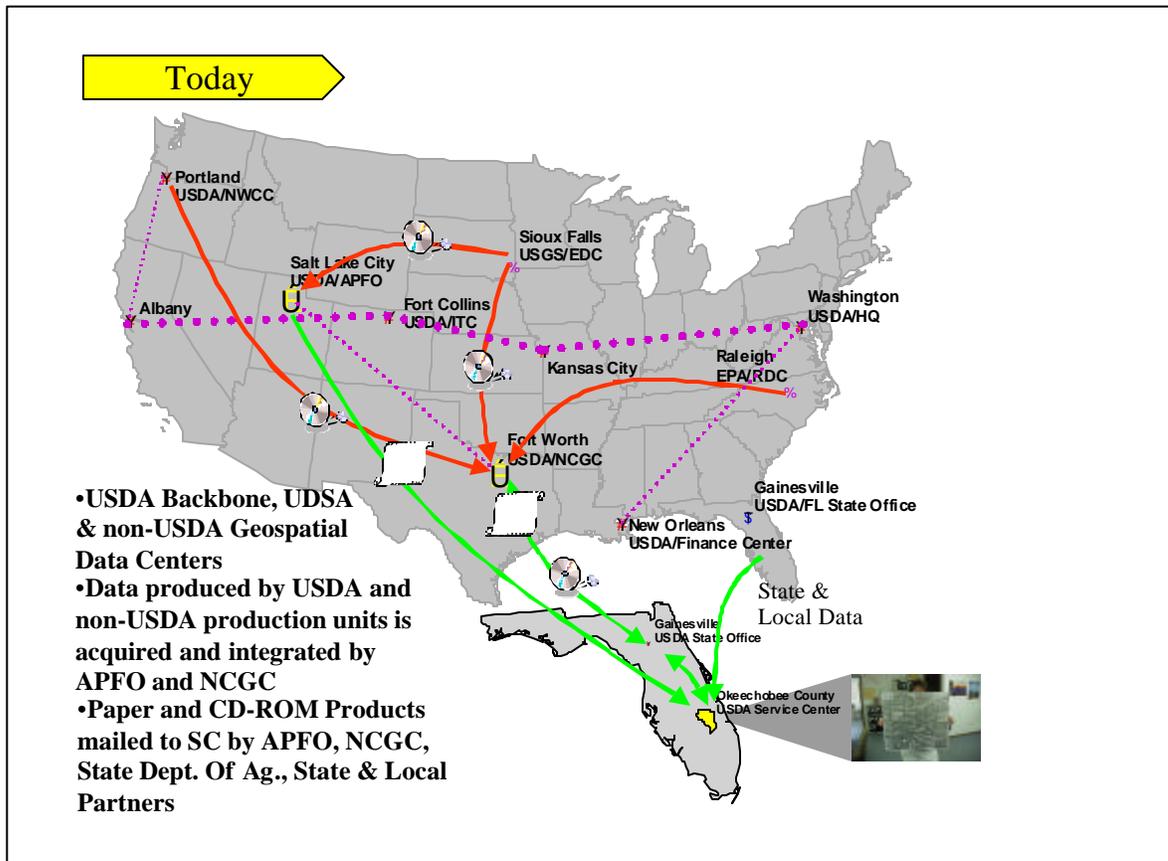
**Table 3-1 Other physical characteristics**

Characteristic	Description
Projection and Transformation	<p>All geospatial data (both raster and vector formats) should be represented in a projected coordinate system for use in the Service Centers. The coordinate reference system to be used is the Universal Transverse Mercator (UTM). Map units of measurement are to be in meters. When a UTM zone cuts through a county, all geospatial data will be projected into the predominant UTM zone for that county.</p> <p>Storage of geospatial datasets in the UTM coordinate system lends itself well to applications at the county level of geography. However many states and state GIS organizations are using data in the State Plane coordinate system and some have adopted custom coordinate reference systems. Although stored in the UTM coordinate, data can be delivered according to a variety of coordinate systems and datums. On-line datasets can be transformed into any standard projection system and convert between NAD 1983 and NAD 1927.</p>
Datum	North American Horizontal Datum 1983 and Vertical Datum 1988
Scale	Digital map data products in use at the Service Center are expected to be at a scale of 1:24,000 or larger. However, smaller map scale (1:50,000 or even 1:100,000) data may be provided and used at the Service Center if appropriate for the application or when that is the only scale at which data is available.
Accuracy	Where appropriate, providing more accurate GPS (Global Positioning Systems) data does not exist, DOQ data at 1-meter resolution should be used as the base for developing <i>geospatial datasets</i> that do not already exist but fall within the domain of USDA. DOQs, MDOQs, and compressed county mosaics produced from DOQs, meet NMAS for the 1:12,000 scale, where 90% of the locations tested must be within 33.3 feet (10 meters) of their true location.
Horizontal	All critical vector <i>geospatial datasets</i> (soils, common land units, and census) will be edge-matched across tile boundaries. Both spatial and attribute data must be consistent between tiles. Other common vector datasets could be edge-matched at a Service Center to the extent determined by cost/benefit.
Vertical	Vertical integration is the spatial alignment of geospatial objects between layers. For example, vector common land unit (CLU) boundaries to match fence lines visible on an ortho photo image or vegetation boundaries from a land cover layer matching stream lines on a hydrography layer.
Topology	All vector data must be topologically clean. Topologically clean implies that vector lines and areas must be free of spatial errors such as slivers and gaps, loop-backs and unnecessary duplicate geometry with nodes properly placed at linear intersections.
Metadata	<p>A minimum set of digital metadata for each <i>geospatial dataset</i> must be provided. All metadata must conform, at least, to the <i>SCI Standard for Geospatial Dataset Metadata</i>. This Standard is a subset of the FGDC Version 2 Content Standard for Digital Geospatial Metadata.</p> <p>Feature metadata must also be generated for each <i>geospatial dataset</i>. This includes information about feature geometry, attributes, symbology, and labels. The <i>Standard for Geospatial Feature Metadata</i>, currently in development, details the documentation necessary for feature metadata.</p>

Characteristic	Description
Attributes	All vector geospatial data will be attributed with at least a primary attribute, or key, such as a unique name or identification number. Additionally, geospatial data may be attributed with a display color and symbology. Geospatial data may have multiple attributes associated with them and if so, these attributes must be formatted and maintained in tables compatible with a relational database system.
Data Model	A geospatial data model must be provided for each <i>geospatial dataset</i> used in Service Centers. At a minimum, this model shall list for each <i>geospatial dataset</i> : the type(s) of <i>feature(s)</i> (i.e., point, line, area, text, raster or grid), <i>feature geometry</i> , <i>feature attributes</i> , including the primary key field used to link the <i>feature</i> to other tabular databases, and the default <i>feature symbology</i> and <i>feature labels</i> . All attribute data must be defined to the physical level.
Naming	<p>All map/feature type files must be named in accordance with Service Center Data Team (SCDT) standards and following any applicable rules and conventions for the GIS platform. The Data Management Team is working on a <i>Standard for Geospatial Dataset File Naming</i>, which outlines the file and directory structure naming conventions for SCI geospatial data and provides guidelines for geospatial data that is obtained from the state or local level.</p> <p>Applications are currently being fielded that depend upon this naming structure.</p>

#### 4. Current data management

All geospatial datasets, both USDA and non-USDA, are acquired, integrated, managed, stored and delivered by the USDA agency charged with their dissemination. The current geospatial data management process, as illustrated in Figure 4-1, is largely a manual process of moving information via Mail (CD-ROM, tape and paper), and File Transfer Protocol (FTP). Data is stored redundantly at the data source site, the data integration site and the end-user site. This operational process demands a great deal of labor and coordination that will increase as more USDA Service Centers acquire GIS technology and applications. The following section characterizes the ownership, storage and distribution of geospatial data management.



**Figure 4-1 Current Geospatial Data Management Operations**

#### 4.1. Ownership

Ownership of geospatial data used by USDA and non-USDA end-users ranges from local private and public organizations to State and Federal Agencies. The majority of data used by USDA comes from national programs such as the USDA Soil Survey, the National Digital Orthophoto Program and the Federal Insurance Mapping Program. Federal mapping programs often take several years to complete and require millions of dollars for both in-house and out-source data collection activities. USDA has a variety of cost-sharing and in-kind exchange agreements with these agencies. However, the data "ownership" is typically retained by the producing agency. Likewise, ownership of data that is produced within USDA is retained by the producing organization. Common Land Unit is an example of distributed, local data ownership.

#### 4.2. Storage

In the current data management environment, geospatial data is stored at the producer site, the integration/distribution site and the end-user site. This not only increases the

amount of storage required for this information but also causes substantial problems with versioning, data integrity and currency. Due to the bandwidth limitations at the integration and distribution sites (APFO, NCGC), the data must be archived or on near on-line for processing. CD-ROM and tape libraries as well as disk storage devices are used. A variety of analog and digital cataloging methods are to manage these storage libraries. Storage at the end-user sites range from digital storage libraries to loosely organized collections of CD-ROMs, tapes or files.

### **4.3. Distribution**

Geospatial data distribution is primarily via paper, FTP or mail. In general, the distribution of paper copies is decreasing while the use of CD-ROM or FTP is increasing each month. Much of the data are distributed as a "push" process, meaning that updated data is sent to end-users on a standing request. However, the implementation of FTP sites, 1 (800) services, and the soon to be implemented Resource Data Gateway allow end-users access to geospatial data with little to no intervention by the distributing organization.

## **5. Dataset management requirements**

To outline a geospatial data management strategy for USDA SCI, a clear set of data management requirements must be collected. These requirements must address both the user requirements, in terms of the anticipated use of the datasets described in section 2 as well as the data management system requirements.

User requirements include the requirements for the datasets that drive the organization, management and delivery of the data. System requirements include the hardware, software, communications, people and services that have been identified to manage geospatial data. In general terms, the system must store, manage and distribute geospatial datasets contained within the dataset categories identified and described in Tables 2-1 and 2-2 and be flexible and scalable enough to respond to changing dataset needs and user requirements.

### **5.1. User requirements (general use case scenarios)**

The following section provides the use-case analysis that describes the current geospatial data management situation. The current geospatial data management process was broken down into the following three scenarios:

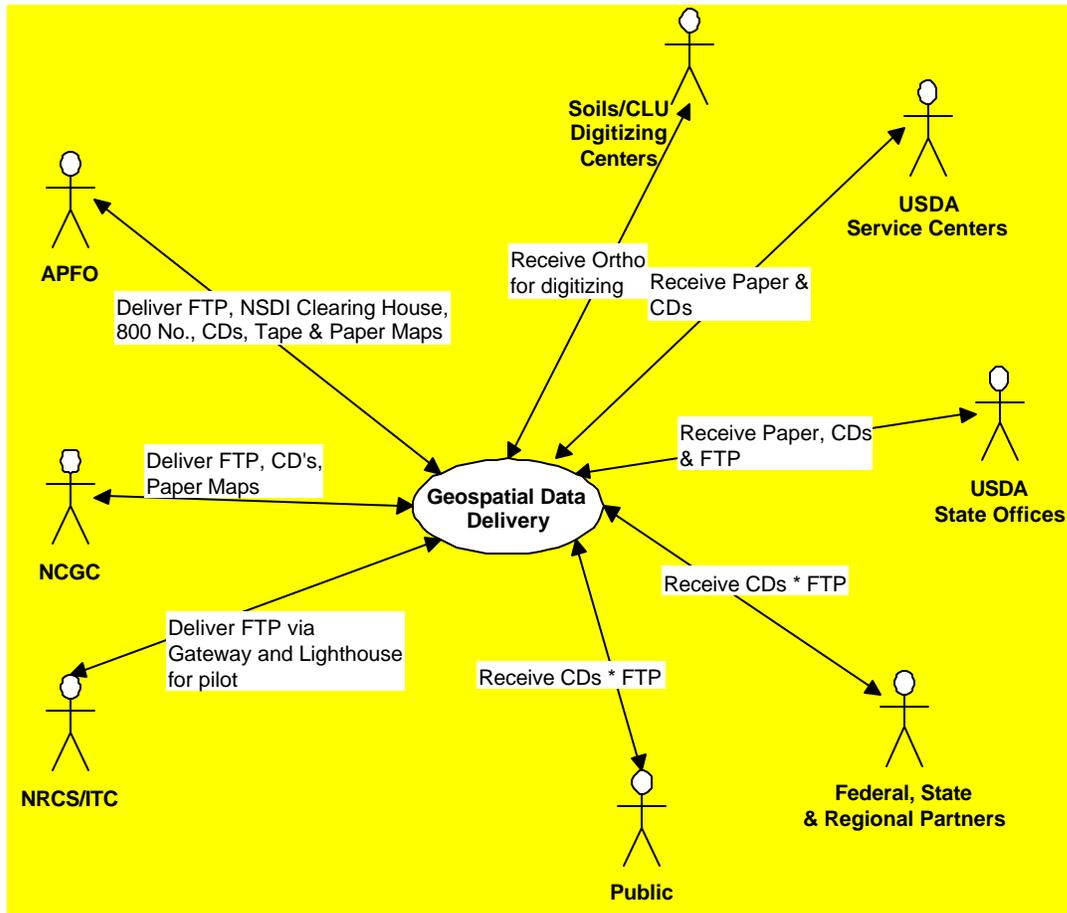
1. Geospatial Data Acquisition
2. Geospatial Data Integration
3. Geospatial Data Delivery

#### **5.1.1. Definition**

The Use-Case model describes what the Geospatial Data Management System will do at a high-level, with a user focus for the purpose of scoping the project. Each use-case

scenario portrays a process to a level detail required to make program planning decisions. This level of detail is not appropriate for actual implementation of any system at this stage.

**5.1.2. Geospatial Data Acquisition**



**Figure 5-1 Geospatial Data Acquisition Use Case Diagram**

**Table 5-1 Geospatial Data Acquisition Use Case Description**

USE-CASE 1	Geospatial Data Acquisition
Goal in Context	Geospatial Data Acquisition is the process of gathering data assets that meet the business needs of the SCI users. SCI geospatial data requirements vary widely by location, agency and program and must be acquired from multiple sources both internal and external to USDA.
Scope & Level	As a result, the scope of the acquisition strategy must be comprehensive enough to include such things as commercial procurement, data exchange partnerships, cost-sharing and

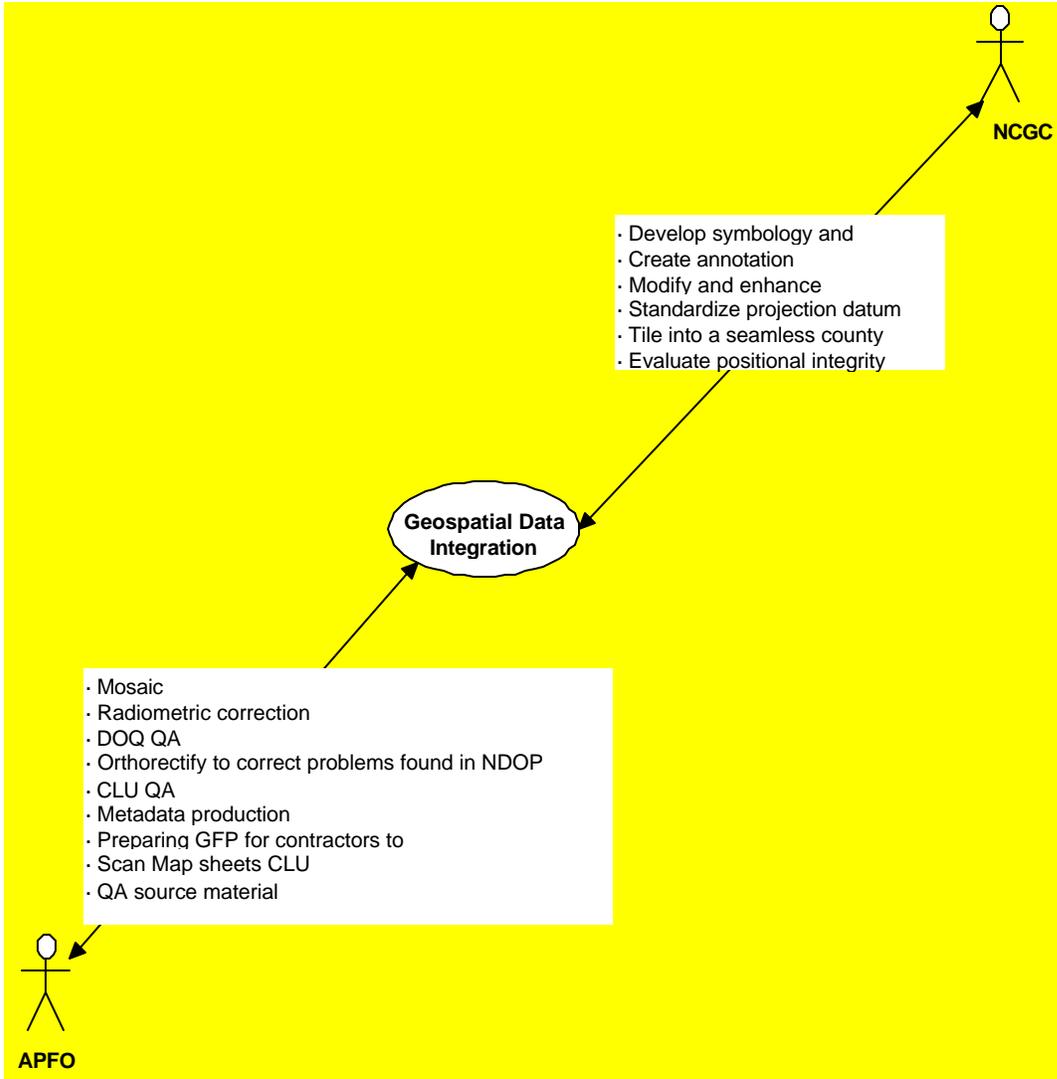
	<p>data development.</p> <p>This is a sub-function of Geospatial Data Acquisition, Integration and Delivery</p>	
Preconditions	<p>Potential providers of geospatial data are preparing data for USDA and other programs.</p>	
Success End Condition	<p>When Data Production Units (APFO, NCGC) of USDA receive data from source.</p>	
Failed End Condition	<p>Data are not received from source.</p>	
Primary, Secondary Actors	<p>Primary Actors:</p> <ul style="list-style-type: none"> <li>➤ Air Photo Field Office (APFO) is the traditional distributor of geospatial information (photo maps) for FSA. APFO is now focused on producing Mosaicked Digital Ortho Quads (MDOQ) and CLU for Service Centers. They are located in Salt Lake City, UT</li> <li>➤ National Cartographic and Geography Center (NCGC) is the primary processor and distributor of geospatial data for NRCS and is located in Ft. Worth, TX.</li> <li>➤ FSA Digitizing Centers - Digitize Common Land Unit (CLU) to be distributed by APFO.</li> <li>➤ USDA Data Stewards include a group of USDA organizations that produce geospatial for various programs (i.e. NASIS, soil digitizing units, National Water and Climate Center, NRI, etc.)</li> <li>➤ Federal Agencies (e.g. US Geologic Survey, Environmental Protection Agency, Fish &amp; Wildlife Service) produce geospatial data that is used for a wide variety of business and scientific applications, some of which belong to USDA.</li> <li>➤ State Agencies (e.g. Dept. of Transportation, Water Management Districts, etc.) partner with USDA on cost sharing and data exchange initiatives within the state.</li> <li>➤ Local/Regional Agencies (e.g. County Assessor Office, Utility Company, etc.) co-operate with USDA state and field service centers to exchange local data.</li> <li>➤ Commercial Vendors (e.g. ETAK, Group1 Software) sell value-added data to USDA for use in GIS applications.</li> </ul>	
Trigger	<p>New or standing geospatial data requirements comes into APFO or NCGC that cannot be filled by current data holdings.</p>	
DESCRIPTION	Step	Action
	1	Request for geospatial data is issued to NCGC or APFO production units (PU) by USDA, public or

		other interested parties.
	2	PU checks current data catalog for data status.
	3	If data is not available, then PU identifies best source of data and issues acquisition request.
	4	Data is procured by various methods
	5	Data is received by PU on electronic media and archived either on tape or CD-ROM.
EXTENSIONS	Step	
EXTENSION 1		In-house production
	3a	Data is not available from any source; PU must produce the data in-house through various means (e.g. scanning, digitizing)
SUB-VARIATIONS		
SUB-VARIATION 1		APFO Data Requirements Management
	1a	APFO has standing requirement to provide 24 x 24 image maps to the SCs. APFO currently publishes approximately 2,800 map sheets per month (36,600 a year). The frequency of this will decrease as more SCs procure GIS.
	1b	NAPP photo acquisition and DOQ production is coordinated with FSA, USGS, NRCS and other cost share partners through the NAPP and NDOP steering committees. Coverage requirements are negotiated. Cost sharing opportunities are prioritized over single agency requirements. In addition, APFO receives data requirements from the Forest Service and NRCS for contracting non-NAPP aerial photography acquisition.
	1c	DOQs are delivered to either APFO and/or NCGC from USGS, according to cost share arrangements.
SUB-VARIATION 2		Check Data Availability Status
	2a	APFO is currently faced with labor intensive process to check on availability of Digital Ortho Quarter Quads (DOQQ). APFO in-house DOQs (based on FSA cost share) frequently do not fully cover a county.
SUB-VARIATION 3		Data Availability
	3a	APFO must fill in void areas in counties by ordering DOQs directly from EDC and/or NCGC.

SUB-VARIATION 4		Procurement Methods
	4a	Established agreement for data acquisition is in place.
	4b	Cost sharing or purchase agreement is arranged.
	4c	An in-kind exchange is made.

RELATED INFORMATION	Geospatial Data Acquisition	
Priority:	Top	
Performance	Currently this could take days to months, but could be instantaneous in the future.	
Frequency	Acquisition is continuous nationwide, but happens days to years for each consumer	
Channels to actors	Ranges from mail, email, ftp, tape, CD-ROM	
OPEN ISSUES	Verify information with APFO, NCGC and DMT.	

### 5.1.3. Geospatial Data Integration



**Figure 5-2 Geospatial Data Integration Use Case Diagram**

**Table 5-2 Geospatial Data Integration Use Case Description**

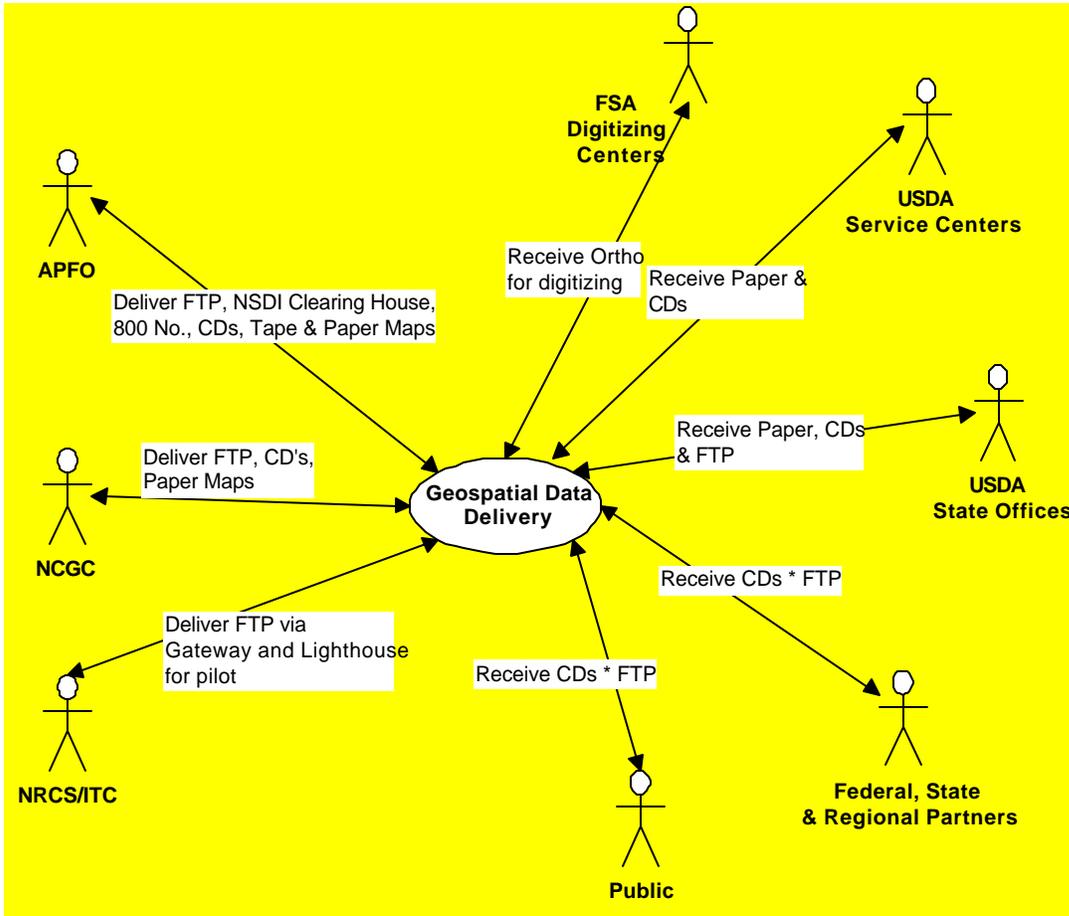
USE-CASE 2	Geospatial Data Integration
Goal in Context	Geospatial Data Integration is the process of adding value to data acquired by USDA to meet the requirements as specified by the USDA Spatial Data Standards for the USDA and non-USDA end-users including the public.
Scope & Level	As a result, the scope of the integration includes all

	<p>processes performed by UDSA production units (PU) (APFO, NCGC) that are required to transform data from the data source acquisition format(s) to the format(s) required by the end-users.</p> <p>This is a sub-function of Geospatial Data Acquisition, Integration and Delivery</p>	
Preconditions	Guidelines are established for geospatial data integration.	
Success End Condition	When geospatial data adheres to user requirements or spatial data standards.	
Failed End Condition	Data does not conform to standard.	
Primary, Secondary Actors	<p>Primary Actors:</p> <ul style="list-style-type: none"> <li>➤ Air Photo Field Office (APFO) is the traditional distributor of geospatial information (photo maps) for FSA. APFO is now focused on producing Mosaicked Digital Ortho Quads (MDOQ) and CLU for Service Centers. They are located in Salt Lake City, UT</li> <li>➤ National Cartographic and Geography Center (NCGC) is the primary processor and distributor of geospatial data for NRCS and is located in Ft. Worth, TX.</li> </ul>	
Trigger	PU obtains requirements and acquired data to be processed for end-user.	
DESCRIPTION	Step	Action
	1	Data set is evaluated for compliance to requirements.
	2	Production sub-unit is tasked with specific integration task.
	3	Data Integration Processing
EXTENSIONS	Step	none
SUB-VARIATIONS		
SUB-VARIATION 1		APFO Integration
	3a	Perform QA on DOQQs as they arrive from NCGC or EDC to access positional and radiometric integrity.
	3b	Mosaic DOQQs into seamless quadrangles.
	3c	Apply radiometric correction to DOQs in quadrangles to provide even contrast of imagery across county.
	3d	Notify USGS of DOQs that do not edgematch within 20 meters.

	3e	Generate replacement ortho imagery for rejected DOQs to meet GIS implementation schedule. Produce metadata for cataloging system.
	3f	QA source material for CLU digitizing. Generate MrSID compressed mosaic of MDOQs for county.
	3g	Scan Map sheets for CLU production
	3h	Preparing Government Furnished Package for contractors and digitizing centers to digitize CLU
	3i	QA Common Land Unit from new Digitizing Centers and from contractors.
<b>SUB-VARIATION 2</b>		<b>NCGC Integration</b>
	3a	Develop symbology and legends
	3b	Create annotation
	3c	Modify and enhance geocoding
	3d	Standardize projection datum units
	3e	Tile into a seamless county theme
	3f	Evaluate positional integrity

<b>RELATED INFORMATION</b>	Geospatial Data Integration
Priority:	Top
Performance	This process ranges from minutes to days
Frequency	Integration is continuous.
Channels to actors	direct
<b>OPEN ISSUES</b>	Verify information with APFO, NCGC and DMT.

**5.1.4. Geospatial Data Delivery**



**Figure 6 – Geospatial Data Delivery Use Case Diagram**

**Table 6 - Geospatial Data Delivery Use Case Description**

<b>USE-CASE 3</b>	<b>Geospatial Data Delivery</b>
Goal in Context	Geospatial Data Delivery is the process of moving geospatial information (hardcopy or softcopy) from the USDA Production Units (PU) to the USDA and non-USDA end-users including the public.
Scope & Level	As a result, the scope includes the distribution of geospatial data via: <ul style="list-style-type: none"> <li>• Mail (CD-ROM, tape, paper)</li> <li>• File Transfer Protocol (FTP)</li> <li>• Feature Streaming via the Internet and USDA Intranet</li> </ul>

	<ul style="list-style-type: none"> <li>Viewing geospatial data in a web browser</li> </ul> <p>This is a sub-function of Geospatial Data Acquisition, Integration and Delivery</p>	
Preconditions	Guidelines are established as guidelines for geospatial data integration.	
Success End Condition	End-users receive standards based geospatial data that can be used to support USDA and non-USDA applications.	
Failed End Condition	Data does not reach end-use	
Primary, Secondary Actors	<p>Primary Actors:</p> <ul style="list-style-type: none"> <li>➤ Air Photo Field Office (APFO) is the traditional distributor of geospatial information (photo maps) for FSA. APFO is now focused on producing Mosaicked Digital Ortho Quads (MDOQ) and CLU for Service Centers. They are located in Salt Lake City, UT</li> <li>➤ National Cartographic and Geography Center (NCGC) is the primary processor and distributor of geospatial data for NRCS and is located in Ft. Worth, TX.</li> <li>➤ Natural Resource Conservation Service/Information Technology Center (NRCS/ITC) stores and distributes data for BPR pilot sites via the Resource Data Gateway (RDG) and the Lighthouse Project.</li> <li>➤ FSA Digitizing Centers - Digitize Common Land Unit (CLU) to be distributed by APFO.</li> <li>➤ USDA State Offices add value, store and distribute geospatial data to USDA and non-USDA end-users.</li> <li>➤ USDA Service Centers receive paper and digital products for use on a variety of Service Center Agency (FSA, NRCS, and RD) applications including Customer Service Toolkit, Wetlands and Easements and Resource Data Gateway.</li> <li>➤ Federal, State and Regional Partners (e.g. US Geologic Survey, State Department of National Resources, Conservation District Offices) provide geospatial data for a wide variety of business and scientific applications.</li> <li>➤ Public accesses USDA geospatial data stores for a variety of applications</li> </ul>	
Trigger	End use has need to use geospatial information for USDA or non-USDA business	
DESCRIPTION	Step	Action
	1	End user (FSA Digitizing Centers, USDA State Offices, USDA Service Centers, Federal, State and Regional Partners, Public) establishes need for

		geospatial information to support business activity.
	2	End user queries or browses USDA data holdings for required geospatial data either on-line via the NRCS/NSDI Clearinghouse Node, APFO Website, NRCS 1-800 number, RDG, email or by phone.
	3	USDA data provider notifies end user on availability of data either on-line or by phone or email.
	4	End user places order for geospatial data providing format and data delivery method requirements by phone, email or on-line.
	5	USDA geospatial data providers (APFO, NCGC, NRCS/ITC) initiates Geospatial Data Integration Task (see Geospatial Data Integration Use Case).
	6	USDA data provider notifies end user of status of order either by phone, email or on-line.
	7	USDA data provider disseminates geospatial data to end user via a variety of methods.
	8	End user receives geospatial data for use in business activity.
<b>EXTENSIONS</b>	<b>Step</b>	
		End User Requirements
	1a	FSA Digitizing Centers use Mosaicked DOQ, source package and scanned county manuscript to digitize CLU.
	1b	USDA State Office, primarily NRCS state offices receive DOQQ and other geospatial products to support Soil Survey, resources assessments, etc.
	1c	USDA Service Centers currently use geospatial in various field offices and in the nine BPR pilot sites to support applications such as Customer Service Toolkit.
	1d	Federal, State and Regional partners use USDA geospatial data to support their own business activities.
	1e	Public uses USDA geospatial information for a variety of activities (i.e. printing maps, education, planning, research, real estate).
<b>SUB-VARIATIONS</b>		
		Geospatial Data Dissemination
	7a	NCGC disseminates 150-200 paper maps/year, 2500 FTP downloads/month, approximately 3000

		Ortho CD-ROMs/year, approximately 350 1-800 calls/year. In addition, NCGC provides DOQ and DRG on-line via the MIT Browser and provides sample data via ArcView IMS, MO IMS, and HTML version.
	7b	APFO disseminates 33,600 paper maps to Service Centers per year. 67,000 photo products are disseminated to support the NDOP program. 523,000 photo products and 1,200 scanned aerial photos are disseminated to other government agencies and the general public. APFO also reproduces DOQs for NCGC. APFO disseminates digital ortho data to digitizing centers. Digital Ortho and CLU data are distributed to all Service Centers with GIS technology and applications.
	7c	NRCS/ITC Resource Data Gateway and the Lighthouse Project will provide geospatial data on-line to the BPR Pilot Sites.

RELATED INFORMATION	Geospatial Data Delivery
Priority:	Top
Performance	Currently dissemination could take minutes to weeks.
Frequency	Dissemination occurs on both regular intervals and as-needed.
Channels to actors	Mail, email, FTP, HTTP
OPEN ISSUES	Verify information with APFO, NCGC and DMT.

## 5.2. Data management system requirements

System requirements are those parameters that have been identified by the SCI as necessary to support the data management requirements identified by the partner agencies in conjunction with the DMT. These requirements must be met whether the data management system resides wholly within USDA, at another federal agency or not for profit agency, an Application Service Provider (ASP) data warehousing facility, or a combination of these. The system requirements are defined in the following sections.

### 5.2.1. Browsing

The data management system must provide the ability to browse on-line data assets so users can determine if the datasets that they require are available. It should consist of a catalog that allows users to browse data holdings on-line within a user friendly, Web-based graphical user interface (GUI). Users should be able to search for dataset by keyword search, geographic coordinates, state name, county name, state or county

Federal Information Processing Standard (FIPS) code and ZIP code. The browse function should provide the user with the ability to view the metadata associated with the datasets and the option to view a “thumbnail” image of the desired product prior to “ordering” the dataset.

### **5.2.2. Storage**

Storage includes providing a physical repository for geospatial datasets and associated metadata records. Components for storage include any hardware, software and Data Base Management Systems (DBMS) that maintain the data warehouse and make it available for query and distribution. A data warehouse storage service is to include any upgrades to system components mentioned above and technical support to those components to minimize system downtime, prevent user disruption and maintain data integrity.

### **5.2.3. Management**

Management includes the ability to store geospatial data and metadata records and be able to distribute data in the required format (described in Table 3-1), providing a secure system for the storage of datasets, making datasets available to users on-line, and designing a system to accept uploads of data from authorized data stewards.

### **5.2.4. Delivery**

At this time, all datasets are mailed to users on CD-ROM from one of the two USDA production centers located in Ft. Worth, TX or Salt Lake City, UT. This method of delivery will change as the Gateway/Lighthouse project enters production, as OpenGIS completes the standardization of Web mapping specifications and Internet delivery becomes more mainstream. While these possibilities are still off in the near-term (Gateway/Lighthouse) and long-term (Web mapping and Internet delivery) and given the current situation of CD-ROM delivery, the following requirements have been identified to handle the current and near-term data management activities.

- The data management system should have the capacity to connect to NSDI Clearinghouses, USDA clearinghouse nodes, state partners, federal partners and Service Centers as identified by the SCI. This connection ensures that customers are able to access all of the USDA datasets identified as important regardless of the data producer or data distributor.
- The system must be able to provide data to users operating on a desktop computer and using an USDA Wide Area Network (WAN), Local Area Network (LAN), modem connection or wireless satellite connection.
- The system must be capable of delivering datasets, both raster and vector formats according to any tiling scheme, including user-defined areas. The system must store geospatial datasets according to a seamless nationwide tiling scheme that will enable the users to define their own “area of interest”. The current tiling scheme for each dataset category are detailed in Tables 1 and 2.

- Finally, the system should provide a service to deliver data on CD-ROM through the United States Postal Service (USPS) first class mail, overnight mail service and electronically through File Transfer Protocol (FTP), E-mail or Web download.

### **5.3. Standards and specifications**

All datasets provided to the data management system will be topologically correct and will be accompanied with FGDC and/or USDA compliant metadata according to the *Standard for Geospatial Dataset Metadata*[A6]. The data management system must deliver data according to the standards developed by the SCI DMT. The relevant standards include the *Standard for Geospatial Data* and the *Standard for Geospatial Dataset File Naming*[A7]. The metadata and dataset formats must not be compromised by the data management service. All three standards listed here are available at <http://www.ssbkc.usda.gov/scdm/DP/Parent.htm>.

Additionally, the data management service must comply with the Open GIS Consortium (OGC) Interoperability Specification (<http://www.opengis.org>).

### **5.4. Metadata**

Provisions should be made for the management, maintenance and delivery of metadata files associated with the geospatial datasets. Metadata will need to be uploaded and refreshed as changes to the geospatial datasets require. These uploads will be performed by the data steward assigned to that particular geospatial dataset. Additionally, geospatial datasets should be packaged with its corresponding metadata record(s) for delivery. Metadata and geospatial data may not have the same update schedule. The data management system must be able to link to the Service Center Data Management (SCDM) Metadata Repository. This repository houses SCI metadata for tabular datasets, geospatial datasets and feature types. The repository is located at the following URL: <http://www.ssbkc.usda.gov/scdm>.

### **5.5. Transactional datasets**

Transactional datasets will be processed outside of the data management system and posted to the appropriate data warehouse according to a predefined update schedule. The system must be capable of accepting updated information, as approved by the data steward, and post these changes to the on-line dataset.

### **5.6. Availability**

The system must be on-line and available on a 24-hour, 7-day basis for download and order placement. System maintenance and upgrade should not be performed during “mission critical” times designated as 6:00 a.m. EST/EDT until 12:00 a.m. EST/EDT Monday through Friday. Additionally, a toll-free number and E-mail address should be provided to report problems with the system. One hundred percent availability may not

be possible, however system resources and administration should be balanced in order to achieve maximum availability and minimize system outage and downtime.

### **5.7. Security**

The service must provide a secure environment for the physical infrastructure of the system components, all data stored within the system and Internet transactions. Security includes network security, database security, geospatial data warehouse, geospatial dataset and geospatial feature security. The Common Computing Environment (CCE) team has documented security procedures in their *Draft USDA SCI and CCE Information Technology Architecture Version 2.0* [A8]. Appendix B describes security requirements, services and standards that must be implemented to ensure a secure infrastructure.

### **5.8. Privacy**

Certain datasets contain sensitive information concerning the financial standing of USDA customers, addresses, telephone numbers and outstanding loan information. Datasets that contain this information may require additional measures to ensure that unauthorized use and distribution of the data does not occur.

### **5.9. Disaster recovery**

The data management system must perform routine backups according to industry standards. This ensures that datasets can be restored to their latest version if the integrity of the on-line version becomes corrupted or is in question. Backups and system downtime should be performed during non-critical hours so as not to take resources away from the on-line search and delivery services. In addition to routine back ups, the system should have a data recovery plan that includes the ability to restore data and should maintain redundant datasets important to USDA business through mirroring of storage devices. Additionally, specific datasets such as USDA land use may require back up and storage requirements that extend beyond industry standards for system recovery and maintenance due to the sensitive nature of the dataset to contain customer information and the temporal characteristics of the dataset. The Common Computing Environment (CCE) team has documented backup and data recovery procedures in their *Draft USDA SCI and CCE Information Technology Architecture Version 2.0* in Section 6.7 titled "Redundancy and Failover".

### **5.10. Performance measures**

Limitations in the telecommunications bandwidth in many Service Centers impedes the delivery of geospatial data in a timely manner. The service must address the situation and be able to provide alternate methods of telecommunication. Detailed performance measures have been documented in CCE's *Draft Investment Center and Server Sizing Modeling Report* [A9].

## 6. Summary

In summary, the USDA Geospatial Data Management Requirements include complex set of variables and IT challenges. Management of geospatial data represents a multitude of sources, formats, infrastructure and business uses. This document lays the framework for the second part of the overall *Geospatial Data Implementation Plan* that will provide recommendations and guidance for the physical location of servers; operations; metadata management; interaction between production units, state offices and Service Centers; and the hardware, software, telecommunications infrastructure, human resources, and their roles and responsibilities.

## Appendix A – Bibliography

- [A1] Service Center Geographic Information Systems Strategy, April 1999
- [A2] Standard for Geospatial Data, December 1999
- [A3] Geospatial Data Acquisition, Integration and Delivery National Implementation Strategy Plan, September 1999
- [A4] NRCS/ITC Evaluation of Service Center Data Themes Disk Space Requirements, June 1999
- [A5] USDA Service Center Initiative Directory Structure and File Naming Convention Change Control Policy, February 1999
- [A6] Standard for Geospatial Feature Metadata, April 1999
- [A7] Standard for Geospatial Dataset File Naming, April 1999
- [A8] Draft USDA SCI and CCE Information Technology Architecture Version 2.0, 2000
- [A9] Draft Investment Center and Server Sizing Modeling Report, May 2000