

NRCS/ITC Evaluation of Service Center Data Themes Disk Space Requirements

**Version 1.0
June 25, 1999**

Prepared For: Wendall Oaks, USDA, Gateway Project Manager & COTR

Copies To: Dave Anderson, USDA, Data Management Team Leader
John Lee, UNISYS, Project Leader
Jean Maloney, UNISYS, Gateway Team Leader
David Tessier, SAIC, Data AID Team

Prepared By: Randy Frosh, UNISYS, Data AID Team

1.0	PURPOSE.....	3
2.0	BASIS OF ESTIMATES	3
3.0	ASSUMPTIONS	3
4.0	SIZE SUMMARY	5
5.0	COUNTY SIZE.....	6
6.0	DIGITAL ORTHOPHO TO QUADRANGLES	7
7.0	GEOSPATIAL DATA LAYER DETAILS	7
8.0	SPATIAL DBMS	8
9.0	MEGABYTES PER SQUARE MILE	8
10.0	SUMMARY.....	10

1.0 Purpose

The purpose of this report is to estimate the disk space requirements for the geospatial data that are needed for use at each of the Service Centers. The geospatial data that are required have been identified in the *Service Center Geospatial Data Standard* and were derived from the *USDA Service Center Geographic Information System (GIS) Strategy*. The disk space requirement estimates are intended for use by the Data Management and Data Aid Business Process Reengineering (BPR) teams.

Geospatial data storage requirements are known for the existing BPR pilot sites. From these, estimates are calculated for storage requirements for each data theme and for the total storage requirements for a 'typical' Service Center. Estimates of the disk space required for total coverage of the coterminous US are also provided.

2.0 Basis of Estimates

Nine BPR sites around the country were selected as test sites for the use of GIS and the Customer Service Toolkit. Several themes of information have been provided to the BPR sites by the NRCS National Cartography and Geospatial Center (NCGC) in Fort Worth, Texas and the FSA Air Photo Field Office (APFO) in Salt Lake City, Utah.

The BPR site data is on-line at NCGC in the form of ESRI shape files, TIFF images, and MrSID compressed images. The UNIX `du` command was used to provide the number 512 byte blocks used by each theme in each county for each BPR site. This raw data is provided in table 2-1. A value of zero means that the theme is not available for the county. Note that the Digital Orthophoto Quadrangles (DOQs) in non-compressed (TIFF) format are not included in the figures in tables 2-1 or 4-1. They are discussed in sections 6.0 and 10.0.

3.0 Assumptions

Several assumptions are made in developing the estimates.

- The BPR sites represent typical counties.
- The BPR sites are typical in terms of extent.
- The BPR sites are typical in terms of complexity of data.
- The BPR sites provide an average size that can be used to infer national coverage size.
- The sizing estimates do not include space for data preparation, only final file storage sizes.

	Size in 512 byte Blocks				
Service Center	Abilene	Los Lunas	Paola	Okeechobee	Okeechobee
County	Taylor TX	Valencia NM	Miami KS	Glades FL	Okeechobee FL
Geospatial Data					
Cadastre	0	628	508	536	1364
Climate-Precipitation	28380	17204	5922	11936	11936
Climate-Temperature	0	0	0	0	0
Common Land Unit	18530	1406	23064	2416	2110
Demographics	15116	6104	8084	4848	8584
Digital Raster Graphics	141524	159960	98372	104724	86356
Elevation	50970	118914	0	6530	14870
FEMA	0	0	0	13956	0
GNIS	132	370	136	74	88
Government Units	328	44	44	204	1934
Hydrography	1356	7450	11230	23490	25474
Hydrologic Units	0	0	5184	180	64686
Land Use Land Cover	0	0	14908	47712	34100
Ortho Imagery	258406	457012	220916	221396	235748
Photo Index	80	112	58	40	58
Plants	0	0	0	0	0
Quad Boundaries	126	96	64	64	68
Soils	79618	22510	118260	84906	37122
Transportation-Roads	15742	7184	0	1725	1754
Transportation-Rail	3530	0	0	42	0
Wetlands-NWI	0	0	10968	36506	68674
Legends	906	32	216	38	30
Metadata	112	198	258	296	424
Total	614856	799224	518192	561619	595380

TABLE 2-1 GEOSPATIAL DATA SIZES BY COUNTY

	Size in 512 byte Blocks					
Service Center	Sacramento	Scottsburg	Snow Hill	Rolla	The Dalles	The Dalles
County	Sacramento CA	Scott IN	Worcester MD	Rolette ND	Wasco OR	Hood River OR
Geospatial Data						
Cadastre	678	0	0	0	0	300
Climate-Precipitation	76054	5232	5560	6640	73004	73004
Climate-Temperature	0	0	0	0	0	0
Common Land Unit	10152	11952	0	30482	0	0
Demographics	52416	3594	12794	9508	15358	7198
Digital Raster Graphics	172486	64692	132004	130004	411396	108628
Elevation	57994	0	26884	6788	98294	32754
FEMA	7924	0	15612	48	1554	454
GNIS	614	126	614	128	672	428
Government Units	140	184	256	632	68	38
Hydrography	8424	5600	14632	3382	17860	18206
Hydrologic Units	0	2544	0	0	0	0
Land Use Land Cover	0	0	0	0	0	0
Ortho Imagery	432726	123862	0	298534	555638	222022
Photo Index	166	42	0	112	172	48
Plants	0	0	0	0	0	0
Quad Boundaries	138	62	84	90	208	64
Soils	217186	114668	246388	405452	90456	34958
Transportation-Roads	0	492	0	0	0	0
Transportation-Rail	0	0	0	0	0	0
Wetlands-NWI	47666	4160	26992	71026	0	0
Legends	176	74	226	38	142	108
Metadata	374	266	164	388	306	298
Total	1085314	337550	482210	963252	1246528	498508

TABLE 2-1 (CONTINUED)

4.0 Size Summary

Table 4-1 provides several summary statistics for the raw data in Table 2-1. The number of counties sampled (County Count) can be used to give a measure of confidence in the column of average megabytes per county (Average Megabytes). The minimum and maximum size in megabytes shows the range by theme for the BPR sites. The All Themes & Counties row at the bottom of Table 4-1 shows the average, minimum and maximum size for the sample counties. It shows that the average size for a county is 350 megabytes. Note that some Service Centers service more than one county and the figures do not include DOQs in non-compressed format (TIFF).

Finally, the coterminous USA column shows the space required for total coterminous national coverage. It is derived by multiplying the number of counties (3111) by the average size for the theme. Using the BPR sites for the calculations, it will take over one terabyte (1.238) to store the

geospatial data for the USA. The 1.238 terabyte estimate does not include Alaska, Hawaii, and US Territories where there are Service Center offices. It also does not include the DOQ TIFF files. The service center data will be developed at the rate of 600 centers a year for four years before the 2300 centers are complete.

Geospatial Data	Summaries in Megabytes					
	Total Megabytes	County Count	Average Megabytes	Minimum Megabytes	Maximum Megabytes	Coterminous USA Total (3111 counties)
Cadastral	2.01	6	0.33	0.00	0.68	1,040.6
Climate-Precipitation	157.44	11	14.31	2.62	38.03	44,525.8
Climate-Temperature	0	0	0	0	0	0
Common Land Unit	50.06	8	6.26	0.00	15.24	19,465.5
Demographics	71.80	11	6.53	1.80	26.21	20,306.9
Digital Raster Graphics	805.07	11	73.19	32.35	205.70	227,689.3
Elevation	207.00	9	23.00	0.00	59.46	71,552.7
Fema	19.77	6	3.30	0.00	7.81	10,252.8
GNIS	1.36	11	0.12	0.04	0.31	383.2
Government Units	1.90	11	0.17	0.02	0.97	537.9
Hydrography	59.62	11	5.42	0.68	12.74	16,862.2
Hydrologic Units	36.30	4	9.07	0.00	32.34	28,230.0
Land Use Land Cover	48.36	3	16.12	0.00	23.86	50,149.3
Ortho Imagery	1,513.13	10	151.31	0.00	277.82	470,734.7
Photo Index	0.44	10	0.04	0.00	0.09	138.1
Plants	0	0	0	0	0	0
Quad Boundaries	0.53	11	0.05	0.03	0.10	150.5
Soils	725.76	11	65.98	11.26	202.73	205,258.7
Transportation-Roads	13.45	5	2.69	0.00	7.87	8,367.7
Transportation-Rail	1.79	2	0.89	0.00	1.77	2,778.1
Wetlands-NWI	133.00	7	19.00	0.00	35.51	59,107.2
Legends	0.99	11	0.09	0.02	0.45	280.8
Metadata	1.54	11	0.14	0.06	0.21	436.1
All Themes & Counties	3,851.32	11	350.12	168.78	623.26	1,238,248.2

TABLE 4-1 GEOSPATIAL DATA SIZE SUMMARIES

5.0 County Size

A key assumption is that the BPR sites represent typical counties in terms of square miles. Most of the themes will vary in disk space requirements with the mapped area. Table 5-1 shows the area in square miles for each of the counties. There are approximately 3,347,670 square miles in the coterminous USA. This provides an average county size of 1070.5 square miles. The average size of the BPR sites is 907.78 square miles, which is 84.76 % of the average. This indicates that the size estimate for US coverage should be increased by 15.34 %, to **1.4282 terabytes**.

Taylor TX	Valencia NM	Miami KS	Glades FL	Okeechobee FL	Sacramento CA	Scott IN	Worcester MD	Rolette ND	Wasco OR	Hood River OR	Average (sq. mi)
919.3	1068.2	590.2	986.1	891.9	995.6	192.7	473.6	939.3	2395.2	533.5	907.78

TABLE 5-1 BPR COUNTY SIZES

6.0 Digital Orthophoto Quadrangles

The reader should note that the largest theme listed in the tables above, Ortho Imagery, only contains the MrSid compressed county mosaic. This theme does not include Digital Ortho Quadrangles (DOQ) or Digital Ortho Quarter Quads, which are a requirement of some Service Center projects. Including single band DOQ data for the US would increase the storage requirements by a factor of 10 for both the individual Service Centers and for centralized database(s) to cover the U.S, or wherever it is necessary to store this data.

The average single band 3.75 minute DOQ is 48 megabytes and a multi-band DOQ is 144 megabytes. There are 4 x 3.75 minute DOQs per 7.5 minute quadrangle. There are 53,818 x 7.5 minute quads in the contiguous US. Assuming 48 megabytes per quarter quadrangle, an **additional** 10,333,056 megabytes, or **10.333 terabytes**, of disk space is required to store single-band DOQ files. Table 10-1 summarizes the geospatial disk space requirements for all data, including DOQ TIFF images.

7.0 Geospatial Data Layer Details

Another assumption is that the quantities of data for the BPR sites are typical and representative. However, not all BPR sites have all the map themes completed. For two themes, plants and climate-temperature, no data is available for any BPR site.

The transportation-rail theme is only available for two of the eleven counties, which does not produce a reliable average size. Similarly, there is only data available for four counties in the hydrologic units theme and three counties for the land use/land cover theme.

The soils geospatial data are a very large vector data set. For most counties, the following maps and associated attribute data are included in the data:

- soils
- prime farmlands
- Soil Survey Area
- Major Land Resource Areas (MLRA)
- hydric soils

Prime farmlands are recently being produced. They duplicate the information in the soils map but have additional attribute fields and will perhaps supercede the soils map.

8.0 Spatial DBMS

Storing the data in a DBMS will require a different amount of space than storing the data in shape files and compressed images. A section in the *NRCS/ITC Evaluation of ESRI's Spatial Data Engine (SDE)*, Version 1.0 June 1, 1999 report compared the size of shape files with the size of the files after being inserted into a DBMS. This information is shown in Table 8-1. It shows that complex features with many coordinates require less space in a SDE/Oracle database than they do in shape files. However, geographically simple features with few coordinates require more disk space in SDE/Oracle than their shape file equivalents.

The majority of the BPR data have not yet been loaded into SDE/DBMS. As a result, an estimate for a factor for an increase or decrease in disk space requirements can not reliably be made. Table 8-1 should be used as an indicator of data storage requirement trends, but the figures are not used in the calculations of the disk space requirements.

Theme	Features	Total Coordinates	Average Coordinates per feature	Size of Shape Files (bytes)	Size of SDE/Oracle DBMS (bytes)	Percent change
qd250k	469	2,361	5	95,209	176,128	+84.9909
qd100k	1,842	46,050	25	959,584	839,680	-12.4954
qd24k	53,818	269,270	5	10,857,088	16,343,040	+50.5287
mad	7,505	374,481	50	7,250,431	4,995,072	-31.1066
county	3,111	2,570,334	826	42,689,408	17,981,440	-57.8785
us_concise	30,368	40,344	1	11,137,802	9,482,240	-14.8644
soils	39,579	6,483,695	156	107,576,987	34,637,824	-67.8018

TABLE 8-1 SIZE REQUIREMENTS OF SHAPE AND SDE/ORACLE FORMATS

9.0 Megabytes per Square Mile

Using the average megabytes per theme and the area per county allows computation of a megabytes per square mile value. Assuming representative sampling, this factor can be used to estimate disk space requirements for the data for a new county of a given size. The values are shown in Table 9-1. The table shows that it takes 0.43845619 megabytes per square mile to store the themes identified by the Data AID team and the *USDA Service Center Geographic Information System (GIS) Strategy* document. A size of zero indicates that no data currently exists for this theme, and the values are unknown.

Geospatial Data	Megabytes per Square Mile
Cadastral	0.00036848
Climate-Precipitation	0.01576630
Climate-Temperature	0
Common Land Unit	0.00689263
Demographics	0.00719055
Digital Raster Graphics	0.08062340
Elevation	0.02533636
FEMA	0.00363046
GNIS	0.00013570
Government Units	0.00019047
Hydrology	0.00503700
Hydrologic Units	0.00999607
Land Use Land Cover	0.01775757
Ortho Imagery	0.16668433
Photo Index	0.00004891
Plants	0
Quad Boundaries	0.00005328
Soils	0.07268086
Transportation-Roads	0.00296294
Transportation-Rail	0.00098372
Wetlands-NWI	0.02092951
Legends	0.00009944
Metadata	0.00015442
Total	0.43752240

TABLE 9-1 DISK SPACE REQUIRED PER SQUARE MILE

10.0 Summary

The following table summarizes the calculations made for the geospatial data requirements for the Service Center BPR projects currently identified by the Data AID team and the *USDA Service Center Geographic Information System (GIS) Strategy* document. This table reflects the additional 15.34% required because the sample set has smaller counties than the national average. It has summarizes storage requirements for DOQ's.

Geospatial Data	Average per Square Mile	Average per County	Coterminous USA (3111 Counties)
Cadastral	0.00036848	0.38	1,200.23
Climate-Precipitation	0.01576630	16.51	51,356.06
Climate-Temperature	0	0	0
Common Land Unit	0.00689263	7.22	22,451.51
Demographics	0.00719055	7.53	23,421.98
Digital Raster Graphics	0.08062340	84.42	262,616.80
Elevation	0.02533636	26.53	82,528.88
FEMA	0.00363046	3.81	11,825.58
GNIS	0.00013570	0.14	441.98
Government Units	0.00019047	0.20	620.41
Hydrology	0.00503700	6.25	19448.861
Hydrologic Units	0.00999607	10.46	32560.482
Land Use Land Cover	0.01775757	18.59	57842.203
Ortho Imagery	0.16668433	174.52	542945.4
Photo Index	0.00004891	0.05	159.28454
Plants	0	0	0
Quad Boundaries	0.00005328	0.06	173.5867
Soils	0.07268086	76.10	236745.38
Transportation-Roads	0.00296294	3.10	9651.3052
Transportation-Rail	0.00098372	1.03	3204.2605
Wetlands-NWI	0.02092951	21.91	68174.244
Legends	0.00009944	0.10	323.87
Metadata	0.00015442	0.16	503.00
Total w/o DOQs (Mb)	0.43752240	403.83	1,428,195.00
DOQs	3.08664115	3321.46	10,333,056.00
Total (megabytes)	3.52416355	3725.29	11,761,251.00

Or 3.73 gigabytes

Or 11.76 terabytes

TABLE 10-1 SUMMARY OF RESULTS