U.S. Department of the Interior
U.S. Geological Survey

The National Map Catalog Technical Discussion Paper

Availability and Growth of *The National Map* in Fiscal Year 2004

December 2004

Larry Moore
Marshall Creighton
Jeff Wendel

Mid-Continent Mapping Center

1 The Catalog Service Checker

The need to quantify Web Map Service (WMS) availability was recognized early in the development of *The National Map* system. In July 2003 a software "service checker" was deployed to monitor the availability of services that contribute data to *The National Map*. Data from this monitoring is stored in Oracle tables in the catalog database.

Service **availability** and **reliability** are synonyms in this document, and are defined for each service as

$$\text{availability} = \frac{\text{'available' data points}}{\text{total data points}}$$

The data used to calculate this value are not continuous; each service is queried about every 17 minutes. The service checker itself is not 100% reliable, so there are some gaps in the data. However, for time spans of weeks enough data points are collected to approximate continuous monitoring.

The basic unit of availability used in this report is one **service-month**, calculated with the above equation for all sample points for one service for one month.

The scope of this document is the availability of individual WMSs that contribute data to *The*

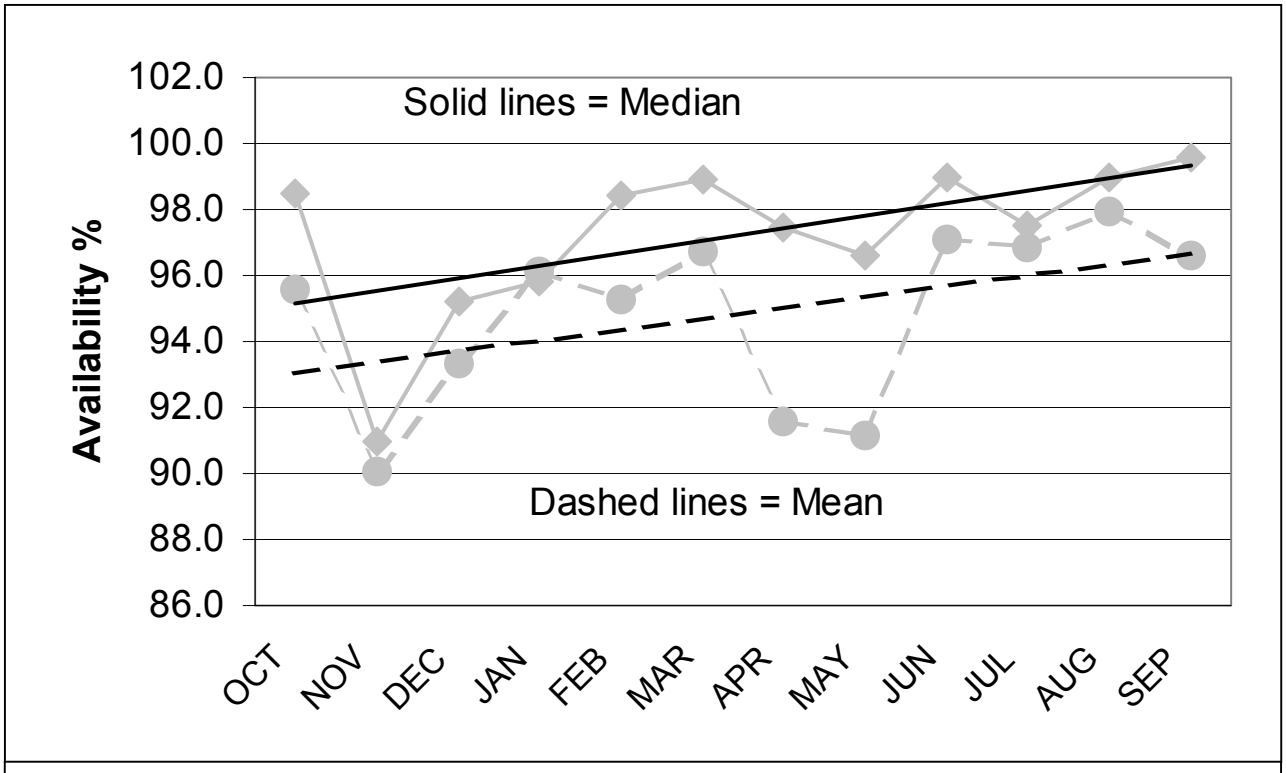

Figure 1. Service availability for FY04. The dashed gray line connects the arithmetic **mean** of all service availability values for each month. The solid gray line connects the **median** of all service availability values for each month. The black lines are linear regression lines. Mean values are lower than median values because it is common for a service to have short periods of extreme unreliability, especially early in the life of the service. These periods drag the mean down, but have little effect on the median. Which statistic is a better measure of reliability is a matter of taste and priorities.

National Map. Overall reliability of *The National Map* system is a much larger issue, affected by many additional factors.

See section 5 for references to documents that contain more information about the operation of the service checker.

2 FY04 Data Summary

2.1 Service Availability

About 120 services contributed data to *The National Map* applications at some time in FY04. Figure 1 illustrates overall service availability by month.

The following table is a different view of the data used to create figure 1. The table lists services that were registered in the catalog and provided data to either the public viewer or the SOLDGR application in FY04. It shows more detail per service than figure 1, but less detail per month. The list is sorted first by the number of months each service was registered with *The National Map*, and second by mean availability. The fourth column is the median of the monthly availability statistics for the service. Medians tend to be higher than means, for the same reasons the median line is above the mean line in figure 1.

Services at the top of the list were available a high percentage of the time all year. Services at the bottom of the list were part of *The National Map* for fewer months, or were not as reliable, or both. Note that this double sorting means that a service can be more reliable than some services above it, and less reliable than some services below it. Note also that if the table were sorted on the median column rather than the mean column, the order would change slightly.

Service Name	Months online	Mean avail %	Median avail %
TerraServer USA	12	99.2	99.9
York County (SC) WMS (NC OneMap)	12	98.9	99.2
Kansas (RNMP-DASC) WMS	12	98.8	99.6
Sedgwick County (KS) WMS	12	98.7	99.6
USGS Catalog WMS	12	98.6	99.5
LA (RAC -new) WMS	12	98.0	99.1
Tahoe Pilot WMS	12	97.9	99.6
Loudoun County (VA) WMS	12	97.3	97.7
USGS Ref WMS (RMMC new)	12	97.2	97.7
USGS Greenness (NDVI) (TEST)	12	97.1	98.2
Washington DC (CUES Region 1) WMS	12	96.9	97.3
Albuquerque NM Pilot WMS	12	96.8	97.1
Denver CO Pilot WMS	12	96.8	97.1
Utah Pilot WMS	12	96.8	97.2
USGS NLCD WMS (EDC)	12	96.7	99.0
USGS NED WMS (EDC)	12	96.6	98.9
USGS BTS Roads WMS (EDC)	12	96.6	99.0
USGS REF WMS (EDC)	12	96.6	98.9
NC OneMap WMS	12	96.4	98.0
Wake County (NC) WMS (NC OneMap)	12	96.3	98.3

Service Name	Months online	Mean avail %	Median avail %
Washington DC (CUES Region2) WMS	12	96.3	97.3
Story County (IA) WMS	12	95.8	98.5
USGS Shuttle Radar Topography Mission	12	95.8	98.5
MSDIS WMS	12	95.6	99.2
Washington DC 10m Hypso WMS	12	95.5	97.0
Washington DC Spot Elevations WMS	12	95.5	97.1
FS Southwestern Region 3 NF WMS	12	95.5	97.7
Washington DC Roads WMS	12	95.5	97.1
Mecklenburg Pilot (Mick_Co) WMS (NC OneMap)	12	95.4	98.3
USGS GTOPO WMS (EDC)	12	95.4	98.5
USGS LANDSAT7 (EDC) WMS	12	95.4	98.5
Washington DC 1m Hypso WMS	12	95.3	97.0
Washington DC 5m Hypso WMS	12	95.2	97.0
Washington DC (Parks) WMS	12	95.1	96.9
133 UA Ortho WMS	12	95.0	97.8
Washington DC Buildings WMS	12	95.0	96.7
Henderson Co NC WMS (NC OneMap)	12	95.0	97.7
Charlotte WMS (National Atlas hydro)	12	94.8	98.4
FS Pike NF Pilot	12	94.7	98.2

Service Name	Months online	Mean avail %	Median avail %
Wash-ID Pilot WMS	12	94.4	95.6
Arkansas (RNMP-CAST) WMS	12	94.2	97.5
Mecklenburg Pilot (sid01) WMS (NC OneM	12	94.1	96.1
BLM PLSS (Boundaries) WMS	12	92.9	92.9
Buncombe County WMS (NC OneMap)	12	91.5	91.1
MetroGIS WMS (MN)	12	90.3	91.4
Missouri (RNMP-MSDIS) WMS	12	88.9	97.9
BLM PLSS WMS	12	86.0	91.3
Oklahoma WMS Server (GEO)	11	99.3	99.8
St. Louis County (MO) WMS	11	98.5	98.3
Buncombe Co (Aerial Photo) WMS (NC OneMap)	11	91.0	91.0
Montana WMS	11	85.4	96.5
Rocky Mountain National Park Pilot	10	97.4	98.6
Missoula County MT Geographic	10	96.3	98.1
Washington DC (CUES DC Orthos) WMS	10	96.2	97.2
Nebraska Department of Natural Resourc	10	95.5	97.0
National Park Service	10	92.9	98.6
FS Bankhead NF	10	88.3	98.2
GNIS WMS	9	97.5	97.6
FS Boundaries WMS	9	96.9	99.0
GNIS/Atlas County Boundaries WMS	9	95.8	97.6
GNIS/Atlas State Boundaries WMS	9	95.6	97.5
Seattle-Tacoma, WA WMS	8	99.2	99.8
Sacramento, CA WMS	7	98.2	97.4
Lubrecht(UMCFC-Montana)	7	97.4	99.7
Texas TNRIS DOQQs UTM Zone14	7	97.4	97.3
Texas TNRIS DOQQs UTM Zone13	7	96.8	97.0
Texas TNRIS DOQQs UTM Zone15	7	96.6	97.4
Shenandoah National Park	7	96.6	97.6
USGS Wetlands WMS	7	95.4	99.2
Shenandoah National Park DRG	7	95.1	97.6
Shenandoah National Park 1 meter image	7	95.1	97.4
Idaho Water Resources (SOLDGR)	7	92.4	96.2
Oahu, HI WMS	7	91.6	97.4
New Jersey 2002 Hi-Res Ortho	6	98.9	99.4
Charleston County 1 Meter Orthos	6	96.2	96.9
Mexico Mde	6	96.1	94.6
Mexico Framework 1M	6	96.1	94.6
Mexican Border, Tamaulipas	6	96.1	94.7
Charleston County Hydrography	6	96.1	96.8
Charleston 1-foot Resolution Imagery	6	96.0	96.7
NHDGeo WMS (RMMC)	6	95.5	94.8
Charleston County Wetlands (NWI Draft)	6	95.0	96.6
Charleston County Maps	6	94.9	96.8

Service Name	Months online	Mean avail %	Median avail %
Charleston County Buildings	6	94.8	96.7
Charleston County Roads	6	94.7	96.7
Charleston Geology	6	94.6	96.5
National Map Corp WMS	6	94.3	99.6
Minnesota LMIC Social	6	86.6	98.6
Minnesota LMIC Health	6	86.6	98.7
Minnesota LMIC Hydro	6	86.5	98.6
Minnesota LMIC Geology	6	86.5	98.6
Minnesota LMIC Boundaries	6	86.4	98.7
Minnesota DOT MapServer	5	99.4	99.7
USGS EDC Ortho 133Urban WMS	5	97.7	99.5
Edgecombe County WMS (NC OneMap)	5	97.1	98.7
RMMC Hazards Service (WMS)	5	96.5	97.1
RMMC Hazards Service (OGCConnector)	5	96.2	95.6
USGS Wetlands Alaska	5	93.5	99.7
OGCConnector v0.5	4	96.0	95.7
NDEP WMS	3	99.8	99.7
NEXRad WMS	3	99.8	99.8
NC OneMap Disaster_Response_Data	3	99.8	100.0
Molokai, HI WMS	3	99.2	99.2
NORM-ED Distance to nearest road	3	99.1	99.0
Merged VDOT, BTS, and CENSUS Roads	3	96.4	95.2
Bozeman / Gallatin MT	3	95.5	99.3
Texas TNRIS 2 WMS	3	90.0	96.5
National Atlas Data (rebtest)	2	99.9	99.8
National Atlas Data	2	99.8	99.7
Kauai, HI WMS	2	99.5	99.2
Rocky's 20 (Backup Shaded Relief) WMS	2	99.3	99.2
Cabarrus County WMS (NC OneMap)	2	32.2	96.1
Hurricane Charley	1	99.9	99.9
Disaster_Response_Data	1	99.9	100.0
Hurricane Charley LANDSAT	1	99.9	100.0
Hurricane Charley_Data	1	99.9	100.0
Hurricane Ivan	1	99.9	100.0
Hurricane Jeanne	1	99.9	100.0
Hurricane Lisa	1	99.9	100.0
Tropical Storm Ivan (New)	1	99.9	100.0
USGS EDC LandCover MODIS	1	99.9	100.0
U.S. Forest Service Roads	1	62.6	89.0

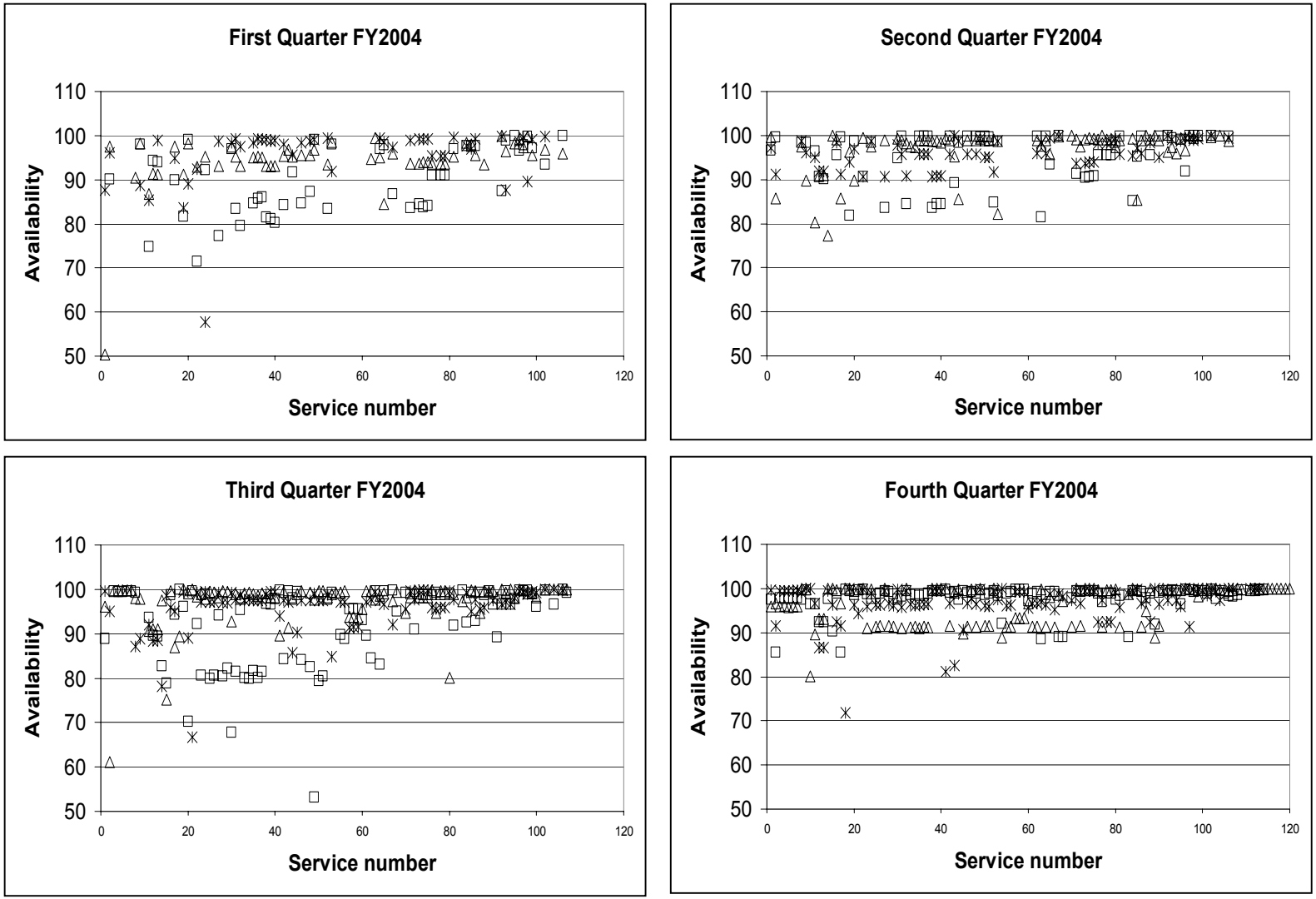

Figure 2. Scatter plots of service availability by quarter. Each point symbol represents one service-month. Within each quarter, stars are the first month, squares are the second, and triangles the third. Services (X axis) are not sorted in any particular order, but are in the same order for all four plots. The general pattern toward higher availability and more services is visually obvious.

Figure 2 is a set of scatter plots of service-month availability points. Four plots are shown, one for each 3-month quarter. Each point on each plot represents one service-month. The graphs visually illustrate that as the year progressed:

- The number of services increased.
- Average reliability improved.
- Variation in reliability between services decreased.

2.2 System Growth

Figures 3 and 4 illustrate the growth of the catalog database in FY04. In both figures, the gray diamonds represent totals for each month, and the black lines are linear regression lines.

The figures illustrate that growth in both services and layers was steady through the fiscal year. Services were added at an average rate of seven per month, and layers at an average rate of 120 per month.

The data includes all services and layers, even though many do not contribute data that are visible in the public viewer. Of the 2,330 layers registered at the end of the year, 850 were visible in the public viewer and 50 more were visible in the SOLDGR system. Most of the remaining layers were either duplicative of some public layer, or judged to be outside the scope of *The National Map*. Inventorying and tracking these layers is not wasted effort, for at least two reasons. First, since they are part of

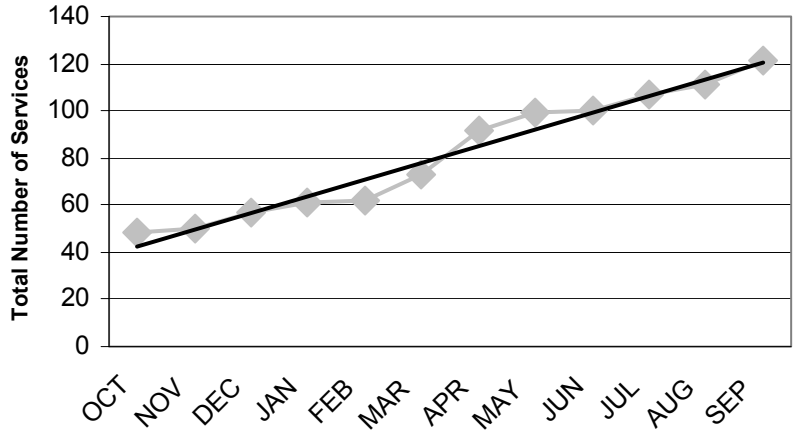

Figure 3. Number of registered services, by month, FY2004

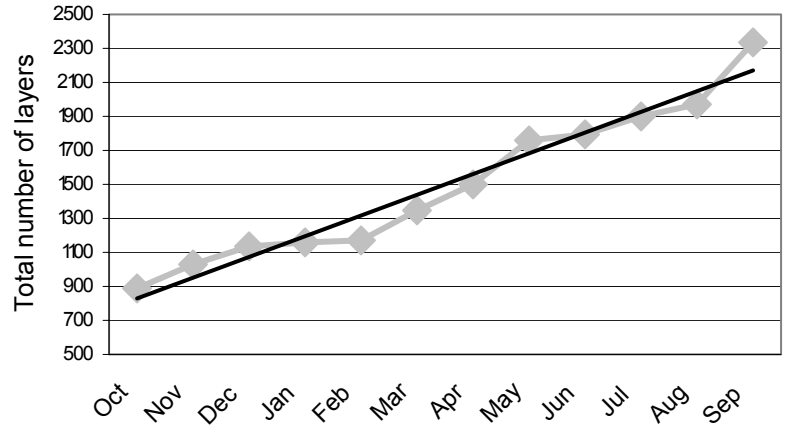

Figure 4. Number of registered data layers, by month, FY2004

services that also manage *The National Map* public layers, the marginal cost of tracking them is almost zero. Second, many have value for specialized scientific applications.

3 Discussion

The metrics presented in this paper are relatively narrow. Even so, they provide several reasons to be optimistic about *The National Map*:

- The reliability of services that contribute to *The National Map* is improving. The variation of availability between services is decreasing. In the last two months of FY04, the median availability of services reached 99%.
- In FY04, *The National Map* grew at the rate of seven services and 120 layers per month. The "doubling time" of the database in FY04 was about 8 months. It is not yet clear if growth is linear or exponential, so no prediction can be made about whether or not this doubling time will be sustained. However, it is very likely that a linear growth rate of at least 100 layers per month can be sustained.[1]
- Though the data holdings are growing fairly rapidly, the resources needed to maintain the system are growing very slowly, or perhaps not growing at all. The catalog support teams (CST) at Mid-Continent and Eastern Region did not grow in FY04. In fact, the resources needed for routine data management of the catalog actually shrank slightly, freeing people to work on training, documentation, and new procedures. This is all evidence that *The National Map's* database-driven design, developed and initially implemented in FY02-03, is robust and scalable.[2]

More and different metrics for evaluating the reliability, growth, and use of *The National Map* are obviously desirable. We do not have a good measure of the overall availability of the public application from an end-user perspective, let alone any measures of customer satisfaction. These are complicated problems that will probably not be addressed in FY05.

However, one new monitoring capability was implemented in November 2004, and will provide some additional data for FY05. We are now sampling and storing information about requests that come to the catalog service. This will not tell us who uses *The National Map* (from the catalog service perspective, the overwhelming majority of requests come from the public viewer, which does not pass along the identity of the "real" user), but it will tell us what data categories and spatial extents are being accessed most frequently.

[1] Whether these growth rates are high, low, or insignificant depends on your perspective. Compared to the size and growth rate of data registries populated with automated crawlers (think Google), *The National Map* doesn't even qualify as background noise. On the other hand, the system is already quite large by traditional GIS standards. It also has goals for data integration, partnership building, data maintenance, service reliability, and cartographic quality that are far higher than anything a fully automated system would attempt. These ambitions make smaller size and lower rates of growth necessary and even desirable. The tradeoffs between speed, volume, and quality are severe, and will be a source of continuing debate for the foreseeable future.

[2] The fact that reports like this one can be written at all is another endorsement of the database approach.

4 Additional Reading

The following documents are posted on the catalog web site, http://mcmcweb.er.usgs.gov/catalog/

- For a more complete discussion of how the service checker behaves, see section 7.2 of "Registering Web Map Services in *The National Map* Catalog." Section 7 of this document also discusses security concerns and system loading.
- Two previous reports on service availability contain more detailed data about the early months of service monitoring, and also more detail about the reliability and performance of the service checker software: "Availability of *The National Map* Web Map Service I" and "Availability of *The National Map* Web Map Service II"
- "Tutorial Introduction to *The National Map* Catalog" contains an explanation of the overall functioning of *The National Map* system, with emphasis on the role of the catalog database and service.