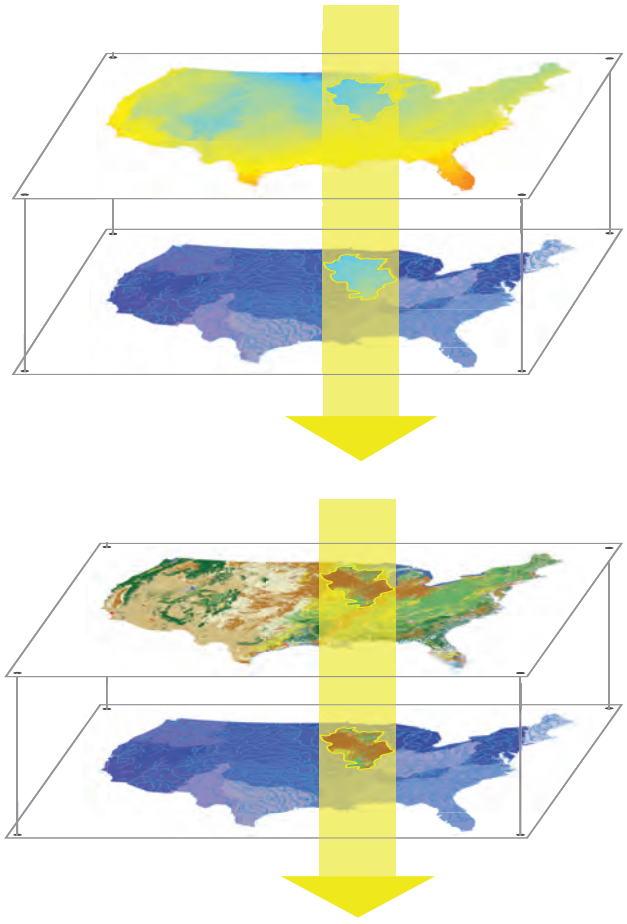
USGS
science for a changing world

Description and Testing of the Geo Data Portal: A Data Integration Framework and Web Processing Services for Environmental Science Collaboration



Open-File Report 2011–1157

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

Description and Testing of the Geo Data Portal: A Data Integration Framework and Web Processing Services for Environmental Science Collaboration

By David L. Blodgett, Nathaniel L. Booth, Thomas C. Kunicki, Jordan I. Walker, and Roland J. Viger

Open-File Report 2011–1157

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

Suggested citation:
Blodgett, D.L., Booth, N.L., Kunicki, T.C., Walker, J.L., and Viger, R.J., 2011, Description and testing of the Geo Data Portal: Data integration framework and Web processing services for environmental science collaboration: U.S. Geological Survey Open-File Report 2011–1157, 9 p.

Contents

Figures

Tables

Conversion Factors and Abbreviations

SI to Inch/Pound

Multiply	By	To obtain
	Length	
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
kilometer (km)	0.5400	mile, nautical (nmi)
meter (m)	1.094	yard (yd)

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

The World Geodetic System 1984 (WGS 84) defines a reference frame for the Earth for use in geodesy and navigation.

Abbreviations

DIF	Data Integration Framework
GDP	Geo Data Portal
GIS	Geographic Information System
NED	National Elevation Dataset
NLCD	National Land Cover Dataset
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
URI	Uniform Resource Identifier
USGS	U.S. Geological Survey
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WPS	Web Processing Service
XML	Extensible Markup Language

Description and Testing of the Geo Data Portal: A Data Integration Framework and Web Processing Services for Environmental Science Collaboration

By David L. Blodgett, Nathaniel L. Booth, Thomas C. Kunicki, Jordan I. Walker, and Roland J. Viger

Abstract

Interest in sharing interdisciplinary environmental modeling results and related data is increasing among scientists. The U.S. Geological Survey Geo Data Portal project enables data sharing by assembling open-standard Web services into an integrated data retrieval and analysis Web application design methodology that streamlines time-consuming and resource-intensive data management tasks. Data-serving Web services allow Web-based processing services to access Internet-available data sources. The Web processing services developed for the project create commonly needed derivatives of data in numerous formats. Coordinate reference system manipulation and spatial statistics calculation components implemented for the Web processing services were confirmed using ArcGIS 9.3.1, a geographic information science software package. Outcomes of the Geo Data Portal project support the rapid development of user interfaces for accessing and manipulating environmental data.

Introduction

Interest in interdisciplinary data sharing is increasing among scientists from the U.S. Geological Survey (USGS), other agencies, and academia. For example, watershed modelers need downscaled general circulation model climate projections and land-cover data summaries to predict streamflow for various climate scenarios. In turn, ecological modelers may be interested in these streamflow predictions for various climate scenarios as they study the effects of altered streamflow conditions on the habitat of biotic communities. As a result, environmental modelers from fields of study including climatology, hydrology, geology, and ecology rely on common, cross-discipline data sources and processing methods.

The Geo Data Portal (GDP) project's purpose was to establish consistency and interoperability among common methods of data access, analysis, and formatting, to assist in model parameterization, model coupling, and data integration. The GDP project achieved this by adopting existing software that implements broadly accepted scientific community-driven standards and by developing connections between them. For example, Web processing services that integrate the geographic reference systems of the climate-and-forecasting community and the geographic scientific community were developed to allow interoperable access to data from either. Thus, the services developed and refined by the GDP project streamline the time-consuming and resource-intensive tasks associated with data access and manipulation that can inhibit the sharing of interdisciplinary environmental science data.

The typical GDP Web processing workflow is illustrated in figure 1. Although the steps shown here may vary in practice, they are representative of a common workflow. The steps represented in the figure are as follows.

1. A user interacts with a user interface to specify required process parameters: 1) a set of polygon features for analysis, 2) a pointer to the gridded dataset to be analyzed, and 3) parameters to configure the GDP Web processing service.

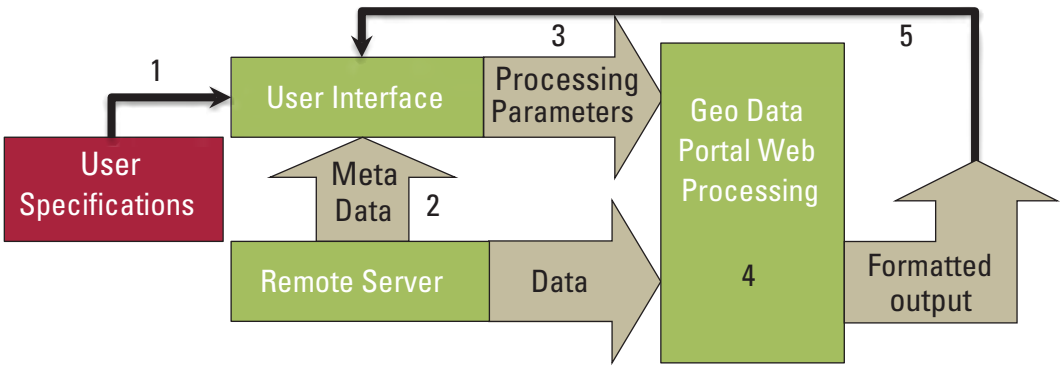


Figure 1. Typical Geo Data Portal Web processing workflow.

2. To assist selection of process parameters in step 1, the user interface accesses discovery metadata from a catalog held on a remote server. Internal dataset metadata, such as data types and time ranges, are found by direct inspection of data resources on remote servers.

3. Based on user input from step 1 and 2, the user interface application generates a request that is submitted to the GDP Web processing services.

4. The GDP Web processing service accesses data and performs calculations and/or transformations as requested in step 3.

5. The GDP Web processing service responds to the user interface application with processing status and, eventually, a file containing the results or a record of a processing error.

Two separate but complementary components of the GDP are described in this report: (1) a data-integration strategy and (2) Web processing and data service software. The data-integration strategy is implemented as a comprehensive system of data-integration standards referred to herein as the Data Integration Framework (DIF). The Web processing and data services software documented in this report implements and enhances existing implementations of standards used in the DIF.

The first section below explains how the components of the DIF work together at a conceptual level. The second section describes implementation details of the Web processing services, and it provides details for readers who want to develop their own software to interact with GDP project services and standards. The final section describes the tests that were performed to confirm that the GDP Web processing services geospatial calculations are consistent with geographic information science software and provides a summary of the results.

Description of the Data Integration Framework

The DIF includes three general types of data: metadata, gridded data, and vector polygon feature data. Web processing services rely on metadata to discover the existence of a gridded dataset and to get a description of the contents of a gridded or feature dataset. Gridded data are the data source that is summarized according to a vector polygon feature dataset. The following sections explain the roles of the three data types in the DIF.

Metadata

The DIF relies on two categories of metadata, (1) data discovery metadata that is used to find data Web services and (2) internal dataset metadata that is found by direct inspection of a data Web service. There are provided by metadata catalogs and gridded data servers respectively. At a minimum, discovery metadata includes a dataset name and a Uniform Resource Identifier (URI) Web address that identifies the location of the dataset. At its most rich, discovery metadata takes the form of an International Organization for Standardization standard metadata record, which is made available by an Open Geospatial Consortium (OGC) standard catalog service that allows a user to search and evaluate various metadata records. Discovery metadata is then used to find internal dataset metadata by querying a remote server about a particular dataset of interest. The names of variables within a dataset and available time ranges are examples of internal dataset metadata.

In contrast to approaches in which each time step of data in a dataset is made available as a separate URI and a separate data Web service, a dataset within the DIF approach is referenced with a URI that resolves to a single data Web service. In the DIF, for example, a gridded dataset holding a time series of temperature, wind, and precipitation can be represented by one URI, and all of the time steps of all of the dataset variables (temperature, wind, precipitation) are accessible through the single URI-resolved data Web service. The use of a single URI per dataset enables the use of the following workflow for data discovery and access:

1. A user interface application is used to query a catalog of discovery metadata to find a data Web service URI.

2. The application uses the data Web service to query the dataset for variable names (like temperature) and content characteristics (like available time range).

3. A Web processing service can then request the user-specified portion of the dataset.

The DIF offers users flexibility in accessing discovery metadata. A simple user-interface implementation using the DIF could offer Web processing services related to a single data source, and in that case no formal datasource catalog is required. For compiling small data-source catalogs, a list of dataset URIs and accompanying discovery metadata made available by a user-interface application may satisfy data discovery needs. To allow users to search for and combine large data-source catalogs, user-interface applications can implement the OGC Catalogue Service for the Web (Open Geospatial Consortium, Inc., 2007a) standard for metadata search and retrieval. It is important to note that the DIF uses discovery metadata only for data Web service URI discovery.

Applications implementing the DIF should rely on data Web service introspection rather than discovery metadata services to access internal dataset metadata. Given a data Web service URI, a user interface implementing the DIF retrieves internal dataset metadata by use of a dynamic query to the data Web service rather than relying on static or cataloged metadata. This is useful if the dataset tends to change frequently, thus making cataloged internal metadata obsolete.

Gridded Data

Commonly, a modeler or analyst needs to determine statistics about gridded data for a set of polygons related to their modeling or analysis. The following paragraphs summarize the types of gridded data included in the DIF and the data Web service protocols used to access them.

Gridded data are collections of cells referenced to geographic axes. The geometry and arrangement of cells in gridded data accessible to GDP Web processing services may be defined in several ways: 1) as an origin and cell size; 2) as a list of distances from the origin of each axis; 3) as two two-dimensional grids of explicitly defined positions so that each grid cell is defined as a point. The first and second cases are regular and rectilinear grids, respectively; the third is actually a set of points that happens to be laid out in a grid not aligned with a particular axis. This third grid type is especially useful if a grid is defined in a projection unknown to processing services, because grid cell locations can be redefined and analyzed using a known coordinate reference system without the need to resample to new grid geometry.

A gridded dataset can contain time-invariant or time-series data. These two data types require special processing considerations. High resolution time-invariant gridded datasets, like digital elevation models, that exceed server capacity must be resampled to a lower resolution or processed using a less demanding analysis method, like tiling. Time-series gridded datasets are generally not as spatially massive per time step as time-invariant gridded datasets, but time-series gridded datasets can be extremely large with regard to the time dimension. Time-series data require functionality to handle specification of time period processing parameters and processing services capable of handling time series appropriately. Two standard Web service protocols have been adopted by the DIF to handle spatially massive or temporally extensive gridded data sources: the OGC Web Coverage Service (WCS) (Open Geospatial Consortium, Inc., 2006a) and the Open-source Project for a Network Data Access Protocol (OPeNDAP; Open-source Project for a Network Data Access Protocol, Inc., 2010). Requests for data from the WCS allow spatial subsetting and resampling of gridded datasets in case they are too large. The OPeNDAP allows flexible access to time-series gridded datasets but requires the requesting Web processing service to request data by grid index position rather than by location or time. Data servers that implement the WCS and OPeNDAP standards, in most cases, also support the OGC Web Map Service (WMS) protocol (Open Geospatial Consortium, Inc., 2006b). Distinct from the WCS and OPeNDAP, the WMS provides a rendered screen-resolution image of the requested dataset. Web processing services do not use Web map services, but applications implementing DIF components are able to take advantage of the WMS for visualization and verification of data.

Vector Polygon Feature Data

Vector polygon feature data have explicit geometry, like points or polygons, which are described using geographic coordinate pairs. The GDP Web processing services analyze gridded data on a per-polygon-feature basis, but the DIF adopted standards are capable of handling any vector features. The shapefile format is an example of a widely used encoding for feature data, like watershed polygons or sampling points. Shapefile is a storage file format that works seamlessly with virtually all geographic-information software and servers. This allows a shapefile created with desktop geographic information system (GIS) software to be loaded into a geospatial server without special preparation or considerations. The DIF uses the OGC Web Feature Service (WFS) protocol (Open Geospatial Consortium Inc., 2005) with the feature data encoded in the Geography Markup Language simple features profile (Open Geospatial Consortium, Inc., 2010). Similar to the OPeNDAP and WCS servers, the geospatial servers that support the WFS also tend to implement the WMSs for visualization. Applications implementing the DIF can access the geospatial server WMS directly to allow users to verify that feature data were transferred to a geospatial server as expected.

Geo Data Portal Implementation Details

Throughout design and implementation of the GDP DIF and Web processing service, a generalized graphical user interface and utility services were implemented as proof of concept to ensure the utility of the services and ease of specification of processing parameters in a Web browser-based user interface application. Important details about the Web processing services and findings from research and development of them are presented below.

Web Feature Service and Time Period Specification

The GDP implementation provides a shapefile upload service that loads a user's shapefile into a geospatial server to make it available as a WFS for processing. A fully formed WFS request is provided to the Web processing service as an input parameter.

Feature data is retrieved by the Web processing service. Processing is performed using the coordinate reference system of the user-supplied feature data. The start and end of the analysis time period is specified using the Extensible Markup Language (XML) Schema standard date-time data type (World Wide Web Consortium, 2004).

Gridded Data Access

Gridded data must be available from a data Web service that meets the OPeNDAP or WCS standard to be accessible to a GDP Web processing service. A data Web service URI and a data-type identifier are required for access because that information is used with information from the WFS and time period specifications to formulate a full request for the subset of gridded data being analyzed. When these requirements are met, Web processing services can then retrieve the gridded data for the user during process execution.

Geo Data Portal Web Processing Service Overview

Geo Data Portal Web processing services are made available using the OGC Web Processing Service (WPS) standard (Open Geospatial Consortium, Inc., 2007b). The WPS standard specifies a language for describing Web processing algorithms. It provides a mechanism for explicit specification of formats and encodings of input data, output data, and process parameters. By using the WPS and DIF standards, software developers are able to rapidly create user interfaces to parameterize and execute Web processing services utilizing remote computer resources and decentralized data repositories. As outlined above, datasets for processing are provided as URIs where Web processing services can retrieve them. The WPS requests can be executed 1) synchronously for immediate mapping or analysis, or 2) asynchronously, in which the processing status is provided and the output file is made available upon process completion. This is best described with examples. A request to upload a shapefile is made synchronously; the Web browser user interface application uploads a file and waits for a server response. A request to run a large analysis is made asynchronously; the Web browser user interface application sends a processing request and receives a link to a processing status document that it can check periodically.

Area-Weighted Statistics Generation

The GDP Web processing services include an algorithm to generate area-weighted statistics of a gridded dataset for a set of vector polygon features. Several parameters discussed in the preceding sections are required in a request to execute this service:

1. a fully specified WFS request,
2. a feature attribute name to label output data,
3. a gridded dataset URI,
4. a data type to be accessed from the gridded dataset, and
5. an optional time range.

By use of the bounding box that encloses the feature data and the time range, if provided, a subset of the gridded dataset is requested from the remote gridded data server. Polygon representations are generated for cells in the retrieved grid. The polygon grid-cell representations are projected to the feature data coordinate reference system and used to calculate per grid-cell feature coverage fractions. Area-weighted statistics, mean, minimum, maximum, standard deviation, variance, count, and weighted sum are then calculated for each feature using the grid values and fractions as weights (West, 1979). The last step is repeated for each time step within the specified time range or all time steps if a time range was not supplied.

Categorical Coverage Fraction Computation

This GDP Web processing service is used with categorical gridded data to assess the coverage fraction of each category in the gridded data for each feature. For example, a gridded dataset of land cover types can be summarized to percentage of each land cover type in each feature provided. The WPS request for computation of the categorical coverage fraction contains similar parameters to those for the request for area-weighted statistics:

1. a fully specified WFS request,
2. a feature attribute name to label output data,
3. a gridded dataset URI, and
4. a data type to be accessed from the gridded dataset.

This service does not process gridded time series data. By use of the feature dataset bounding box, a subset of the gridded dataset is requested from the remote gridded data server. The location of each grid-cell center is then projected to the feature dataset coordinate reference system. For each grid cell in the subsetted grid, the grid-cell center is tested for inclusion in each feature in the feature dataset. If the grid-cell center is in a given feature, the count for that cell's category is incremented for that feature. After all the grid-cell centers are processed, the coverage fraction for each category is calculated for each feature.

Testing of Geo Data Portal Web Processing Services

To show the GDP Web processing services perform geospatial calculations consistent with geographic information science software, several tests have been developed as examples that are intended to demonstrate a broad range of service functionality. The GIS processing for these examples were carried out using ArcGIS 9.3.1 desktop software, and the results of the tests of the GDP Web processing services and ArcGIS 9.3.1 were compared. ArcGIS 9.3.1 was used owing to its wide use and a general consensus that it represents the geographic information science standard and is applicable for the selected use cases. It was necessary for the user doing the testing to download and manipulate raw data into appropriate formats for use with ArcGIS 9.3.1, while these same data were accessed directly from data Web services by the GDP Web processing service.

Testing of Area-Weighted Statistics Generation

The generation of area-weighted statistics of a gridded dataset for a set of polygon features was tested with the GDP processing services and ArcGIS 9.3.1. The method that ArcGIS 9.3.1 uses to calculate statistics of gridded data per feature is substantially different than the method associated with the GDP Web processing services. ArcGIS 9.3.1 first converts the feature data into a gridded representation of the features with the same layout (for instance, cell size) as the gridded dataset and then performs analysis on a raster basis. The GDP area weighted statistics Web processing service automatically converts all data to a vector representation. When the cell size of the gridded dataset is very small relative to the size of the features, ArcGIS 9.3.1 and the GDP Web processing service return very similar results; however, as the size of the gridded data cells approach or exceed the size of the (vector) features, major differences are likely to occur in the derived statistics. To achieve equivalent results using the different analysis methods, the gridded dataset was resampled to a fine enough resolution to accurately define the feature geometry using a gridded representation.

Figure 2 illustrates the process of resampling the coarse gridded dataset to a finer resolution so that polygon features can be more accurately defined. Figure 2*A* shows an example set of polygon features overlaying a grid whose cell size is so coarse relative to the size of the features that the ArcGIS method would produce different results than the GDP method. Figure 2*B* shows the polygon features converted to the coarse cell size of the gridded data; note the pixilation in the feature boundaries. Figure 2*C* shows the entire analysis zone resampled to a grid 100 times finer than the original input gridded data. Figure 2*D* is an exploded view showing that the fine resolution version of the features more accurately represents the detail of the input feature data polygons.

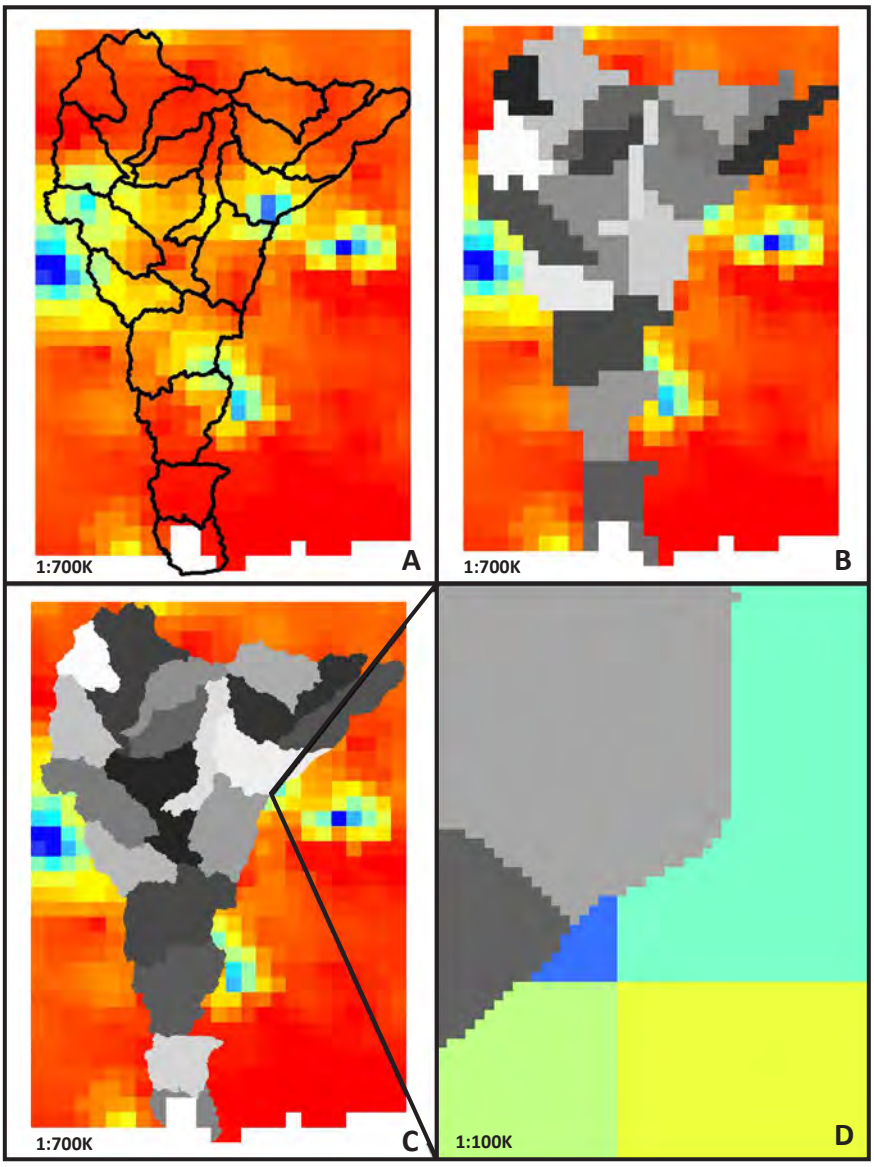


Figure 2. Examples of more accurate definition of gridded polygon features by resampling a gridded dataset to finer resolutions. *A*, The original 0.125-degree grid with feature polygons. *B*, Feature polygons resampled to 0.125 degree grid. *C*, Feature polygons resampled to a 0.00125-degree grid. *D*, View showing a detail of a 0.00125-degree grid that can more accurately reproduce zone polygons for analysis.

The general workflow used in ArcGIS 9.3.1 was as follows:

1. Gridded data were downloaded.
2. The gridded data were imported from their native format into ArcGIS GRID format.
3. The gridded data were resampled to a very fine resolution.
4. The resampled gridded data were projected into an Albers equal area projection.
5. The (vector) feature data were projected into an Albers equal area projection.
6. The feature data were rasterized using the same grid layout as the resampled gridded data to verify consistent polygon geometry reproduction.
7. Statistics, mean, minimum, maximum, and standard deviation of the gridded data were calculated for the feature data using a built in ArcGIS 9.3.1 tool.

The general workflow used with the GDP Web processing service was as follows:

1. Feature data were uploaded to the GDP WFS.
2. Gridded data URIs and processing parameters were selected.
3. The request was submitted to the Web processing service.

Summary of Datasets Used for Testing GDP Web Processing Services

Tests of the ability of the GDP Web processing services to perform coordinate reference system conversions and pass data between software components were conducted. The weighted statistics analysis outlined in the area-weighted statistics generation section was applied to three datasets, each with different cell sizes and spatial referencing characteristics. The categorical coverage fraction analysis was applied to one dataset.

Gridded Meteorological Observations Dataset

One of the primary uses that the area-weighted statistics Web processing service serves is the analysis of gridded meteorological and/or downscaled general circulation model climate change scenarios according to a modeler's areas of interest. As previously mentioned, watershed modelers often use data resulting from this analysis to predict streamflow for various climate scenarios. The Gridded Meteorological Observations dataset (Maurer and others, 2002) uses a 12-kilometer (km) cell size, with position expressed as the latitude and longitude of the cell centers. This "unprojected" data lacks datum information. For grids lacking datum specification, both ArcGIS 9.3.1 and the GDP processing services presented here assume latitude-longitude pairs using the World Geodetic System 1984 geodetic datum. Weighted statistics analysis was applied to the dataset using the GDP processing service and ArcGIS 9.3.1. For the ArcGIS 9.3.1 analysis, the grid was resampled from its native 0.125-degree resolution to a projected resolution of about 140 meters (m). The feature data used for this analysis are shown converted to the analysis grid in figure 3.

Figure 3. Feature data used for the gridded meteorological test case shown converted to the grid resolution for analysis.

River Forecasting Center Quantitative Precipitation Estimates Dataset

Radar rainfall estimates can be attributed to watersheds using the GDP area-weighted statistics generation service. Multisensor quantitative precipitation estimates are derived by the National Weather Service River Forecasting Centers and distributed by the National Precipitation Verification Unit (National Precipitation Verification Unit, 2010). The positions of these data are expressed on a polar stereographic projected grid with a projected spatial resolution of 4.762 km. Projected to an Albers Equal Area projection, the grid has a resolution of about 4.3 km in the vicinity of the analysis presented here (89 degrees west and 43 degrees north) (Fulton, 1998). Weighted statistics analysis was applied to the dataset using the GDP processing service and ArcGIS 9.3.1. The 4.3-km grid was resampled to a resolution of about 43 m. The feature data used for this analysis were rasterized according to the layout of the analysis grid (shown in grey) in the background of figure 4.

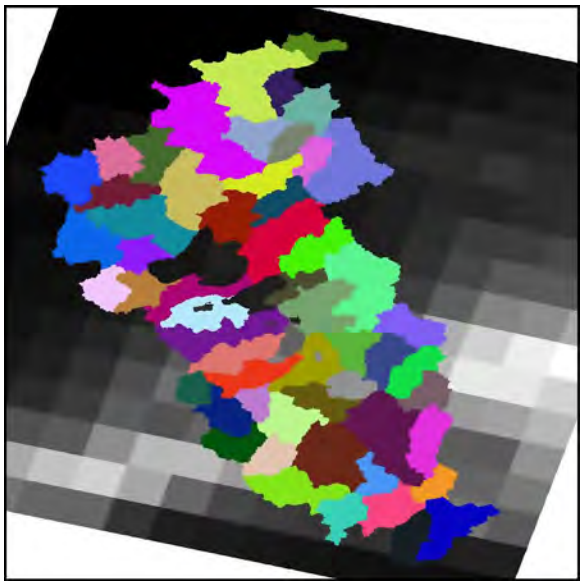


Figure 4. Feature data used for the River Forecasting Center Quantitative Precipitation Estimate test case shown converted to the grid resolution for analysis.

National Land Cover Dataset

The National Land Cover Dataset (NLCD; U.S. Environmental Protection Agency Multi-Resolution Land Characteristics Consortium, n.d.) is the best available nationally contiguous source for land-cover classes, impervious surface estimates, and forest canopy information. As was mentioned previously, watershed modelers often analyze land cover data and create land cover data summaries to predict streamflow for various climate scenarios in areas of interest. The NLCD is a 30-m resolution raster dataset projected using an Albers equal area projection based on the North American Datum of 1983 (NAD 83). For the test presented here, computation of the categorical coverage fraction was applied to one land cover class, deciduous-forest, by using the GDP processing service and ArcInfo 9.3.1. While the GDP Web processing service returns all classes, for testing purposes the output from one was considered. Because the categorical coverage fraction Web process calculates statistics based on grid cell centers, the NLCD grid was not resampled prior to calculating statistics in ArcGIS. In addition, comparability of analysis results would not be hampered since the NLCD cell size was sufficiently fine relative to the feature sizes.

National Elevation Dataset Digital Elevation Dataset

The National Elevation Dataset (NED) uses high precision ground surface elevation data for the United States and is a 1/9th arc-second (roughly 3-m) resolution grid based on the NAD 83 datum (U.S. Geological Survey, n.d.). Weighted statistics analysis was applied to the dataset using the GDP processing service and ArcGIS 9.3.1. Because the NED is of such high resolution, no resampling was performed prior to calculating weighted statistics in ArcGIS.

Testing Results

The testing results presented in table 1 demonstrate that analyses performed using the GDP Web processing service produce results that are equivalent to analyses performed using standard GIS software. The area weighted mean difference of spatial means in table 1 represents the average difference of the GPD Web processing service analysis and the ArcGIS 9.3.1 processing using the area of the feature to weight its contribution to the average. The mean of spatial means is the overall average value of the set of features used for the test. The mean of spatial means and area weighted mean difference of spatial means were used to calculate the standard error between the two. A pooled two-sample T-test was conducted for each set of analysis output to test for a statistical difference between the GPD Web processing service and ArcGIS 9.3.1 analysis results. The T distribution value was used to calculate the T distribution probability (p-value) associated with that T value. A p-value less than or equal to 0.05 was considered to represent a significant difference in the results of the analyses.

The largest standard error of results found using the two processing methods, 2 percent, was the result of the test of the categorical coverage fraction algorithm. Three factors may have contributed to the relatively large difference of that test case: the feature dataset had quite small polygon features, the land cover dataset was binary (the variables were "forested" or "not forested"), and the ArcGIS 9.3.1 and GDP Web processing service analysis methods differed slightly. Test results of the area-weighted statistics generation services showed standard errors of less than 1 percent. All T-test p-values are much greater than the 0.05 statistical difference threshold. These very small standard errors and high p-values indicate that the GDP Web processing service's geospatial calculations are consistent with standard geospatial methods.

Summary

The GDP data integration framework and Web processing services were developed based on the open-standard data and Web service specifications that were found to best facilitate interdisciplinary environmental data integration. The GDP Web processing services allow environmental modelers and other environmental data users to extract aerially weighted and categorical coverage statistics from gridded data sources for polygonal features of interest.

The software has been implemented to enhance scientific community-supported open-source software from the geographic, atmospheric, and oceanographic science communities. Adoption of open standards and technology from multiple scientific communities has allowed development of generalized spatial processing services and user interface implementations that are capable of accessing any dataset made available by one of the framework standards. GDP Web processing service GIS calculations were confirmed to be consistent with ArcGIS 9.3.1 geospatial analysis software.

Table 1. Results of testing show small errors between spatial averages derived using ArcGIS 9.3.1 and the GDP Web processing service.

[mm/d, millimeters per day; mm, millimeters; m, meters; %, percent]

Dataset	Area weighted mean difference of spatial means	Mean of spatial means	Standard error	T-test p-values
Gridded Meteorological Observations	0.0196 mm/day	23.0 mm/day	0.085 %	0.997
River Forecasting Center Quantitative Precipitation Estimates	.0320 mm	12.6 mm	.254 %	0.993
National Land Cover Dataset Categorical Percent Coverage	.0057 %	.28 %	2.028 %	0.962
National Elevation Dataset Digital Elevation Model	.0127 m	286.8 m	.004 %	0.999

References Cited

Fulton, R.A., 1998, WSR–88D polar-to-HRAP mapping: Silver Spring, Md., National Weather Service, Office of Hydrology, Hydrologic Research Laboratory Technical Memorandum, 33 pages.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., and Nijssen, B., 2002, A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Journal of Climate, v. 15, no. 22, p. 3,237–3,251.

National Precipitation Verification Unit, 2010, NPVU River Forecasting Center data access: National Weather Service, accessed January 15, 2011, at *http://www.hpc.ncep.noaa.gov/npvu/data/*.

Open Geospatial Consortium, Inc., 2005, OpenGIS Web Feature Service Implementation Specification: accessed June 10, 2010, at *http://www.opengeospatial.org/standards/wfs*.

Open Geospatial Consortium Inc., 2006a, Web Coverage Service Implementation Standard: accessed June 10, 2010, at *http://www.opengeospatial.org/standards/wcs*.

Open Geospatial Consortium, Inc., 2006b, OpenGIS Web Map Service Implementation Specification: accessed June 10, 2010, at *http://www.opengeospatial.org/standards/wms*.

Open Geospatial Consortium, Inc., 2007a, OpenGIS Catalogue Service for the Web Implementation Specification: accessed June 10, 2010, at *http://www.opengeospatial.org/standards/specifications/catalog*.

Open Geospatial Consortium, Inc., 2007b, Web Procesing Service: accessed June 10, 2010, at *http://www.opengeospatial.org/standards/wps*.

Open Geospatial Consortium, Inc., 2010, Open GIS Geography Markup Language (GML) Encoding Standard: accessed June 10, 2010 at *http://www.opengeospatial.org/standards/gml*.

Opensource Project for a Network Data Access Protocol, Inc., 2010, OPeNDAP Web site: accessed July 21, 2010, at *http://opendap.org/*.

U.S. Environmental Protection Agency, Multi-Resolution Land Characteristics Consortium [n.d.], 2001 National Land Cover Dataset: accessed August 19, 2010, at *http://www.epa.gov/mrlc/nlcd-2001.html*.

U.S. Geological Survey [n.d.], National Elevation Dataset: accessed August 20, 2010, at *http://ned.usgs.gov/*.

West, D. H. D., 1979, Updating Mean and Variance Estimates: An Improved Method: Communications of the ACM, v. 22, no. 9, p. 532–535.

World Wide Web Consortium, 2004, XML Schema Part 2: Datatypes (: XML Schema Working Group, accessed February 15, 2010, at *http://www.w3.org/TR/xmlschema-2/#dateTime*.

Glossary

Area-weighted statistics—Statistics based on a sample population weighted by the fraction of each sample intersecting the area being used for analysis.

Categorical coverage fraction—The area of a given analysis region covered by a single category in a categorical data type normalized against the total area of the analysis region.

Dataset bounding box—A geospatial rectangle that fully encloses a given dataset.

Data Web service—An Internet available method of describing and providing access to data resources.

Derivative data—A calculated summary of a dataset.

Discovery metadata—Information used to describe a dataset's contents and location or access method.

Downscaling—A process used to refine global climate model projections of future conditions from course grid and time resolution to finer grid and time resolution.

Implementation—A piece of software developed to conform to, or implement, a particular standard or design guideline.

Internal dataset metadata—Information inherent to a dataset that is found by inspecting the dataset directly.

Polygon feature—A spatially distinct entity defined using many precise location pairs arranged in a line forming a closed ring.

Uniform Resource Identifier—A universally unique Internet address.

User interface application—A computer program, often embedded in a Web page, meant to provide human access to programmatic tools.

Web processing service—An Internet available method of describing and providing access to processing algorithms and other server resources.

Publishing support provided by the U.S. Geological Survey
Science Publishing Network, Columbus Publishing Service Center

For more information concerning the research in this report, contact the
Director, Wisconsin Water Science Center
U.S. Geological Survey
8505 Research Way
Middleton, Wisconsin 53562
http://wi.water.usgs.gov/