



CIESIN

Center for International Earth Science Information Network
Columbia University

61 Route 9W
Palisades, New York 10964 USA

Telephone: +1 845 365 8988
Facsimile: +1 845 365 8922

**Provision of Context-Sensitive Client Support,
Access to Socioeconomic, Hazards, and Biodiversity
Data, and Web Processing Services
in the
GEO Architecture Implementation Pilot – Phase 2 (AIP-2)**

CFP Issuance Date: 26 June 2008

Kickoff Response Due Date: 1 September 2008

Robert S. Chen

bchen@ciesin.columbia.edu

+1 845 365 8952

Gregory Yetman

gyetman@ciesin.columbia.edu

+1 845 365 8982

Table Of Contents

| | | |
|----------|--|----------|
| 1 | Overview | 3 |
| 2 | Proposed Contributions | 3 |
| 2.1 | Societal Benefit Area Alignment and Support..... | 3 |
| 2.2 | Component and Service Contributions | 6 |
| 2.3 | Architecture and Interoperability Arrangement Development..... | 9 |
| 3 | Description of Responding Organization..... | 9 |

CIESIN Response to the GEOSS AIP-2 CFP

1 Overview

CIESIN is pleased to respond to Phase 2 of the GEOSS Architecture Implementation Pilot (AIP) Call for Participation (CFP). CIESIN plans to contribute 3 major enterprise components:

- 1) A standards-compliant web-based client that supports at least four OGC specifications: Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), and Web Map Context (WMC);
- 2) A set of data and associated data services related to population, urban area extents, species location, protected areas, and hazard distributions; and
- 3) A Web Processing Service (WPS) that permits geospatial analysis of user-defined areas for selected variables on the topics listed above.

CIESIN will focus on two of the Societal Benefit Areas and scenarios outlined in the CFP: Disaster Response and Protected Areas. CIESIN has extensive experience in linking data and information products with applied research and decision making in both of these arenas.

In addition, CIESIN will work to ensure the visibility of the above components in the available GEOSS catalogs, service registry, and portal interfaces, and will support development and testing of service chaining.

CIESIN also plans to contribute to definition of specific requirements for digital rights management, consistent with the GEOSS Data Sharing Principles and with the proposed implementation guidelines for the Data Sharing Principles under development by task DA-06-01.

CIESIN is the World Data Center for Human Interactions in the Environment of the International Council for Science (ICSU), an Associate (University) member of the Open Geospatial Consortium (OGC), and a member of the Global Spatial Data Infrastructure Association (GSDI). CIESIN also operates the Socioeconomic Data and Applications Center (SEDAC), one of the distributed active archive centers in the Earth Observing System Data and Information System (EOSDIS) of NASA.

2 Proposed Contributions

CIESIN has been developing and implementing a wide range of OGC-compliant data, tools, and services relevant to GEOSS. These are widely applicable to the GEOSS SBAs, but are especially relevant to the four SBAs to be addressed in AIP-2. However, for resource reasons, CIESIN will focus on two of the four SBAs as described below.

2.1 Societal Benefit Area Alignment and Support

CIESIN plans to support two of the four SBAs identified in the CFP:

- 1) Disaster Response SBA; and
- 2) Biodiversity SBA: Protected Areas.

Disaster Response

Key issues throughout the disaster life cycle are understanding of the population potentially exposed to hazards, identification of those actually exposed, assessment of the vulnerability of exposed populations to both direct and indirect effects of hazards, and the locations and condition of populations most in need of disaster relief, rehabilitation, and reconstruction efforts. Therefore, provision of geospatially referenced population data is important to disaster management and decision making. CIESIN's Gridded Population of the World (GPW) version 3 dataset is currently accessible through a WMS service, and will be made available for selected areas through a WCS as part of AIP-2. CIESIN also provides access to a number of hazard-related datasets that delineate areas of high-risk from one or more natural hazards and will expand availability of these datasets through WMS and WCS services.

Disaster managers not only need to display population data in conjunction with hazard data, but also to obtain usable information from the integration of these data. Therefore, CIESIN will provide a WPS designed to support spatial queries in which, for example, statistics on populations affected are returned for user-defined areas, e.g., a buffer zone around a hurricane path. A prototype service was demonstrated at GSDI-9 in Santiago, Chile and is being implemented by CIESIN as part of its NASA-supported Socioeconomic Data and Applications Center (SEDAC).

Because disaster managers typically have specific regions and hazards of primary interest, user interfaces should be customizable to focus on their specific needs. When a specific hazard occurs, they should be able to visualize events easily, bring in relevant data, and save views of the data for later use or use by others. The WMC specification enables such functionality in WMC-enabled map viewers. CIESIN will make its WMC-enabled map client available and adapt it as needed for use in the Disaster Response scenario.

CIESIN will ensure that these data and services are registered and visible in the appropriate catalogs, service registries, and portals, and also make sure that they are compliant with the GEOSS Data Sharing Principles and proposed data sharing guidelines being developed by GEOSS Task DA-06-01. CIESIN will help organize and participate in service-chaining scenarios that utilize different combinations of CIESIN-provided and other AIP-2 components.

CIESIN has extensive experience in working with the disaster management community and will draw on its partnerships and contacts in the development and testing of the scenarios developed for this SBA. For example, CIESIN is currently working with the UN International Strategy for Disaster Reduction and with the UN Office for the Coordination of Humanitarian Affairs on two different multi-hazard risk assessment projects. CIESIN has strong contacts with the UN Geographic Information Support Team (GIST), UN World Food Programme, ReliefWeb, the United Nations Environment Program (UNEP) Post-Conflict and Disaster Management Branch, the ProVention

Consortium, the Humanitarian Information Unit of the U.S. Department of State, and other governmental and nongovernmental organizations involved in humanitarian assistance. CIESIN's SEDAC is also part of a network of NASA-supported data centers that provide relevant remote sensing data that may be useful to access in support of this scenario.

Protected Areas

Management of protected areas requires data not only on the protected areas themselves, but on the human pressures (including urbanization and land use change) and the spatial distribution of key species surrounding the protected areas. In addition to CIESIN's GPW version 3 dataset mentioned above, CIESIN also offers a unique database on urbanization, the Global Rural-Urban Mapping Project (GRUMP) collection, which combines census and remote sensing data to improve understanding of urban settlements and spatial extent. GRUMP has already been used widely in scientific studies of land use, land cover, and biodiversity change, and is potentially very useful for management of protected areas. CIESIN is also in the process of releasing a new gridded product on the spatial distribution of species. Both of these datasets are of significant potential value to protected area scenarios, and will be made available via WMS, WFS, and WCS for selected regions.

Protected area and biodiversity decision-makers increasingly require information about protected areas that places them in a bioregional context. CIESIN computes, on an annual basis, a national-level indicator that quantifies the extent of protected areas by biome. CIESIN also collaborates with Yale University in producing the Environmental Performance Index, which includes specific biodiversity indicators derived from spatial data on protected areas and other information, including the Human Footprint (Sanderson *et al.*, 2002). CIESIN will make selected variables from these global datasets available through a WMS.

As in the case of disaster response, the ability to perform spatial queries on combined datasets is an important function in support of decision making. Protected area managers are likely to want to estimate the total numbers of people residing within a certain distance of protected area borders, or in areas where specific species are known to live. CIESIN will provide a WPS that is able to compute such statistics for selected population, land use, and biodiversity variables for user-defined areas.

Protected area managers also have specific regions of interest and may want to group different types of species together when visualizing and analyzing data. CIESIN will make available its WMC-enabled map client, adapting the user interface if necessary to enable users to customize the map visualizations they generate and save these for later use or for sharing with other users.

CIESIN will ensure that these data and services are registered and visible in the appropriate catalogs, service registries, and portals, and also make sure that they are compliant with the GEOSS Data Sharing Principles, proposed data sharing guidelines being developed by GEOSS Task DA-06-01, and possible data sharing approaches developed by the GEO Biodiversity Observation Network. CIESIN will help organize

and participate in service-chaining scenarios that utilize different combinations of CIESIN-provided and other AIP-2 components.

CIESIN has extensive experience in working with the biodiversity and protected area community and will draw on its partnerships and contacts in the development and testing of the scenarios developed for this SBA. For example, CIESIN has worked closely with the Ramsar Bureau on Wetlands of International Importance, the World Conservation Union (IUCN), and Birdlife International on spatial data needs of wetlands managers. CIESIN has collaborated with NatureServe in making the species grid data available, and has worked with the World Conservation Monitoring Center (WCMC), the Millennium Ecosystem Assessment (MA), and the Global Biodiversity Information Facility (GBIF) on various biodiversity data activities. In its capacity as the World Data Center for Human Interactions in the Environment, CIESIN is also part of a collaboration with the World Data Center for Biodiversity and other ICSU World Data Centers and groups to establish a new WDC for Human Health and Biodiversity in South Africa. CIESIN has also conducted research on the use of remote sensing and geospatial data to support management of international wetlands.

2.2 Component and Service Contributions

CIESIN expects to register three types of components and services with GEOSS:

- 1) A set of WMS, WFS, and WCS services with global and regional data as described in section 2.1 (some of these have already been registered);
- 2) A map client that supports the WMC specification; and
- 3) A test version of a WPS capable of performing spatial queries between population data provided by CIESIN and one or more user-selected variables.

CIESIN's population data services will include WCS and WMS services for global surfaces (excluding Antarctica) of population counts (persons), densities (persons per square km) and a quality measure of population-weighted mean geographic unit area (square km). The data are available at two resolutions: 2.5 arc-minutes and 30 arc-seconds. The population variables are available for the years 1990-2015 in increments of 5 years for GPW at 2.5 arc-minutes and 1990, 1995, and 2000 for GRUMP at 30 arc-seconds. Additionally, vector locations for populated places with associated population estimates and a vector extent of urban extents will be made available via WFS and WMS.

Natural hazard data services presently available as WMS layers will be made available as WCS layers and registered in the GEOSS registry. These layers include information on the occurrence and mortality and economic risks of six hazards: flood, drought, cyclones, landslides, earthquakes, and volcanoes. The data are available globally at a resolution of 2.5 arc-minutes, but only selected regions may be offered depending on server and storage capacity. Improved versions of these layers and additional hazards (fires, tsunamis) based on ongoing project work will be added and registered if time permits.

Conservation-related layers to be made available include national-level indicators related to conservation management that will be available as a WFS and WMS service. Selected WFS, WMS and WCS layers for specific regions of interest during the AIP pilot from

CIESIN's large collection of species occurrence grids will be published to support different conservation scenarios. The species grids include data for global amphibian distributions, and for birds and mammals in the Americas at a resolution of 30 arc-seconds (coverage) or as vector extents.

The CIESIN map client will contain pre-loaded WMC documents for the data layers available from CIESIN and selected AIP-2 service providers. The client has the flexibility to allow users to generate their own WMC documents based on any publicly-available WMS service and save the document for sharing with others. For WFS services, visualization of simple feature types and selection based on the OGC query specification are supported. Currently, the CIESIN map client supports WMS version 1.1.1; however, a test version that will support WMS 1.3 may be made available during this phase of the AIP if time permits.

CIESIN will develop a publicly available test WPS that accepts a simple point, line, or polygon geometry and buffer distance and returns the estimated population and land area within the given geometry. The service will accept requests from clients that can supply the requisite geometry or through service chaining (server-based submission). A measure of data quality (input geographic unit area) and other population characteristics, such as poverty indicators and age/sex ratios, are planned as additions to the WPS once the initial capacity has been established. In a client-based environment, the WPS will have the ability to return basic population characteristics of user-selected thematic layers, such as land use/land cover and protected areas.

These three sets of components and services will support the Disaster Response scenario in the following possible ways:

- 1) The population, hazard, and related data and coverages will provide important information complementary to real-time or near-real time satellite imagery, sensor data, and model-based forecasts, allowing users to assess the potential level of threat to population concentrations and other important resources.
- 2) The map client will allow users to access both CIESIN and distributed data sources and, utilizing the WMC specification, preserve the selection of data and data sources and the spatial extent of interest for later reuse or for sharing with other users. CIESIN will also be able to predefine contexts of interest for specific user scenarios and needs.
- 3) The WPS will enable users to define and submit a spatial query through interoperable clients and receive both a map and selected statistics resulting from the query. A user might, for example, be interested in estimating the total number of people or the number of poor people residing within a certain distance of the recent and projected path of a hurricane.

The three sets of components and services will support the Protected Areas scenario in the following possible ways:

- 1) The population and species grid coverages will enable protected area managers to examine the location of people and important species relative to current or potential protected areas. The national-level indicators will provide useful context

for monitoring and decision making regarding protection of existing areas and development of new areas.

- 2) The map client will allow users to access both CIESIN and distributed data sources and, utilizing the WMC specification, preserve the selection of data and data sources and the spatial extent of interest for later reuse or for sharing with other users. CIESIN will also be able to predefine contexts of interest for specific user scenarios and needs.
- 3) The WPS will enable users to define and submit a spatial query through the CIESIN map client or other interoperable clients and receive both a map and selected statistics resulting from the query. A user might, for example, be interested in estimating the total number of people or the number of poor people residing within a certain distance of a protected area, or the proportion of a given species' range that lies within protected territory.

The three services may also be of substantial utility to the air quality and energy scenarios, e.g., enabling users to assess populations affected by an air quality alert or within a specified distance of a proposed energy facility. Since the services rely on approved OGC specifications, interoperability with other data and services developed for other SBA scenarios should not be a problem. However, at this time CIESIN does not plan to specifically address the data or user interface needs of these scenarios.

The data service (WMS, WCS, WFS) components proposed will be registered in the GEOSS gateway and made available for use via the portal and other OGC-capable clients. The well-established discover-find-bind paradigm for these OGC services will enable multiple clients and users to utilize the services without the need for any interaction with CIESIN staff or special arrangements.

The CIESIN map client will allow the creation and sharing of WMC documents that can be used both in CIESIN's client and the portal. In addition, the map client support for WFS and filter query operations will permit analysis of feature data from distributed sources.

Relatively low levels of hits per hour are supported with the existing SEDAC infrastructure; however, the mapping servers and clients can support much higher levels of use without any additional resource deployment. The distributed architecture of SEDAC's systems allows the separation of map requests and web application services across machines. Although stress testing of map services has not been completed in the current environment, it is anticipated that services could handle several thousand requests per hour without much in the way of diminished performance.

Components provided to this effort are already part of SEDAC's operational infrastructure or will be incorporated into SEDAC's external beta test environment as part of the pilot. It is expected that the components will be maintained and improved as part of SEDAC's ongoing operations through at least July 2013 (the end date of CIESIN's current contract with NASA to operate SEDAC). In the event that SEDAC operations terminate at any time, CIESIN has a collaborative arrangement with the Columbia University Libraries to ensure the long-term preservation and accessibility of SEDAC data through the SEDAC Long-Term Archive (LTA). However, the levels of

service provided for these data are subject to determination by the SEDAC LTA Board, based on available resources. The data would be available for other entities to access and provide higher levels of service, if needed.

2.3 *Architecture and Interoperability Arrangement Development*

The GCI is a good baseline for data interoperability. For data modeling and analysis to take place in a distributed environment, additional service binding to custom (WPS, other) implementations is required. This is currently achieved through interoperability arrangements. As interoperability arrangements become more common, it may be possible to surmise a set of documentation requirements that will help automate service binding to service chains and model outputs. CIESIN's test WPS and any resulting interoperability arrangements for service chaining will contribute to the development of requirements that can help make service chaining a more routine part of the GEOSS GCI.

The idea of data bundles or thematic areas that contribute to the SBAs is an important point raised in the proposed architecture. The use of the WMC standard to share assembled interactive map documents is a sensible way to implement this idea. It would be useful if the WMC could be extended to include data and not just map services. In this way, a combination of WMS layers and selected features or layers from a WFS and rendered WCS layers could be combined seamlessly for visualization and use in a map client. Using the CIESIN map client, it should be possible to demonstrate data bundles that combine WFS and WMS layers that would be ideal inputs for an extended WMC standard.

As an Associate (University) Member of the Open Geospatial Consortium (OGC), CIESIN has contributed to various OGC standards initiatives, including past OWS testbeds, the earth observation and natural resources working group, and the data preservation working group. CIESIN has also participated in a number of standards-development efforts of the U.S. Federal Geographic Data Committee (FGDC), including *a)* testing and implementing a WFS that utilized the InterNational Committee for Information Technology Standards (INCITS) standard for administrative boundaries and *b)* development and modification of profiles for the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) and the International Standards Organization (ISO) metadata standard, TC211 standard 19115.

Based on the experience in this pilot, e.g., with interoperability arrangements for WPS, CIESIN will recommend additions/revisions to the architecture to better facilitate rapid and reliable integration of modeling and observational data and support for spatial query operations.

3 Description of Responding Organization

CIESIN is a research center within the Earth Institute at Columbia University, based at the Lamont campus in Palisades, New York. CIESIN has been designated the World Data Center for Human Interactions in the Environment by ICSU, a GEO Participating Organization. CIESIN is also active in other ICSU activities and bodies, such as the ICSU Committee on Data for Science and Technology (CODATA). The programmatic contact for this proposal, CIESIN Director Dr. Robert S. Chen, is Secretary-General of

CODATA and also task lead for GEO Task DA-06-01 on the GEOSS Data Sharing Principles.

As noted above, CIESIN is an Associate (University) Member of the Open Geospatial Consortium (OGC). CIESIN has participated in a number of OGC initiatives, including phase 1 of the Architecture Implementation Pilot (AIP-1), as part of the NASA contribution. CIESIN data and components were highlighted in several of the scenarios and portal demonstrations developed for AIP-1 (http://www.ogcnetwork.net/pub/ogcnetwork/GEOSS/AI_Pilot_Demo/index.html).

CIESIN is also a full member of the Global Spatial Data Infrastructure Association (GSDI), another GEO Participating Organization. The technical contact for this proposal, Gregory Yetman, is a member of the GSDI Board. CIESIN contributed to GEO-related activities at GSDI-9 and GSDI-10.

CIESIN operates the Socioeconomic Data and Applications Center (SEDAC), one of the distributed active archive centers in the Earth Observing System Data and Information System (EOSDIS) of NASA. SEDAC provides operational support for the archiving and dissemination of a wide range of data and information products and maintains a variety of online and offline services. SEDAC is committed to expanding accessibility to its data and services through open standards and is actively developing and implementing geospatial and other data services using OGC-compliant specifications. SEDAC's external advisory group, the SEDAC User Working Group, has endorsed SEDAC's active participation in GEOSS activities. SEDAC mission areas include Hazard Vulnerability Assessment (HVA), Population, Land Use and Emissions (PLUE), and Environmental Sustainability (ES). SEDAC data holdings are applicable to a wide range of SBAs, including the four selected for AIP-2. SEDAC is an active member of the Earth Science Information Partners (ESIP) Federation and the US FGDC Historical Data Working Group. CIESIN's contract to operate SEDAC for NASA has just been renewed for five years, which should allow for continued provision and updating of data and operation of services through at least July 2013.

CIESIN will continue to participate in AIP telecons and meetings and expects to contribute to the kickoff meeting in September 2008. SEDAC staff and system resources will be utilized to support modification and testing of the components described in this proposal. SEDAC plans to implement the components as part of its operational infrastructure. System resources available to support this effort include licenses to the ERDAS RedSpider software suite, ESRI ArcServer 9.3, ArcSDE 9.3, Oracle Enterprise, and BEA WebLogic. CIESIN also works with open source products such as GeoNetwork, Geoserver, Mapserver, and Fedora. CIESIN hardware resources include 8 application/mapping servers, 3 database servers, and over 25 terabytes of online storage.

Dr. Robert Chen, CIESIN Director, will serve as Programmatic/Business contact (bchen@ciesin.columbia.edu, +1 845 365-8952) and Gregory Yetman, CIESIN Senior Staff Associate, will serve as Technical contact (gyetman@ciesin.columbia.edu, +1 845 365-8982). In addition, CIESIN's Associate Director for Information Technology and SEDAC System Engineer, Sri Vinay (sri@ciesin.columbia.edu, +1 845 365-8967) will serve as alternate Technical contact.