Open Geospatial Consortium Inc.

Date: 2005-3-9

Reference number of this OGC™ document: OGC 05-030

Version: 1.1

Category: OGC™ White Paper

Author: Brandon Fisher, Editor: Carl Reed

Server Architecture Models for the National Spatial Data Infrastructures (NSDI)

OGC Document Number 05-030

DOCUMENT STATUS

This document is version 1.1 of the Server Architecture Models for the National Spatial Data Infrastructure (NSDI). Mr. Brandon E. Fisher of SAIC, under contract to the OGC, prepared this document.

Mr. Sam Bacharach, Dr. Carl Reed, and Mr. Mark Reichardt of the Open Geospatial Consortium, Inc (OGC) provided additional content and review of the document. Carl Reed provided a major edit for version 1.1 of this document prior to submission to the OGC membership for discussion and approval.

The document fulfills the requirements set forth in Minerals Management Task Order 32102, under cooperative agreement 1435-04-03-CA-70297.

POINTS OF CONTACT

OGC invites comments and recommended additions to this document. Please send all comments regarding this document to the following OGC staff:

Mark E. Reichardt
President, OGC
mreichardt@opengeospatial.org
+1 (301) 840-5443

Sam Bacharach
Executive Director, Outreach and Community Adoption
sbacharach@opengeospatial.org
+1 (703) 352-3938

1. Introduction/Purpose

This report is as an initial analysis of the current (late 2004) disparate server architectures of the National Spatial Data Infrastructure (NSDI) and the Geospatial One Stop (GOS) Portal. In order to yield the maximum value from current and future investments in GOS, it is necessary to understand and transition toward a sound and cost effective data provider server architecture model. To that end, it is also necessary to ensure the robustness, accuracy/currency (up-to-date-ness), and availability of national spatial data assets as well as access to those assets.

This document will attempt to address the relevant issues associated with different implementation architectures as communities move forward with development or enhancement of their systems architecture in support of local needs and broader NSDI objectives. This document applies to the operational interests of government, industry and academia to improve and simplify the management, discovery, access, sharing and application of geospatial information and services.

Specifically, the report addresses issues associated with the management of geospatial data via either centralized, distributed, or a combination of implementation architectures . The report draws from successful implementations of SDI's in the community, representing different producers and users of the data, and their respective needs regarding availability and accuracy – or fitness for use – of the data.

2. Organization of Report

The report begins by reviewing the current GOS and NSDI server architectures. Next, example operational reference architectures will be described, discussed, and compared. Based upon the discussions of the reference architectures, initial findings and conclusions are discussed. Finally, architecture guidelines and recommendations are provided for consideration by implementing organizations.

Hopefully, this document will be broadly referenced. Therefore, the document will be periodically enhanced as additional reference implementations are added and collective knowledge of effective data provider architectures grows. To facilitate initial communication and outreach, a template PowerPoint briefing for use by readers of this document is included.

The report was compiled with as much input as possible from interested parties as possible given the time constraints for this effort. Also, a great deal of information has been gathered and consolidated from other sources. The "Credits and References" section at the end of the report notes each source of information included in the report.

3. Current GOS Portal Architecture[1]

The ISO/RM-ODP[2] modeling approach defines five architectural viewpoints for specifying interoperability requirements for open, distributed processing. In a generic way, the model identifies the top priorities for architectural specifications and provides a minimal set of requirements—plus an object model—to ensure system integrity. Five standard *viewpoints* are defined; the viewpoints address different aspects of the system and enable the 'separation of concerns' (See Table 1).

Table 1 - RM-ODP viewpoints

Viewpoint Name	Definition of RM-ODP Viewpoint
Enterprise	Focuses on the purpose, scope and policies for that system.
Information	Focuses on the semantics of information and information processing.
Computational	Captures component and interface details without regard to distribution
Engineering	Focuses on the mechanisms and functions required to support distributed interaction between objects in the system.
Technology	Focuses on the choice of technology.

For the purposes of this document, we will emphasize the Enterprise Viewpoint. An Enterprise Viewpoint provides a high-level system concept with supporting use cases to help describe the architecture. The system concept illustrates the operational setting, major system components, and major interfaces. The Use Cases provide descriptions of the behavior of the system from the point of view of Users. For The GOS Portal, the System Concept is in this section.

[1] From "GOS-Portal Implementation Architecture," 2003-05-04

[2] The **Reference Model for Open Distributed Processing** (RM-ODP) is an international standard for architecting open, distributed processing systems. It provides an overall conceptual framework for building distributed systems in an incremental manner. The RM-ODP standards have been widely adopted: they constitute the conceptual basis for the ISO 19100 series of geographic information standards (normative references in ISO/DIS 19119), and they also have been employed in the OMG object management architecture.

An intergovernmental project managed by the Department of the Interior in support of the President's Initiative for E-government, Geospatial One Stop builds upon its partnership with the Federal Geographic Data Committee (FGDC) to improve the ability of the public and government to use geospatial information to support the business of government and facilitate decision-making.

The vision of the GOS Portal is to enable users to discover, view and obtain desired geospatial data for a particular part of the country, without needing to know the details of how the data are stored and maintained by independent organizations. Figure 3.1 below depicts users from all sectors of government and society being able to access The GOS Portal. The GOS Portal, in turn, is able to access information and services from a variety of Providers distributed across the network. As providers increasingly support standardized protocols for accessing their content and services, other Portals can be linked with the GOS Portal. Such portal linkages will provide additional functionality or more specialized views of the information. Indeed, The GOS Portal itself could run at multiple sites in order to provide redundancy and avoid bottlenecks at a single location.

Thus, Geospatial One-Stop is an enterprise information portal (EIP). An EIP is a concept for a Web site that serves as a single gateway to a company's information and knowledge base for employees and for its customers, business partners, and the general public as well. A portal implementation does not store or maintain the data and its associated services. Rather a portal provides a gateway to distributed content and services accessible at many locations nationwide and maintained by the agency or organization that is responsible for specific content and services. For example, the US Bureau of Transportation might maintain a service providing interstate highway data, a state agency might serve data about the highways under its jurisdiction, and a city agency might serve urban street data. A user should be able to view a map including roads from all of these jurisdictions simultaneously, letting the Portal automatically contact the necessary services, access the required content or service, and process or fuse the data as required. Furthermore, the User should be able to view detailed documentation, or metadata, about the data and its source(s) if desired.

The GOS Portal builds upon the Clearinghouse Network used in the US National Spatial Data Infrastructure (NSDI). That network catalogs data that have been documented according to the metadata standard published by the US Federal Geographic Data

Committee (FGDC). Users can search the Clearinghouse or the individual catalogs and be referred to specific geospatial resources. The Portal enhances existing Clearinghouse capabilities by providing direct access to a subset of the data in the catalog—specifically and potentially to those data services that use specific types of standardized access methods.

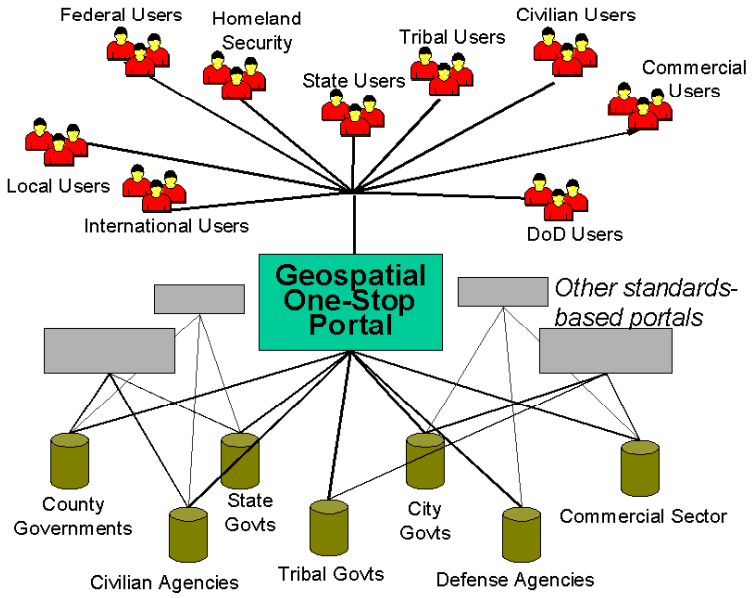


Figure 3.1: The Geospatial One Stop Portal

A major goal of the Geospatial One-Stop is to leverage open standards that have been or will be defined collaboratively by a variety of stakeholders, are freely published, and are able to be implemented by any vendor or organization. Three broad classes of standards and specifications are relevant to the Portal and the services it accesses (see the "Contract for Interoperable Geospatial Portal Components" web page at http://www.fgdc.gov/geoportal/ for a comprehensive discussion of applicable standards):

1. **Framework Standards:** There are seven geospatial data themes that are considered to be of fundamental importance to many applications. Known in the U.S. as Framework Data[3], these themes are: Elevation, Orthoimagery, Hydrography, Transportation, Government Units (administrative boundaries), Cadastral (property boundaries), and Geodetic Control. Framework Data content standards are now under development by another component of the Geospatial One-Stop initiative (in particular, see the related GOS Transportation Pilot described below). Data sources

[3] see www.fgdc.gov/framework/framework.html

wishing to be classified as Framework Data shall, at minimum, be able to exchange data in a manner that complies with these emerging standards. The Portal shall be able to access both Framework Data sources and other, non-Framework data.

2. **Service Specifications and standards:** Access to data and maps are provided according to open consensus standards and specifications. For example, the OpenGIS® Web Map Service, Web Feature Service, and Web Coverage Service specifications define standard interfaces and methods for requesting spatial data via the web for a given geographic area of interest. Some organizations will offer only a Map service, while others will also offer Feature or Coverage services to support data analysis, maintenance and update across the web. A summary of OGC web service standards is provided as Annex B of this document.

3. **Metadata Standard:** Metadata shall be published that provides detailed information about data and services. In particular, data will be documented according the FGDC Content Standard for Digital Geospatial Metadata[4] (CSDGM). Access to the metadata is through services such as the OGC Catalog Service. Furthermore, the GOS Portal may access or maintain other registries that support discovery, access and use of web services applications, schemas, styles, symbols etc., necessary in applying the data for a particular use.

3. Taxonomy of Geospatial Server Architectures

There is no “one size fits all” geospatial enterprise or server architecture that is appropriate for all organizations. Organizations will develop their architectures and systems to best fit the data quality, security, accessibility and related factors associated with their business environment and processes. For instance, the architecture for a bear census maintained by a rural county in central Pennsylvania will be different than the architecture used by a private national weather service providing weather information to the FAA. What are some of the obvious, and maybe not-so-obvious, reasons these architectures would be different?

1. **Amount of data:** Weather forecasting is hugely complex, and requires the maximum amount of data about current conditions. For this reason, a national weather servicing organization would need many terabytes of storage for its data, high-speed network access, and perhaps access to GRID applications for weather modeling. In

[4] see http://www.fgdc.gov/metadata/contstan.html

contrast, a county-level organization that is collecting and maintaining small, countywide datasets would likely suffice with only a few gigabytes of disk space, low bandwidth and no complex modeling requirements.

The amount of data collected, maintained, and accessed by an organization impacts the architecture not only as a factor for disk space, but also for processing power, computer memory (RAM), required database software, and communications bandwidth, user interface design and so forth.

3. **Location of data sources:** For many applications, the geospatial data required to provide a service or solve a problem may be distributed. For instance, to get a realistic "picture" of the weather situation in a metropolitan area, data is accessed from US Government satellite resources, local Weather Forecast Offices (NEXRAD Radar Scenes), Atmospheric and Surface Observation System (ASOS) observations from local airports, and other sensors managed by other private and public concerns. Alternatively, for other applications, such as emergency response, it may be desired to maintain certain data elements centrally to insure access to that data when an emergency event occurs.

2. **Criticality of data:** If the FAA, or an associated air traffic controller loses access to real time weather data, it can result in a costly and/or dangerous situation. For this reason, a data center providing real-time weather data would be considered "mission-critical." Mission critical data and applications are expected to be available 100% of the time, 24 hours a day, 7 days a week. These applications must employ reliable redundancy and fail-over measures ensuring consistent availability. Conversely, there are many geospatial resources that do not have such time criticality yet need to be accessed on demand by a range of applications and users.

3. **Currency and Accuracy of data:** The currency of data is also of potential importance. The weather data in Consideration 2 is critical only if it is timely – yesterday's weather information is of no value to management of today's flights.

4. **Security and Privacy needs:** Increasingly, there are authentication and security concerns related to the access and use of geospatial data. Today, it's not only data that provides the location of military and intelligence installations that are considered in need of security measures. Information providing the location of water treatment

plants will now contain some level of security in terms of who can access the data, the level of detail provided, and so forth.

Data security measures are not only driven by military and anti-terrorism efforts. The National Biological Information Infrastructure (NBII) regards the known locations of certain threatened and/or endangered species as data not necessarily for public consumption and therefore requires some level of security.

Yet another driver affecting security is the protection of private or proprietary data. Whether the infrastructure must protect proprietary data licensed from another organization or company, or house its own proprietary data, it must be protected from theft and unauthorized distribution.

Business Processes: Information technology infrastructure decisions related to geospatial resources should be made within the context of an organization's business process environment. The business process requirements should guide the selection of hardware and software resources that are dedicated to geospatial data systems. Within the business process context, all of the portal requirements should be captured before any decisions are made regarding technology procurement. If one follows the RM-ODP process, at a minimum the enterprise, information, and component viewpoints need to be understood and documented before implementation decisions are made.

5. **Data rights:** Many geospatial data centers or portals will provide access to a collection of data with many different owners. It is not uncommon for data centers to purchase licenses to utilize geospatial data from another provider. In this case, there may be restrictions on how this data can be shared. Complicating the matter is the situation where a data center contains both freely available data and data that are proprietary and forbidden for redistribution.

   Also, private organizations that collect, compile, and sell/distribute geospatial data will store, and manage access to, and/or distribute that data differently than a public data store might.

6. **Organizational objectives/policy:** Loosely related to data rights, the objectives and policies of an organization will have a significant effect on the architecture. For instance, organizations like NBII and GeoStor that have a policy objective to provide

free access to much of their public data, should take great steps to implement and employ existing data storage and transmission standards – Thus providing the least resistive means to access and utilize their geospatial data.

Private organizations and companies on the other hand, are not as obligated to employ these standards because data is not generally shared broadly. However, metadata and service standards benefit businesses that wish to improve the ability to mobilize new technology solutions with minimal integration and customization. This is particularly true as more applications that implement standards appear in the marketplace, and more public and private organizations continue to employ and rely upon standards. The bottom line is, however, that private companies collecting and generating geospatial data as a business proposition have more freedom with their architectures, whereas public sites expected to widely provide and distribute public data should utilize software and hardware that adhere to appropriate and applicable geospatial standards.

7. **Budget.** The amount of financial resources will inevitably factor in to the implemented architecture. Given a particular organizational objective/policy, along with the characteristics of the data (amount of data, amount of data to be distributed/transferred, etc.), there will be a range of acceptable architecture parameters (security level, necessary hard/software, bandwidth requirements, etc.) for the target system. Because these parameters come with their respective costs, there will be a range of an expected cost to build and maintain such a system.

   The budget available to the organization will determine how aggressive the organization can be when designing and implementing their geospatial data system architecture. It is important that an organization establish a budget for their geospatial data center sufficient enough to meet the requirements driven by the organization's geospatial objectives, business processes, and security/technical issues.

   The use of standards based technology is one way that organizations have addressed budget constraints. A benefit of employing standards is the ability to rapidly – and easily – expand system capabilities. Systems can be implemented in iterations, and as a result of applying standards, the integration time and costs associated with system expansions is minimized.

All these factors, and more, will drive the architecture of each geospatial data server infrastructure. These factors will drive organizational decisions from as general as the overall architecture – i.e. centralized, decentralized, hybrid, etc. – to as detailed as the amount of memory in the application server.

Oftentimes these factors are in 'conflict' with one another in that maximizing one necessarily minimizes another. The user must weigh the comparative benefits of each factor and places where tradeoffs must be made.

4. Overview of Reference Architectures

Following are four reference architectures along with information regarding their respective organizations business and policy drivers.

4.1. Reference Architecture No. 1: Centralized Spatial Data-Center (Warehouse)

4.1.1 Overview

The example of a centralized spatial data center that which will be used as a first reference architecture is the USDA Geospatial Data Warehouse (GDW).

The single most important driver of the architecture design for the GDW is the mission critical nature of the data. It is absolutely necessary that the data of the GDW be available continuously. It is for this reason that the USDA has chosen a centralized architecture for the GDW. The data of the GDW is obtained from various sources, through various means.

Because there is so much data stored in the GDW, specialized data marts break out subsets of that data. This significantly minimizes the time, and increases the likelihood of finding and accessing specific sets of data.

GDW Architecture

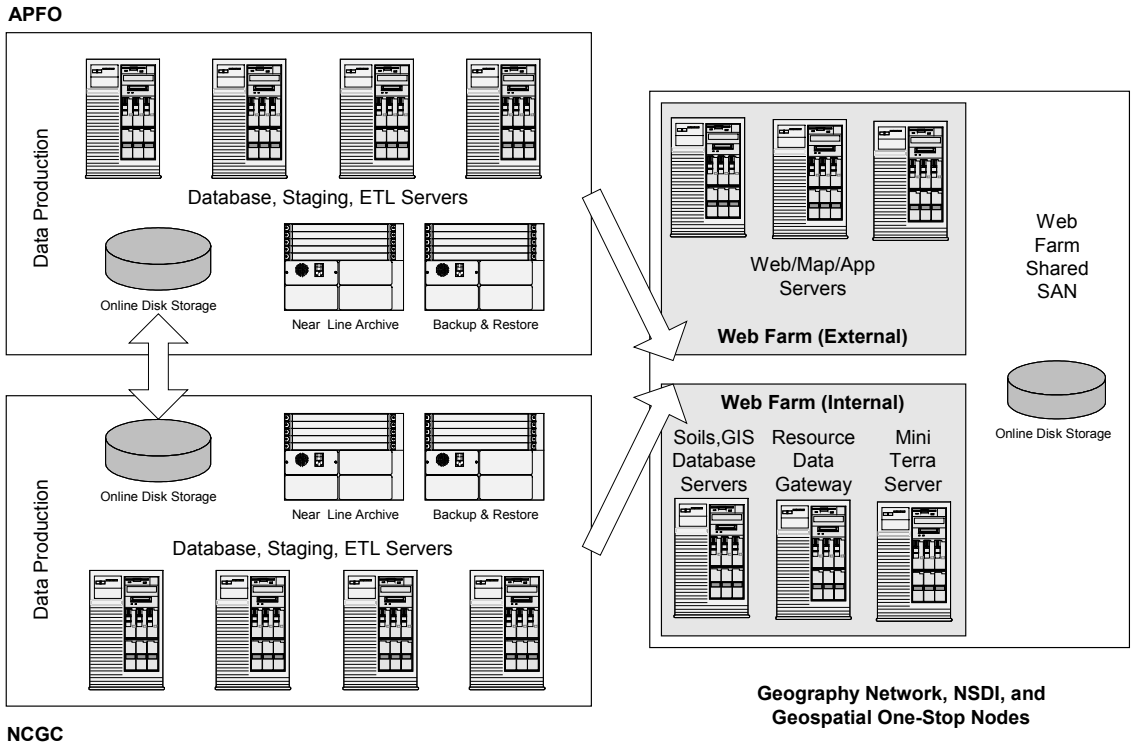


Figure 4.1: Computational architecture of the GDW.

Figure 4.1 provides an overall computational architecture diagram of the GDW.

DATA MANAGEMENT

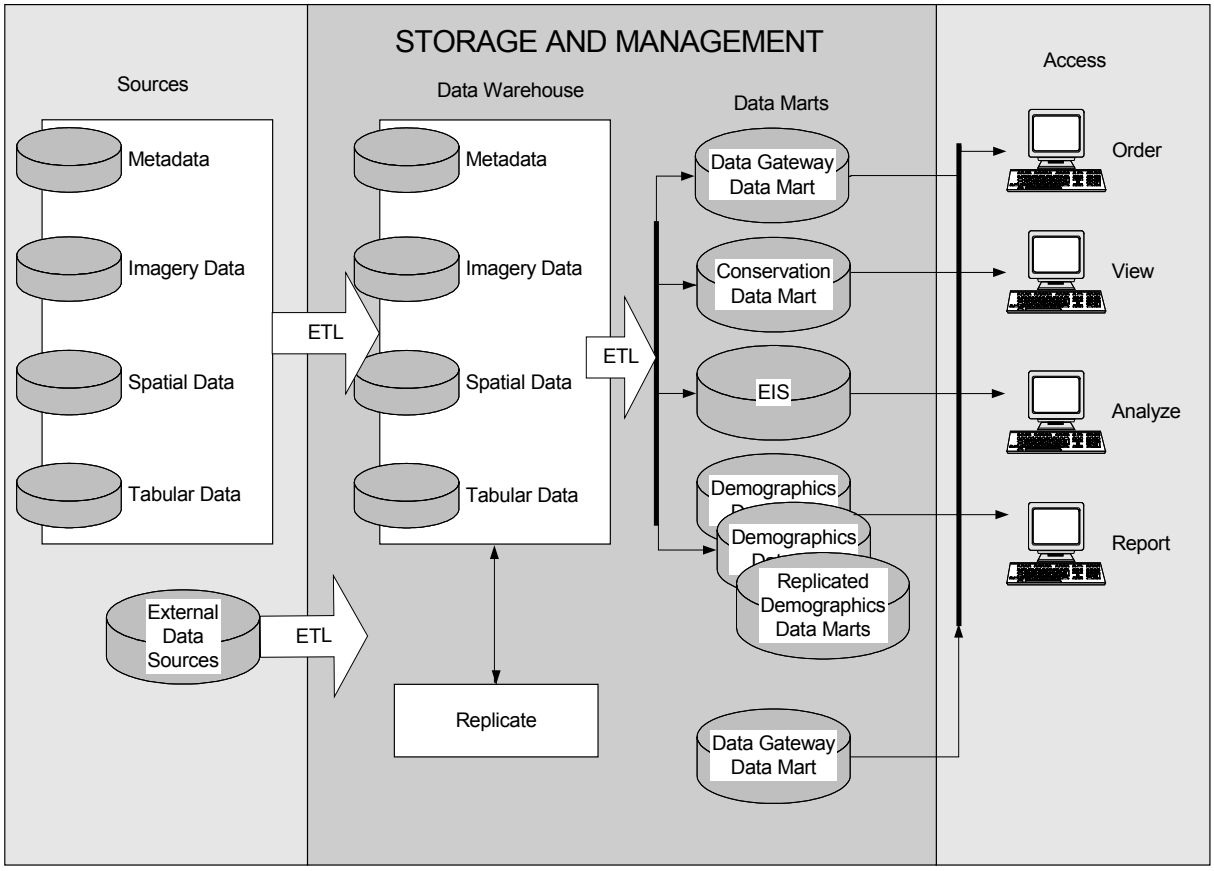


Figure 4.2: Data management components of the GDW.

Although the data marts are considered to be managed by the individual data centers, they are essentially separate entities operate independently.

Figure 4.3 provides a high level data flow diagram showing how raw flows into the digital data production services process all the way through to the end process where it is delivered to the consumer in its final form. The data production process receives the raw data and performs various tasks, which include parsing, assembling, enhancing, and formatting the data. Once the data has made it through this process, it is put in the data warehouse. The data warehouse essentially ingests the processed data and stores it. Based on a series of business rules specific to each data theme, the data is replicated to its sister data center, and transformed for use in the data marts. As it ages, the data is also archived in a near-line state.

The data marts receive the transformed data from the data warehouses through some combination of pushing, queuing, and pulling the data at a yet-to-be-determined interval. The

data marts are the active repositories for finished and current data that make that data available to the consumers. In some cases, the term ‘current data’ means only the most recent version of that particular data theme; in others cases, it means all versions of the data that need to be accessible to the consumer(s). Various Web Services make the data available to the consumer subject to the request made through a given Web application. The Web application then formats the data accordingly and delivers it to the consumer as specified.

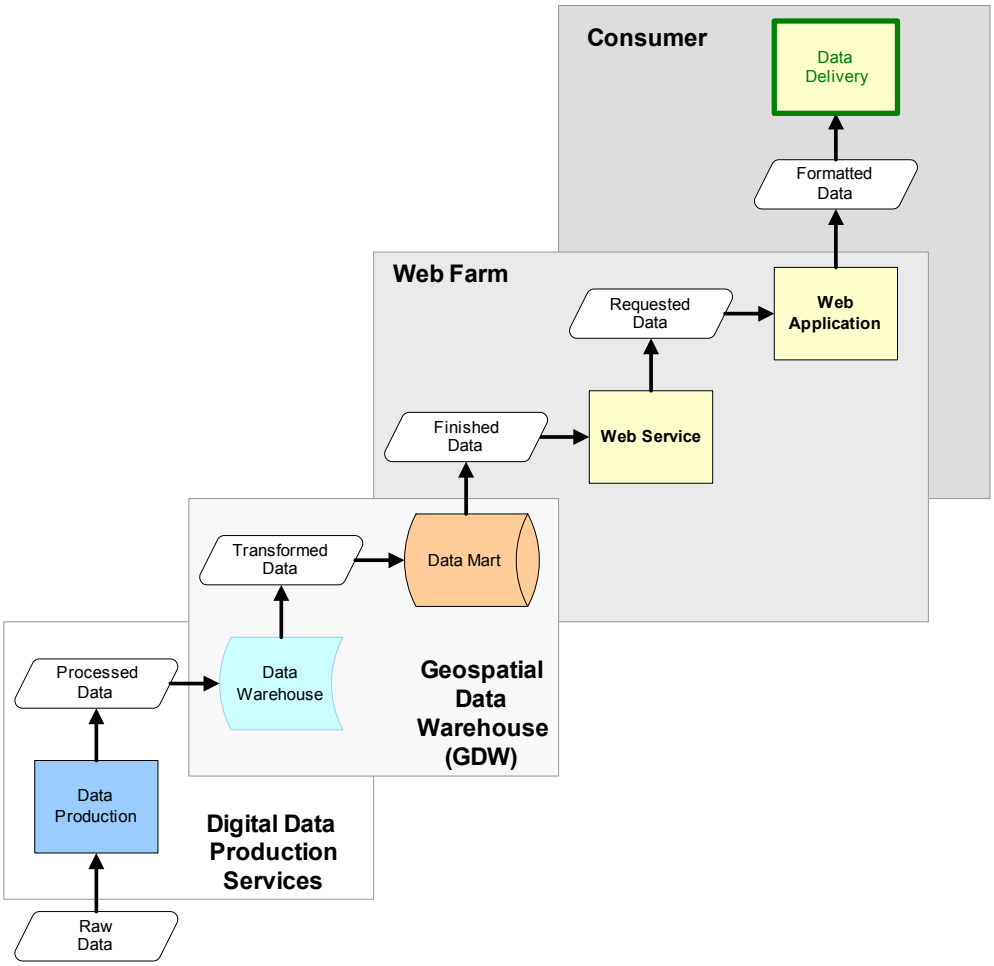


Figure 4.3: GDW data flow diagram.

Consumers are the users of the data that reside in the data marts. They are made up of USDA employees at the data centers and the Field Service Centers, as well as contractors, 3rd party stakeholders, and public users. Depending on the service(s) used, the data may be accessed in several different ways: it may be packaged and delivered as an FTP download or on a CD, it may consist of a live feed through a Web application, it may be initiated by a

data center employee as a batch request, or it may be directly accessed by internal production and/or development employees.

4.1.2 Commentary on Reference Architecture No. 1

For mission critical applications where availability is priority, a centralized implementation architecture is often viewed as the best solution. When you own or manage the entire system from end-to-end, you can ensure access, availability, and fitness-for-use of the spatial data assets.

When so much data is stored in a single location, mining that data in real-time can be an expensive and time-consuming operation. Data marts bring together the various types of data to produce more specialized sets of data – or data access – thus increasing performance of data access.

There is however a price tag associated with guaranteed access and availability. Also, managing this much data can be expensive as well. An often-overlooked expense is the cost of telecommunications. A relatively large, centralized system, as that of the GDW, requires a large capacity for communications bandwidth. The cost of sufficient bandwidth at the two large facilities in Fort Worth, TX and Salt Lake City, UT costs the USDA $240,000 a year, at *each* location. Data management software along with online[5] and near-online[6] storage comes at a cost of $1M to $1.5M per location – with maintenance costs around $1M per year. Table 4.1 itemizes FY02 and FY03 costs for the GDW.

INFRASTRUCTURE ITEM	Total Costs	FY 02	FY 03
Storage (Online)	$ 588,000.00	$ 294,000.00	$ 294,000.00
Storage (Near-Online)	$ 1,100,000.00	$ 550,000.00	$ 550,000.00
Servers (data, web, applic.)	$ 3,350,000.00	$2,000,000.00	$1,350,000.00
Other Hw/Sw	$ 330,000.00	$ 250,000.00	$ 80,000.00
Data Mgt Software	$ 1,900,000.00	$1,000,000.00	$ 900,000.00
ETL/OLAP Software	$ 1,100,000.00	$ 550,000.00	$ 550,000.00
Telecommunications $/yr	$ 510,000.00	$ 230,000.00	$ 510,000.00
Telecommunication Security	$ 210,000.00	$ 210,000.00	$ -
Physical Security	$ -	$ -	
Support Services - $/yr	$ 3,720,000.00	$1,000,000.00	$3,720,000.00
Implementation of Servers, communications, and replication	$ 240,000.00	$ 240,000.00	$ -
TOTAL	$13,048,000.00	$6,324,000.00	$7,954,000.00

[5] Online storage cost is based on APFO estimate of $21,000 per TB

[6] Near Online cost is based on NCGC needing 40 TB and APFO needing 20 TB to come to a total of 40 TB each.

Table 4.1: GDW costs

4.2. Reference Architecture No. 2: Distributed Spatial Data-Centers

4.2.1 Overview

For many business reasons, it is often most appropriate to employ an overall distributed architecture to deliver sets of geospatial data and services. For instance, the Pacific Forestry Centre (PFC), of the Canadian Forest Service, found it to be a high priority to allow local agencies to remain autonomous and to maintain full ownership and control of their spatial data. Yet, the PFC was faced with the problem of generating nationally and internationally mandated reports on forestry assets – on a relatively frequent basis – that required geospatial data and information from these autonomous agencies.

The PFC needed data and information from many sources, but needed to access it from a single location. Another component in the PFC's system architecture approach is the fact that, at this time, the system stores no "mission critical" information, or information that could cause a significant impact – financial, health, or other, if the information is temporarily unavailable.

Based on these factors, the most logical approach for the PFC was a distributed architecture. The following is an architecture diagram at the main project office of the PFC 's geospatial data server system.

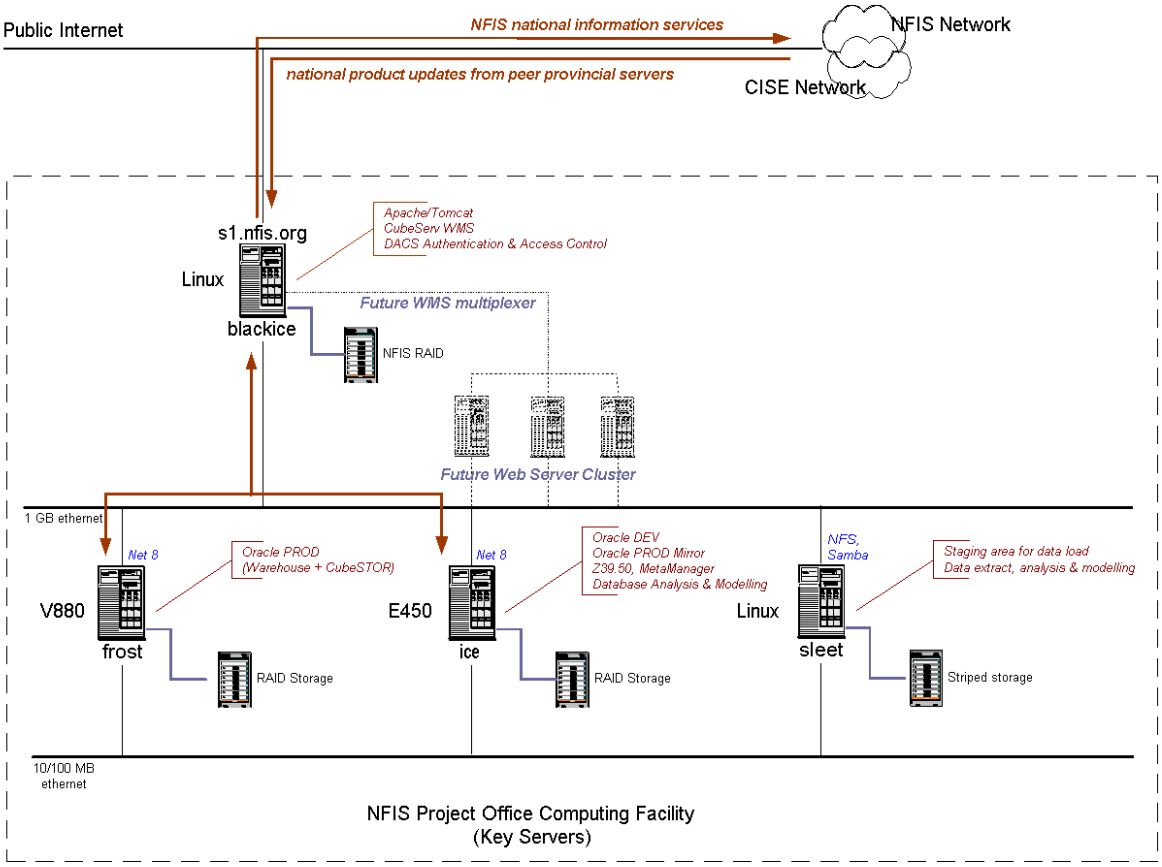


Figure 4.3: NFIS National Project Office Computing Facility.

There are no geospatial data collected, generated, or stored at the main project office. The primary storage facilities depicted in figure 4.3 are used to store information and reports generated *from* data obtained from the node agencies.

And because server implementations vary at each of the partnering agencies, each node employs an OGC Web Map Service (WMS) connector such as those provided by ESRI, Intergraph, University of Minnesota Map Server, or Cubewerx. As necessary, other additional standards such as the OGC Catalog interface standard are used to make data available to the main project office.. This standards based approach allows “plug and play” connectivity for the main project office to the partner agency nodes, regardless of the agency’s server architecture.

Figure 4.4 illustrates an example of the architecture at one of the NFIS nodes -- Newfoundland & Labrador.

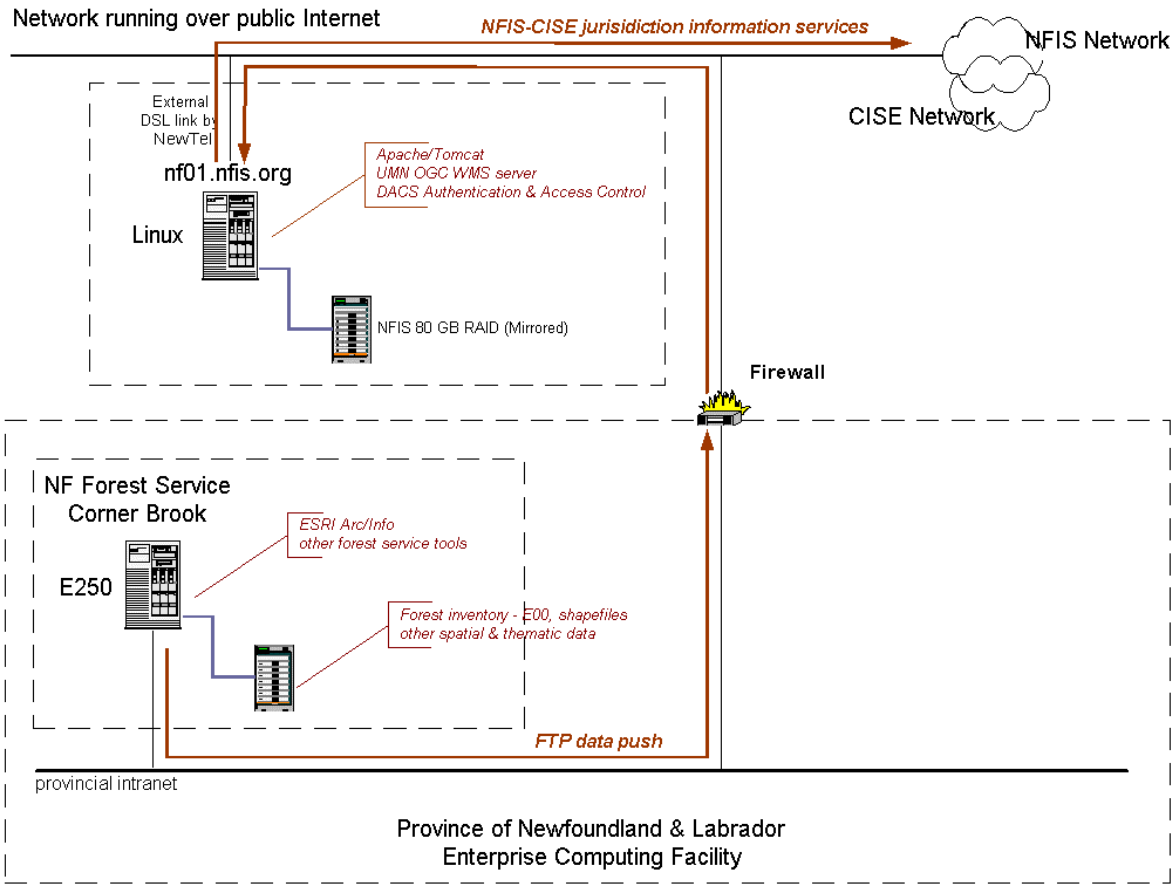


Figure 4.4: Newfoundland & Labrador NFIS Server Configuration

Not all agency nodes have the same architecture/configuration as Newfoundland & Labrador. Some agency systems are very small and may not even have separate database and web servers. The Newfoundland & Labrador example is a very practical and pragmatic solution as it takes two key -- yet relatively inexpensive -- steps towards creating a secure and reliable node. One step is the use of a separate database and web server, and the second is the use of a firewall between the two servers. Because in general, the cost of hardware today is relatively cheap compared to the software, adding an additional server to act as the database server will not significantly add to architecture costs. The additional processing capability gained, however, will greatly increase the performance and reliability of the system. Also, hardware firewalls such as a network router have become so inexpensive that they've become a commodity.

The standards-based approach employed by the NFIS and necessitated by this independence of the agency nodes such as Newfoundland & Labrador also affords them the

ability to directly link up with a portal like the GOS. Each partnering agency can be individually registered with the GOS portal, and makes at least a subset of data available to the public finding the agency's data via the GOS portal.

Geospatial data acquired via the GOS portal is accessed directly from NFIS partnering agencies via the same open standards-based interfaces used by the PFC main project office to obtain the most current geospatial data.

4.2.2 Commentary on Reference Architecture No. 2

The decentralized model is the perfect fit for an organization facing the set of drivers such as those facing the PFC:

1. Policy Objective: Reduce effort/cost of collecting data from existing, disparate agencies.
2. Disparate agency geospatial capabilities already in existence, and already collecting and storing geospatial data.
3. Minimal impact on the agencies a high priority.
4. Meet agency requirements to maintain local ownership and control of their respective datasets.
5. Agency datasets determined to be not mission critical to the policy objective.
6. Because a reduction in overall costs was the objective driving the project, implementing and maintaining the system at a minimal cost was a primary goal.

In order to permit interoperability between all agencies and subsystems, the PFC has made geospatial interoperability standards central to their systems. The use of the OGC WMS enabled Web Mapping Services is critical and fundamental for interaction with partners. The PFC considers it important not to isolate itself by using proprietary or non-standard applications and protocols.

With the adoption of software that is true to existing standards, the PFC is able to make a decentralized model work for them. They were able to build the hub of the system for less than $250,000 (US) of materials and caused minimal impact to their partners.

A major challenge facing the PFC is the creation of seamless and consistent geospatial data from data maintained in various data models, many of which when integrated do not provide

consistent feature descriptions and attributes. This hampers the process of data processing for analysis and reporting.

4.3 Reference Architecture No. 3: Combination Spatial Data-Centers

4.3.1 Overview

The National Biological Information Infrastructure (NBII) is in a similar situation as the PFC – Dispersed agency nodes collecting their own geospatial data. However, none of their data is considered mission-critical with regards to cost or human well-being.

One major difference between NBII and the PFC is that NBII was established with a single program objective to create an infrastructure for sharing biological data over the internet. As it turns out, some NBII related nodes are too small to build and/or maintain their own geospatial data servers, or are funded directly by the program office. Therefore NBII has established architecture at their hub in Denver, CO for the purpose of maintaining the geospatial data collected by these nodes. Furthermore, NBII does license a small amount of data from other organizations and stores that data at the hub location in Denver.

This architecture can be classified as a hybrid centralized and decentralized geospatial data architecture.

Beyond the Denver, CO. hub location, there are seven existing nodes collecting geospatial NBII data, with five more in various stages of development. The data, the services, and the models – the resources that would be reported to GOS – are all stored and updated at the node level. As in the case of some of the nodes, data is not even always centrally stored at the node. NBII characterizes their nodes as a “virtual infrastructure of partners.” Of the data that can be accessed via an NBII node, some of it is collected by, and stored at the node, and some of it is owned by partner agencies.

Each node covers either a specific geographical area and the biological issues within that area, or a specific biological issue over an entire geography. For instance, the Bird Conservation node maintains geospatial information related to bird conservation over all of North America.

In figure 4.5, the flow chart illustrates the NBII’s data discovery process.

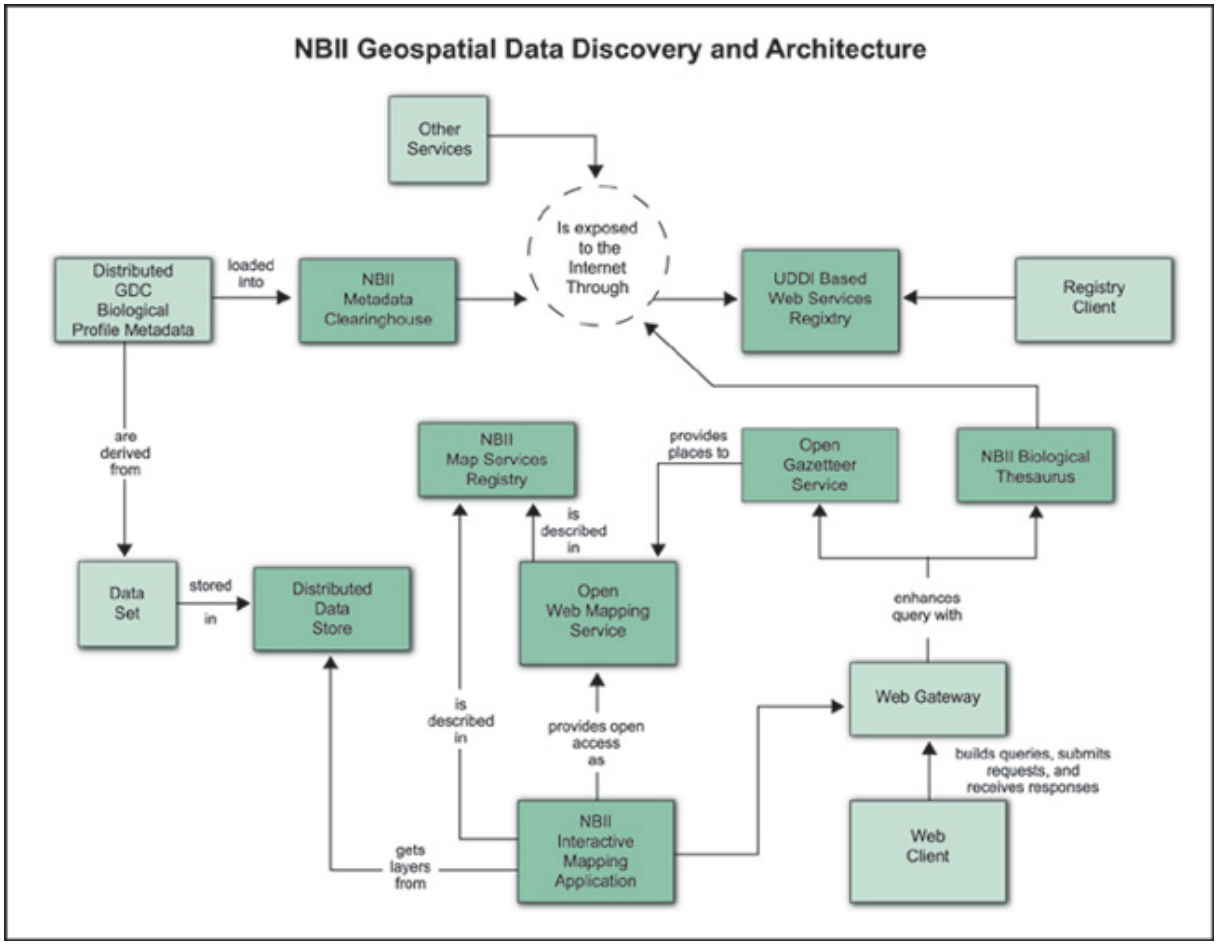


Figure 4.5: Data discovery architecture of the NBII.

NBII does not send down a mandate to its nodes as to what software to use in their geospatial systems. The only requirement is that standard applications are used to transfer data between the node and the main data center. In NBII's case, this includes ESRI and Minnesota Map Server for WMS servers, FGDC for data/metadata, UDDI for web service registries, and Dublin Core for cataloging of resources.

The main data center primarily uses Microsoft software – Windows and SQL Server for its database – but throughout the entire system, a wide variety of software is employed including Windows, Linux, Solaris, Oracle and Oracle Spatial, SQL Server, and MySQL.

Below is an "enterprise" level diagram of the NBII geospatial server architecture.

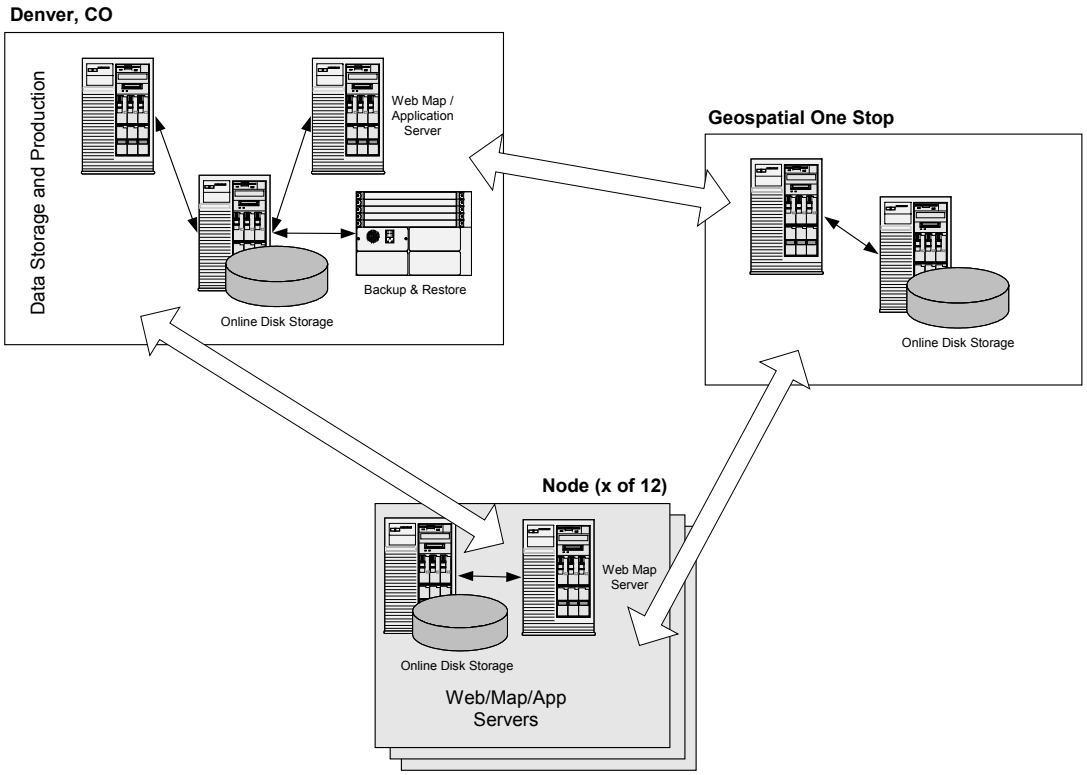


Figure 4.6: Computational architecture of the NBII.

NBII nodes are not required to have any plans to maintain high availability. They are expected to attempt to adhere to "best practices," which in this case would include some type of back-up and recovery plan, but none is required.

Because there is no "mission critical" data at any of the nodes, server reliability is not considered an overly important issue. NBII encourages all nodes to strive for maximum availability of their servers and data, but there are no actions expected to be taken automatically if a server goes down.

Although the data housed here is also considered not to be mission critical, the data center in Denver is held to a somewhat different standard. Being a USGS site, the central data center has a security plan, with a relatively rigid backup and recovery process.

4.3.2 Commentary on Reference Architecture No. 3

At NBII, sharing information is the program objective. NBII was built specifically to share biological information on the internet. The chosen architecture fits NBII well for the following reasons:

1. Not all nodes are owned/funded directly by the NBII program office. This requires NBII to work with the organizations at the nodes that implement their architectures in a way that is most beneficial for them. In most cases, this translates to the node organizations storing and maintaining their own data. As a result, the NBII's servers simply establish a linked relationship to the nodes.
2. Data not considered "mission critical." When an organization's data is, or can be, considered not to be mission critical, there is more freedom with regards to the overall system architecture. A hybrid-decentralized architecture was determined to be best for NBII, and the attributes of the data allowed for that option.
3. Some very small NBII related organizations. These organizations do not receive the necessary funding to "stand up" and maintain their own geospatial data servers. In these cases, NBII takes it upon itself to obtain the data collected by these organizations, and includes them in the data stores of the main data center.

No precise cost data was available, but an estimate puts the cost at "somewhere over $1M to maintain the central Denver node." This estimate includes a staff of about six or seven Full Time Equivalents (FTEs), as-needed training, as well as software and hardware updates.

The cost of each node varies greatly, as their systems and needs vary. Some nodes have very little support (< 0.1 FTE with no admin and no DBA), and some nodes nearly reach the staff level of that at the central node in Denver.

Due to sometimes minimal funding for their NBII related geospatial systems, NBII nodes are often very pragmatic with their resource levels. As a result, NBII often finds itself assisting node organizations with their geospatial data systems.

4.4 Reference Architecture No. 4 Centralized Local - Regional Government

4.4.1 Overview

MetroGIS was formed as a "regional forum to promote and facilitate widespread sharing of data"[7] in a seven county area of Minneapolis-St. Paul, Minnesota. With goals to reduce overall costs and to support cross jurisdictional decision-making, the Metropolitan Council, an agency established as a regional planning and operational agency for the twin city areas provides staffing and financial support for MetroGIS operations. It should be noted that MetroGIS is not an incorporated organization, and cannot own data or manage funds. But it does support consensus decision-making, involves elected officials in its processes, and coordinates best practices for stakeholder voluntary compliance.

The Metropolitan Council, on behalf of MetroGIS manages and serves up geospatial data for use by the Metropolitan Council and a variety of stakeholders in the twin cities area. In addition, MetroGIS operates a metadata clearinghouse node – part of the NSDI Clearinghouse, and publicly available web map services.

The Metropolitan Council / MetroGIS Architecture can best be defined as a centralized architecture with replication to support secondary usage. Metropolitan Council supports internal geospatial data discovery, archive and distribution for staff of the Metropolitan Council, as well as public access to geospatial metadata and holdings via an external MetroGIS server. The server architecture developed by Metropolitan Council is shown in figure 4.7 below.

For external (public) support, MetroGIS utilizes ESRI ArcIMS, FME, ISite, and DataFinder /Café, a web-based application developed for metadata publication and search, and to support web map visualization via the OGC Web Map Service and vendor proprietary capabilities. MetroGIS is presently investigating replacing or augmentation their operational web services capability to support other open web services standards including OpenGIS Web Feature Service and Web Coverage Service.

From a staffing perspective, GIS Web Server operations are supported on a part time basis by a GIS Web Developer, GIS Database Administrator, and Information Systems department staff.

[7] See http://www.metrogis.org/

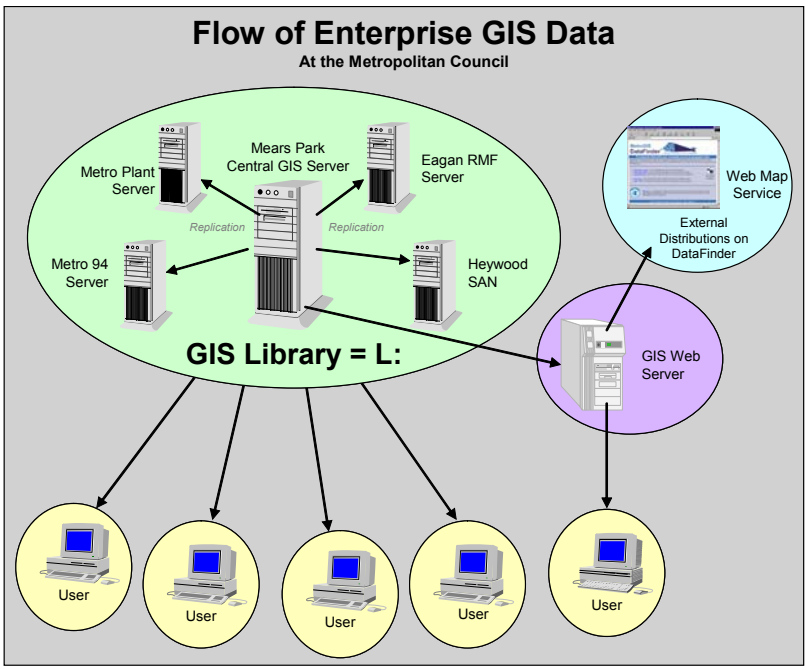


Figure 4.7: MetroGIS Server Architecture

Although MetroGIS offers access to their server on a continuous basis, 24/7 operations are not guaranteed should an operational interruption occur outside of normal business hours. To date, no user requirements for continuous or “mission critical” operations have explicitly defined by the user community.

4.4.2 Commentary on Reference Architecture #4

The MetroGIS architecture evolved to support as a centralized approach to support both Council and external operational needs, which at the time did not emphasize a fully web services approach. MetroGIS serves as a regional repository for stakeholder organizations with and without their own geospatial capabilities.

Data availability through MetroGIS is not considered “mission critical”, which may limit continuous operations in the event of a system failure. No fail over support is currently provided.

Costs associated with the development and implementation of the GIS Web Server for external use approximates $110K, with a maintenance cost of $10.5K annually. MetroGIS receives a level of Information Systems operations and maintenance support free of charge from the Metropolitan Council.

Staff levels required to support sustained operations approximates 0.4 FTE, divided between the following skills: GIS Web Designer, GIS Database Administrator, and IS Department Staff.

5. Conclusions

When designing a geospatial data architecture that fits a particular need, there are many factors – some more important than others – that will determine the optimal architecture for both performance and budget. Aside from budget, business requirements and the level of criticality of the data that is being managed appear to be the most significant determinants of geospatial system architectures.

For systems containing mission critical data, security is typically an important concern. Whenever you have mission critical data, you'll have individuals or organizations interested in accessing that data. In some cases, intruders will either want to steal, or corrupt, or disrupt this mission critical data.

Systems that require constant availability and highly accurate/reliable data, and as a result high levels of security, are costly. Depending on the amount of data and the level of security, and the storage facilities, these systems can run in the tens of millions of dollars to build, and millions of dollars a year to maintain.

Although there are many standards employed for formatting, cataloging, and transmitting data, there appears to be a contingent of software and hardware providers that supply products to meet these needs. This includes:

- Computer operating systems: Found a relatively even distribution between Windows, Linux, and Sun
- Database software – Primarily Oracle (and Oracle Spatial) along with MS SQL Server. There are some instances of MySQL storing geospatial data.
- Application Hardware – Primarily distributed between Sun and PC based servers with Cisco networking equipment.
- Application Software: Much of this software employs the implementation of "standards" that are becoming more common today. Almost all geospatial software vendors utilize OGC, FGDC, and other geospatial standards where applicable.

Also, most vendors that provided application software before the standards were developed are implementing the standards-based technology in new releases.[8]

Typically, systems that are not bound by the needs of highly critical data can put together their system on a smaller budget. For instance, the NFIS/PFC put three $3000 dell servers behind a Cisco load balancer, resulting in excellent performance at a minimal cost. They have also employed Apache Tomcat and Minnesota Map Server – both are examples of free, open-source web servers (MN Map server a "servlet") capable of handling typical internet traffic.

There is one necessary, potentially significant cost that is unavoidable for any reliable geospatial information system – telecommunications. At a minimum, reliable bandwidth is going to cost a data center a few hundred, to several thousand dollars a month. And as more and more bandwidth becomes necessary, the cost can reach into the tens of thousands of dollars a month. The USDA's mission critical GDW absorbs a cost of nearly $250,000 a year at each of its main data centers.

For systems such as that of the NBII, a hybrid centralized/decentralized – which will be a more common architecture – bandwidth requirements can be much less. And because of this, telecommunication costs can be significantly less expensive, yet still reliable. Table 5.1 provides approximate costs associated with some referenced architectures.

[8] The OpenGIS website has a list of vendors that supply OGC compliant software applications. Go to the "Registered Products" page under the "Resources" tab at www.opengis.org

	GDW	GeoStor	NBII	NFIS	MetroGIS
Build Costs	**$9,000,000.00**	**$1,000,000.00**		**$480,000.00**	**$110,000.00**
Hardware	$3,000,000.00			$100,000.00	$10,000.00
Software	$5,000,000.00			$200,000.00	$94,000.00
Staff	$1,000,000.00			$280,000.00	
Maintenance (yearly)	**$4,000,000.00**	**$200,000.00**	**$1,000,000.00**		
SW/HW upgrades			$50,000.00		$10,300.00
Staff	$3,500,000.00	$200,000.00	$900,000.00	$280,000.00	$25,000.00
Telecom	$500,000.00		$50,000.00		
Notes	"Staff" includes all support services. Costs are total of two data centers. Software costs include Online and Near Online storage, data mgt, and misc. software.	Amounts do not include value of free equipment and sponsorships received	Very rough estimates. Growing system, difficult to accurately estimate build costs.	Staff costs based on 3-5 FTEs at approx $80000/year	Cost and staff information are for external (NSDI) operations only, and do not include internal Metropolitan Council Support

Table 5.1: Build and Maintenance costs associated with referenced architectures.

As stated in the introduction to this report, and supported throughout the text, there appears to be no one-size-fits-all solution for geospatial data systems and portals. It is believed that an organization can prioritize and analyze the factors listed in section 3 and illustrated in section 4, and using the basis for analysis provided with the factors, determine a pragmatic solution that fits the type of geospatial system approach (i.e. mission critical, secure, distributed, centralized etc.) that best fits organizational needs.

In each of these examples, the organization(s) described have leveraged the resources of other participating organizations to some extent, and have made their geospatial resources available for discovery and reuse. Having done this, they have been successful. For example in the case of the USDA GDW none of the participating organizations alone would have had the resources to enable them to implement their architecture. They were able to document and articulate a viable business case to U.S. Office of Management and Budget through this cross-agency partnership. These approaches to partnership are challenging for a number of reasons, but the overriding challenge is trust. Organizations need to be able to establish trusted relationships such as service level agreements to ensure that these data will be available for their applications, especially when there is a mission critical operational

requirement. In some cases, organizations are unwilling or unable to explore the establishment of these partnerships. Unfortunately without these, agencies will develop their own stand-alone systems that replicate each other to some extent. When viable partnerships are developed, the participating organizations and the consumers of geospatial information benefit

Appendix A: Discussion on Information Interoperability

Despite efforts by the Federal Geographic Data Committee (FGDC) and others to encourage broad use of data content standards for improved geospatial data sharing, the reality is that communities across the nation usually collect and maintain their data (transportation, hydrography, etc) using data models established by them to meet local needs. (A data model lists and defines the types of entities represented in the data, including their attributes and relationships.) While unique data models serve local needs within a jurisdiction, the use of different data models among neighboring jurisdictions hinders sharing of data and cross-boundary collaboration, and the use of different data models among overlapping jurisdictions results in redundant data collection and management. OGC's geospatial software interface and encoding standards (OpenGIS Specifications) support web-based discovery, access and integration of data, but this "technical interoperability" does not guarantee the "semantic interoperability" necessary for applications like emergency management and homeland security.

Geospatial One Stop, the National Map, and other programs that facilitate the sharing of data across the web will need to continue to promote the use of data content standards, but there will always be differing data needs about the same geographic area. The cartographer, the highway maintenance manager, the FedEx dispatcher and others will never be able to do their jobs with a committee designed data model. Fortunately, a degree of data sharing is possible without perfect adherence to those standards. OGC's XML-based Geography Markup Language (GML) provides a way to accomplish partial translation between data models, so that collaborating organizations can make the best possible use of each other's data, despite differences in their data models. A common theme such as transportation can be defined by the GOS Framework standards and each user then maps his or her data to the standard. Software using off the shelf XML technologies is then able to translate many individual data models “as needed” through the standard to the model of the requesting user. This method will enable the integration of data from many sources, legacy and new, into a semantically consistent data set for use in decision support, analysis and visualization. This is referred to in OGC as "information interoperability."

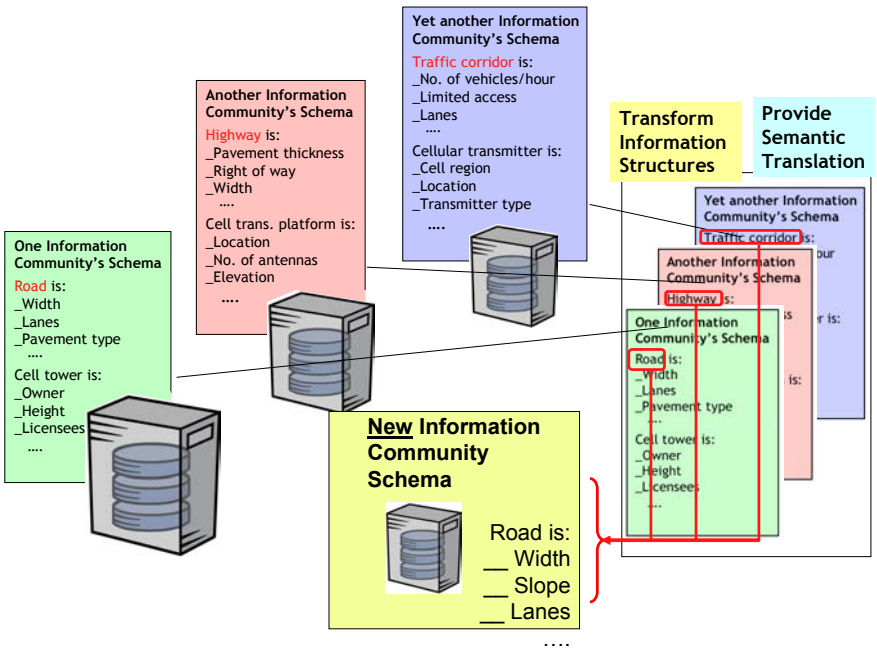
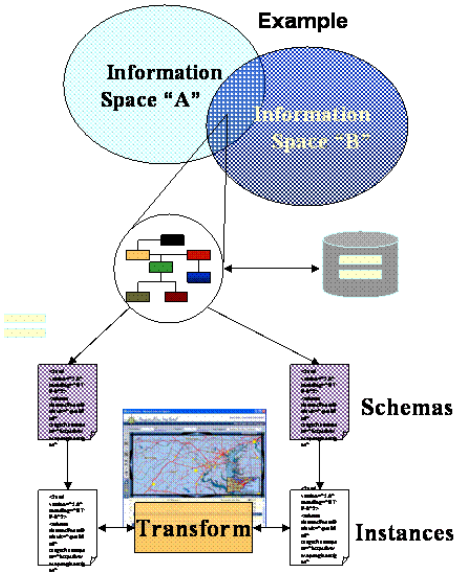


Figure A.1 – The Information Interoperability Challenge

One-to-many mapping of data models is made possible by XML tools. XML tools prototyped in OGC's Geospatial One Stop Transportation Pilot and Critical Infrastructure Protection pilot projects create a GML "application schema" from a UML representation of a local data model (see http://www.opengis.org/initiatives/?iid=8). After establishing a mapping between similar elements in two dissimilar GML-encoded data models, it is possible to translate – "on the fly" – between them, so that county or state data can be translated to a regional or national model, and vice versa. Given that this capability is possible, the next decision to make is whether to pay the price for "on the fly" translation or to institute some kind of an update cycle (e.g., daily, monthly, yearly) that translates local data to update a remote store of the translated version. This decision will relate directly to the provider architecture in place.



Through this process, data thus becomes "as useful as possible" between data sharing partners who use different data models. Of course, there will be cases where certain elements of one model do not map to the other model. But the XML tools make these inconsistencies plain in all their details, so that it is easy for data managers to focus on correcting only the critical data model elements that can't be translated. As an alternative to forcing the broad adoption of a single data model, the key benefit of this approach is that it *minimizes cost and effort for organizations that wish to share data*. It makes it easier for people at the local level to accommodate regional and national standards in an affordable and practical way, and it makes it easier for people at the regional national level to work with local data that hasn't been converted in all its details to the national standard.

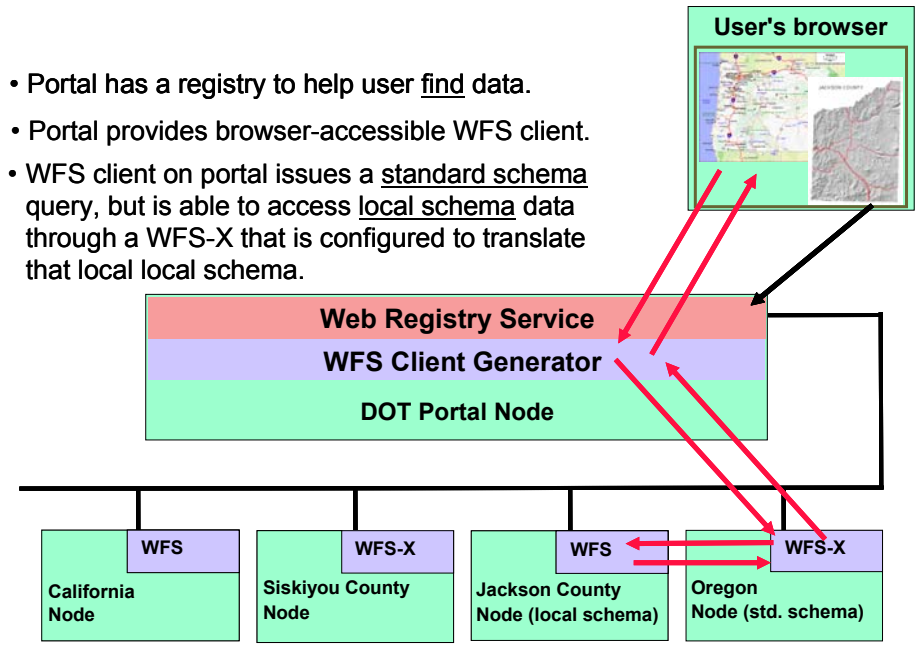


Figure A.2 – Geospatial One Stop Transportation Pilot Prototype

Appendix B: OpenGIS Specifications Relevant to NSDI Server Architectures

The following is a brief summary of the relevant OpenGIS Specifications applicable to Data Server architectures and standards-based portals. By including OpenGIS Specifications in GIS and related programs, data sharing with other organizations and jurisdictions becomes much faster and easier. Organizations also maximize their ability to rapidly adapt to new technologies regardless of vendor, and adaptation requires less integration support. Please note that this summary was prepared in April 2004. Consult www.opengeospatial.org for the latest OpenGIS Specifications, which are freely available for download (click on the "Documents" tab, then "OpenGIS® Specifications").

OpenGIS Catalog Services Implementation Specification v2.0

The OpenGIS Catalog Services Specification defines common interfaces to discover, browse, and query metadata about data, services, and other potential resources. Profiles of the OGC Catalog 2.0 specification provide implementation interface definitions for specific technology platforms, such as HTTP, ISO 19115/19119 and OASIS ebRIM.

OpenGIS Coordinate Transformation Services Implementation Specification 1.0

A key requirement for overlaying views of geodata ("maps") from diverse sources is the ability to perform *coordinate transformation* in such a way that all spatial data use the same *spatial reference system*. This specification provides a standard way for software to specify and access coordinate transformation services for use on specified spatial data.

OpenGIS® Filter Encoding Implementation Specification 1.0

Bundled with the Web Feature Service (WFS) specification is the Filter Encoding Specification, which defines a standard encoding for query predicates using XML. Using XML encoding, a query operation can be defined that retrieves objects that lie in a particular region, or that deletes object instances that lie in a particular region and have a particular value for some specified non-spatial property.

OpenGIS Geography Markup Language Implementation Specification (GML 3.1)

GML 3.1 defines a data encoding in XML – an XML "namespace" – for geographic data and its attributes. GML provides a means of encoding spatial information for both data transport and data storage, especially in a Web context. It is extensible, supporting a wide variety of

spatial tasks, from portrayal to analysis. It separates content from presentation (graphic or otherwise), and permits easy integration of spatial and non-spatial data.

OpenGIS Grid Coverages Implementation Specification 1.0

In the OGC context, a "coverage" is a function or any set of entities that exhaustively cover a plane. A grid coverage is a specific case of coverage in which a set of grid values covers the surface. Examples of a grid coverage are satellite images, digital elevation models, and digital orthophotos. This specification specifies interfaces that provide for requesting and viewing a grid coverage and performing certain kinds of analysis such as histogram calculation, image covariance and other statistical measurements.

OpenGIS Styled Layer Descriptor Implementation Specification (SLD v1.0)

A basic tenet of OpenGIS Specifications (and of XML) is the separation of content from presentation. Such separation enables a client to instruct that a particular “view” be created of a feature collection. The SLD is an encoding specification for associating presentation rules with properties of features.

OpenGIS Simple Features Specifications

OpenGIS Simple Features for OLE/COM (1.1), Simple Features for CORBA (1.0) and Simple Features for SQL (1.1) specify interfaces for OpenGIS Simple Features that are tailored for three different distributed computing platforms other than the World Wide Web. The OpenGIS Simple Feature Specification application programming interfaces (APIs) provide for publishing, storage, access, and simple operations on Simple Features (i.e., features described using vector data elements such as points, lines and polygons.)

OpenGIS Web Coverage Service Implementation Specification (WCS 1.0)

The Web Coverage Service specification applies the Grid Coverages specification to the Web. It extends the Web Map Server (WMS) interface to allow access to geospatial "coverages" that represent values or properties of geographic locations, rather than WMS generated maps (pictures). Future versions of the WCS specification are expected. They will, for example, expand supported coverage types beyond grid coverages only.

OpenGIS Web Feature Service Implementation Specification (WFS v1.0)

In contrast to the OpenGIS Web Map Service Implementation Specification, which delivers a picture, a WFS implementation in a client supports the dynamic access and exploitation of feature (vector) data and associated attributes. It describes data manipulation operations on

OpenGIS Simple Features (e.g., points, lines, and polygons) so that servers and clients can communicate at the feature level.

OpenGIS® Web Map Context Documents Implementation Specification (WMC 1.0)

The Web Map Context Documents Specification is a companion Specification to the OpenGIS Web Map Service 1.1.1 Implementation Specification. It describes how context information can be defined in XML and saved so that web maps created by users can be reconstructed and augmented by the user or other users in the future.

OpenGIS Web Map Service Interface Implementation Specification (WMS 1.3)

The OpenGIS Web Map Service Specification (WMS) provides uniform access by Web clients to maps rendered by map servers on the Internet. Thus, WMS is a service interface specification that enables the dynamic construction of a map as a picture, as a series of graphical elements, or as a packaged set of geographic feature data. It answers basic queries about the content of the map. And it can inform other programs about the maps it can produce and which of those can be queried further.

Annex D: Acknowledgements

The authors wish to acknowledge the support of the following individuals and organizations in the compilation and review of this document:

- ***MetroGIS***

  Randall Johnson
  MetroGIS Policy / Staff Coordinator
  230 East 5th Street
  St Paul, MN 55105
  (651) 602-1638
  randy.johnson@metc.state.mn.us

  Mark Kotz
  (651) 602-1644
  mark.kotz@metc.state.mn.us

  Alison Slaats
  (651) 602-1363
  alison.slaats@metc.state.mn.us

- ***National Forest Information System (NFIS)***

  Dr. Robin Quenet
  (250) 363-0127
  rquenet@pfc.cfs.nrcan.gc.ca

- ***National Biologic Information Infrastructure (NBII)***

  Donna Roy
  USGS Center for Biological Informatics
  12201 Sunrise Valley Drive, MS 302
  Reston, VA 20192
  Droy.usgs.gov
  http://www.nbii.gov

- ***University of Arkansas, Center for Advance Spatial Technologies (CAST)***

  Dr. Fred Limp
  Director, CAST
  12 Ozark Hall
  Fayetteville, AK 72701
  (479) 575-7909
  fred@cast.uark.edu

- ***National States Geographic Information Council / State of Oregon***

  Cy Smith
  Statewide GIS Coordinator
  Information Resources Management Division

Oregon Department of Administrative Services
955 Center Street NE, Room 470
Salem, Oregon 97301
(503) 378-6066
cy.smith@state.or.us