NGA Homeland Security
Enterprise GIS Support Project

Bay Area Regional GIS Council
Homeland Security Pilot Summary

Draft National Geospatial-Intelligence Agency
March 2005

Prepared for:

Brad Lucas, Contracting Officer's Representative
National Geospatial-Intelligence Agency
4600 Sangamore Road
Bethesda, Maryland 20816-5003
Tel.: 301-287-6572

Prepared by:

ESRI
380 New York Street
Redlands, California 92373-8100

"

Document History

Revision History

Date	Description	Person Responsible
3/16/05	Initial Draft	Mike Ruth
3/24/05	Updates pertaining to the Pilot Summary	Sergio Rodriguez
3/28/05	Major Updates to the structure of the document	Taylor Sims
3/29/05	Updates to the text of the document	Taylor Sims
3/30/05	Updates based on comments by Peter Bottenberg	Taylor Sims
3/30/05	Updates based on comments by David Kehrlein	Taylor Sims
3/31/05	Last Updates to the Draft	Taylor Sims
4/4/05	Sensitive Items removal	Michael Young / Taylor Sims

Table of Contents

Appendixes

1.0 About This Document

1.1 Purpose

This document is intended as a resource to be used in the creation of data repositories in support of homeland security (HLS), homeland defense (HLD), activities at a regional or state level. It will be used as a source document for a more comprehensive effort by a consortium of federal agencies to create a template that can be used nationwide. It represents an accounting of the efforts to date setting up the Bay Area Regional-Homeland Security Data Server (BAR-HSDS).

This document will also serve as guidance for the upcoming collaboration between the National Geospatial-Intelligence Agency (NGA), the State of Colorado, and the United States Geological Survey (USGS). The intent is to follow the processes developed for the BAR-HSDS as described in this document and incorporate new findings based on the implementation experience in Colorado.

1.2 Intended Audience of This Document

This document is intended primarily for use by the NGA and USGS Homeland Security Enterprise GIS Support Project management team, their partner agencies, and the ESRI Homeland Security Enterprise GIS Support Project team.

1.3 Document Overview

The document is organized into five sections and 14 appendixes as summarized below.

- **Section 1—**About This Document: Provides an overview of the Bay Area Regional GIS Council Homeland Security Pilot Summary document.
- **Section 2—**Project Homeland Overview: Presents an overview of the stakeholders and the architectural framework utilized.
- **Section 3—***The National Map*/Palanterra™ Synchronization Task: Provides a detailed breakdown of the task's operations, systems, data, and security.
- **Section 4—**The BAR-GC Pilot Summary: Provides a detailed description of the project outcomes.

- **Section 5—**Lessons Learned Summary: Provides a description of the findings from the activities performed.

- **Appendix A—**Geodatabase Design for Bay Area Regional GIS Council Pilot Project: Provides a detailed description of the data model used for the BAR-GC effort.

- **Appendix B—**Homeland Security Data Model Diagrams: Detailed diagrams of the data model used for the BAR-GC effort.

- **Appendix C—**Bay Area Regional GIS Council Pilot Project Data Migration Report: Provides a thorough description of the process used to gather, convert, map, and import data into the BAR-GC Data Model.

- **Appendix D—**Bay Area Regional HSDS ETL and Data Replication Framework: Provides an overview of the Extract, Transform and Load (ETL) process used for the BAR-GC pilot.

- **Appendix E—**Bay Area Regional GIS Council Memorandum of Understanding: A copy of the Memorandum of Understanding (MOU) used to share data for this pilot project.

- **Appendix F—**Bay Area Regional GIS Council Risk Assessment/Management Plan: A copy of the risk assessment management plan used on this BAR-GC pilot.

- **Appendix G—**Site Suitability Matrix: Provides an outline of questions to ask in determining a host site.

- **Appendix H—**Bay Area Regional GIS Council HSDS Host Site Data Center Survey: Provides an outline of questions to ask in determining a host site.

- **Appendix I—**Bay Area Regional HSDS Security Framework: Provides a framework intended to guide the implementation of access and security for the BAR-HSDS.

- **Appendix J—**The National Map/Palanterra™ Synchronization System Architecture Design: Provides a detailed description of the entire system architecture.

- **Appendix K—**Bay Area Regional GIS Council Viewer Release Notes: Provides a detailed description of the BAR-GC viewer and its implementation.

- **Appendix L—**Glossary: Defines important terms used in the BAR-GC HLS Pilot Summary document.

- **Appendix M—**Acronyms and Definitions: Defines various acronyms used in the BAR-GC HLS Pilot Summary document.

2.0 Project Homeland Overview

Project Homeland's primary goal is to improve and enhance the ability of the National Geospatial-Intelligence Agency to support homeland security and homeland defense customers with timely, relevant, tactical, operational, and strategic data products and services. The efforts conducted as part of this project will provide a proof of concept for the larger implementation of COTS-based solutions to support HLS and HLD missions.

Four primary objectives will be addressed as part of the NGA HLS enterprise GIS effort.

1. Utilize a rapid integration and implementation process to improve and enhance the existing NGA HLS data model.

2. Automate the sharing of spatial databases between the NGA Palanterra™ Systems and the USGS National Map system. The synchronization efforts will focus on (1) synchronization between state and local geospatial data sources and *The National Map* of the USGS, enabled by the HLS data model and (2) synchronization of *The National Map* with NGA Palanterra™ systems.

3. Improve and enhance the operational capabilities of Palanterra™, NGA's Web-based Geospatial-Intelligence Decision Support System (analytical, visualization, and dissemination systems). These efforts will provide a proof of concept for the larger integration and implementation of the COTS-based solutions to support the homeland security and homeland defense missions.

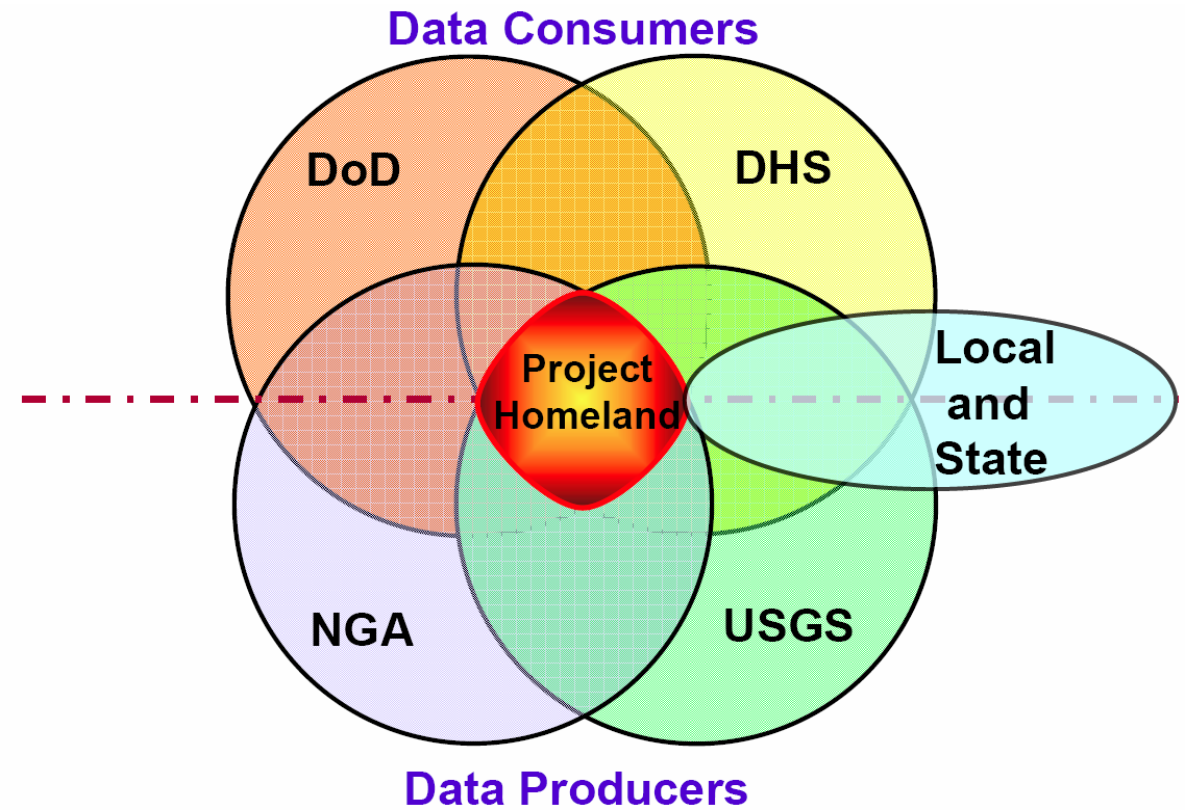
4. Provide training and technical resources in support of objectives 1, 2, and 3.

This document focuses specifically on the BAR-GC pilot.

2.1 Stakeholders

Project Homeland's primary stakeholders are the NGA, Department of Homeland Security (DHS), Department of Defense (DoD), and USGS; their roles as data consumers or data producers are illustrated in figure 2-1.

Figure 2-1
Project Homeland Primary Stakeholders



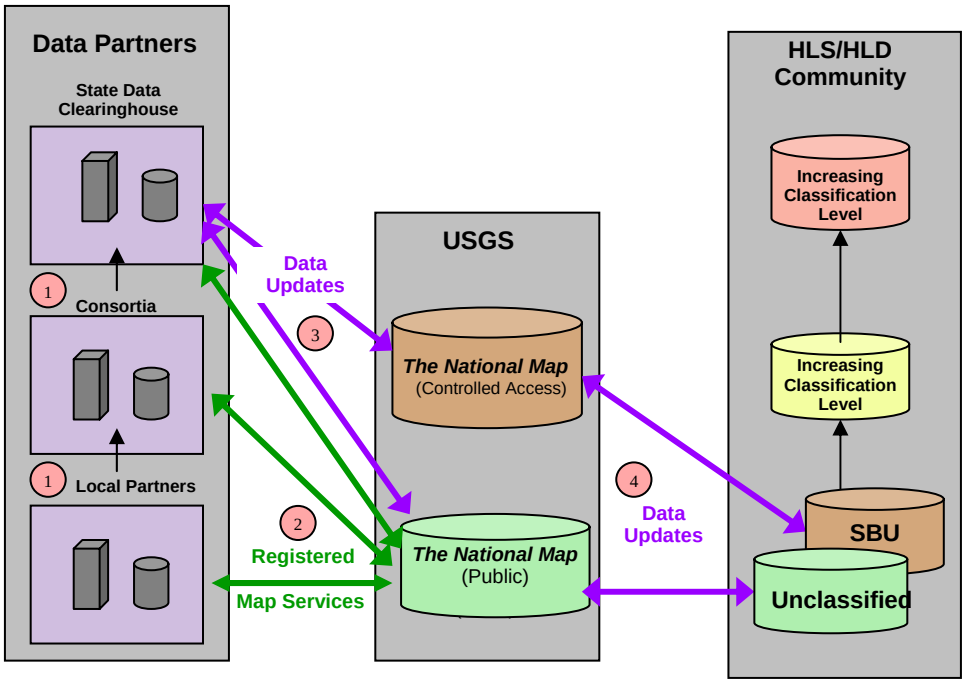
Each organization utilizes various applications to help distribute its data. For this project, the organizations will utilize systems as follows:

- NGA—Palanterra™: Views, collects, and distributes data
- USGS—The National Map: Centralized baseline of national GIS data
- DHS—COMFORCE: Primarily consumes data from NGA and USGS
- DoD—NorthCOM: Primarily consumes data from NGA and USGS
- Local and State—BAR-GC: Provides detailed data and consumes USGS data from *The National Map.*

2.2 System Concept

The foundation of Project Homeland is the concept of providing end-to-end, Web-enabled, and services-oriented geospatial intelligence between multiple agencies. Figure 2-2 depicts the overall system concept in terms of synthesis and collaboration activities.

Figure 2-2
Synthesis and Collaboration Activities



1. Spatial data flows from local partners to regional GIS consortia to state-level partners.

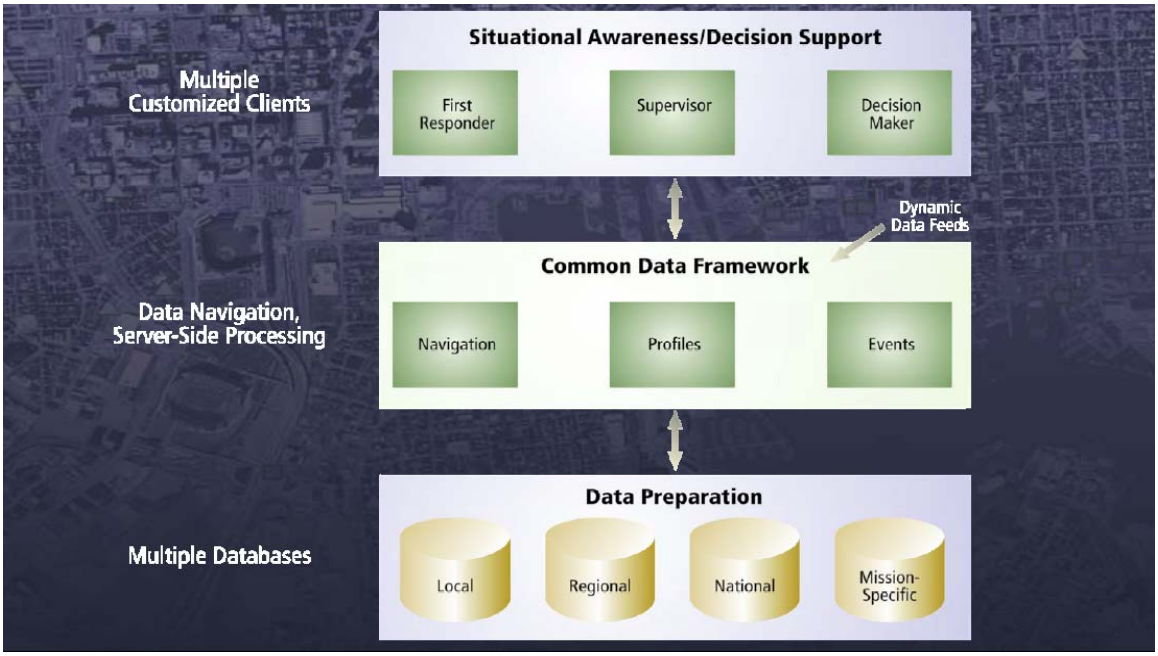
2. The local partners, consortia, and state agencies will create Web mapping services and register them with *The National Map* catalog.

3. The partner data will be loaded onto the controlled access and publicly accessible data stores, based upon the partner-identified access privileges. After the initial data load, updates made by the USGS or the data partners will be passed to each other on an ongoing basis.

4. The data from USGS will be loaded onto the Homeland Security (HLS) and Homeland Defense (HLD) data stores, based upon the partner-identified access privileges. After the initial data load, updates made by the USGS or the HLS/HLD community will be passed to each other on an ongoing basis.

At a conceptual level there are three tiers in the system, which are illustrated in Figure 2-3 below.

Figure 2-3
Concept View of Information Flow



The top tier (Situational Awareness/Decision Support) is a set of customized interactive geospatial presentations that will be designed to present the information that a particular type of user needs for either situational awareness or decision support. These users will be operators, analysts, managers, or decision makers.

The middle tier (Common Data Framework) organizes multiple data sources into a common data framework that feeds the upper (presentation) tier. The goal of this framework is to enable multiple coherent views of the current situation. Since different users require different information, there will not typically be just one common operating picture (COP) but multiple pictures, each filtered and presented to meet the information needs of its intended audience. The framework will also provide capabilities to navigate through the system's multiple datasets to quickly discover relevant data. Real-time data feeds, such as events or tracking data, will be processed at the common data framework tier and the results made available to the presentation tier. This common data framework is a core component of providing a federated services-oriented architecture (SOA).

The lower tier (Data Preparation) consists of the many data sources for the system. These include *The National Map* from USGS, NGA's own data holdings, data from other agencies that has been supplied to NGA to support this mission, and mission-specific data held by partner agencies.

3.0 *The National Map*/Palanterra™ Synchronization Task

The National Map/Palanterra™ Synchronization task's primary goal is to improve and enhance the availability of geospatial data to support NGA in its support of HLS and HLD customers with timely, relevant, tactical, operational, and strategic data products and services. This objective will be accomplished through several proof-of-concept activities that test and demonstrate key components of a future, more-comprehensive implementation of COTS-based solutions and *The National Map* to support HLS and HLD missions.

This task has three main objectives:

- Support NGA and USGS in the data synthesis between state, local, and national systems.
- Support NGA and USGS in the data collaboration and communication between the two agency databases.
- Collaborate with NGA and USGS in reviewing and updating the maintenance strategy for *The National Map* database based on the needs of NGA.

The National Map/Palanterra™ synchronization architecture has evolved in parallel with the development work on the project.

3.1 Operations

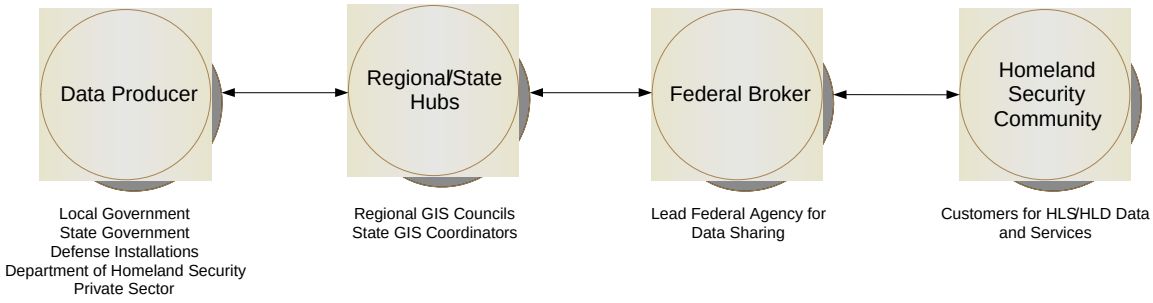
Operationally, task activities are focused on the processes for sharing and maintaining State and Local data with the USGS and fulfilling special requests for data to support the operational use of Palanterra™.

There are four main role types within Project Homeland.

- **Data Producer—**Creates data within local jurisdictions and makes available through Web Services, maps, and datasets
- **Regional/State Hub—**Consolidates data from the data producers
- **Federal Broker—**Limits the need for every federal agency to go to every state for the information it needs to fulfill HLS/HLD responsibilities
- **Homeland Security Community—**Utilizes information similar to local first responders

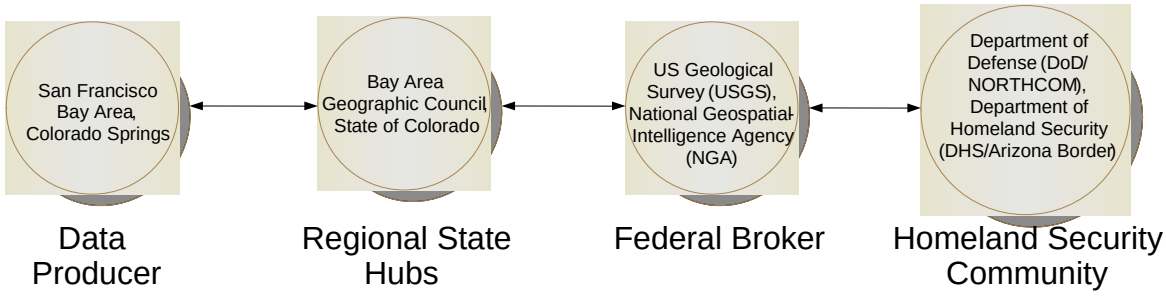
The relationship of data flow between the types of roles is shown in figure 3-1 below.

Figure 3-1
Types of Roles within Project Homeland



Project Homeland's primary stakeholders that fit each of these roles and their relationships to each other are illustrated in figure 3-2 below.

Figure 3-2
Stakeholders within Project Homeland



The USGS acts as a liaison supplying state and local data to members of the HLS/HLD community in support of their missions. In order to support this role, USGS has worked with NGA to instantiate an instance of *The National Map* data holdings within the NGA classified systems. Additionally, USGS receives requests from NGA for data to support a specific event or mission. USGS routes these special requests for data to its local partners and acquires the available resources for provision to NGA.

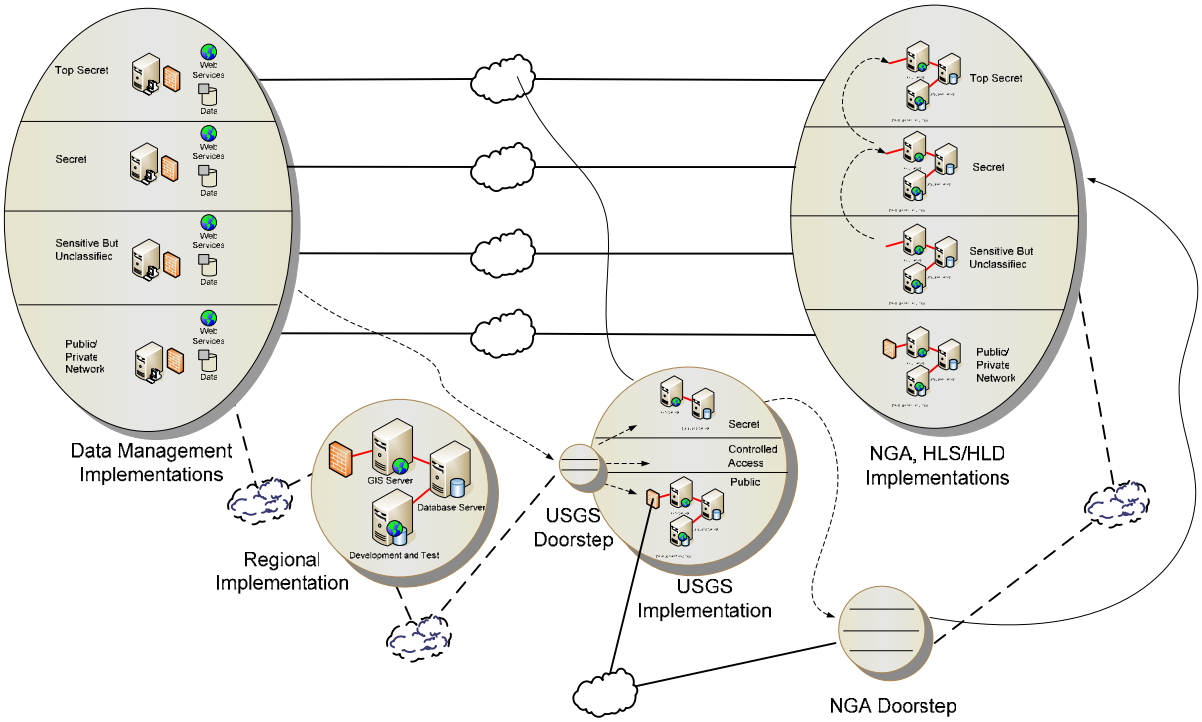
3.2 System

The system resources associated with *The National Map*/Palanterra™ synchronization activities include the local partner, USGS, and HLS/HLD elements required to support the data synthesis and collaboration activities.

There are four major computational elements in the system, as illustrated in figure 3-3.

- Data Management (Local) Implementations
- Regional Implementations
- USGS Implementation
- NGA, HLS/HLD Implementations

Figure 3-3
Overview of System



3.3 Data

One of the key components of the program is the flow of data and metadata between stakeholders. This presents a number of technical and organizational challenges, and the

project approach is to test these parts of the system through a series of pilot projects. Some of the key technical challenges are as follows.

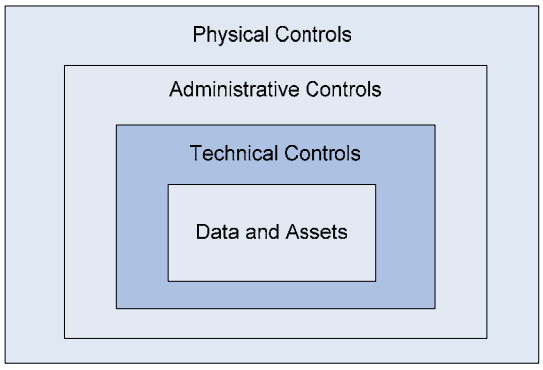
- Aggregation of data from local providers into regional/state systems has implications for data sharing, security, and database design best practices.
- Groups at each level need flexibility to self-define the database design that works for them, but at another level organizations have data requirements that may be different from what local providers currently make available.
- The needs of HLS/HLD users are not well defined for the local data producers. There is a need for better guidance on the information needs of the HLS/HLD community.
- There are design, organizational, and funding aspects to these technical considerations. The goal of this project is to make progress on the technical aspects and document implications in the other areas.

3.4 Security

Figure 3-4 depicts security controls in a layered approach to protect assets.

- **Administrative Controls—**Policies, standards, procedures, guidelines, screening personnel, security awareness training
- **Technical Controls—**Logical access controls, encryption, security devices, identification, authentication
- **Physical Controls—**Facility protection, security guards, locks, monitoring, environmental controls, intrusion detection

Figure 3-4
Layered Security Approach



4.0 Bay Area Regional GIS Council Pilot Summary

4.1 Pilot Overview Summary

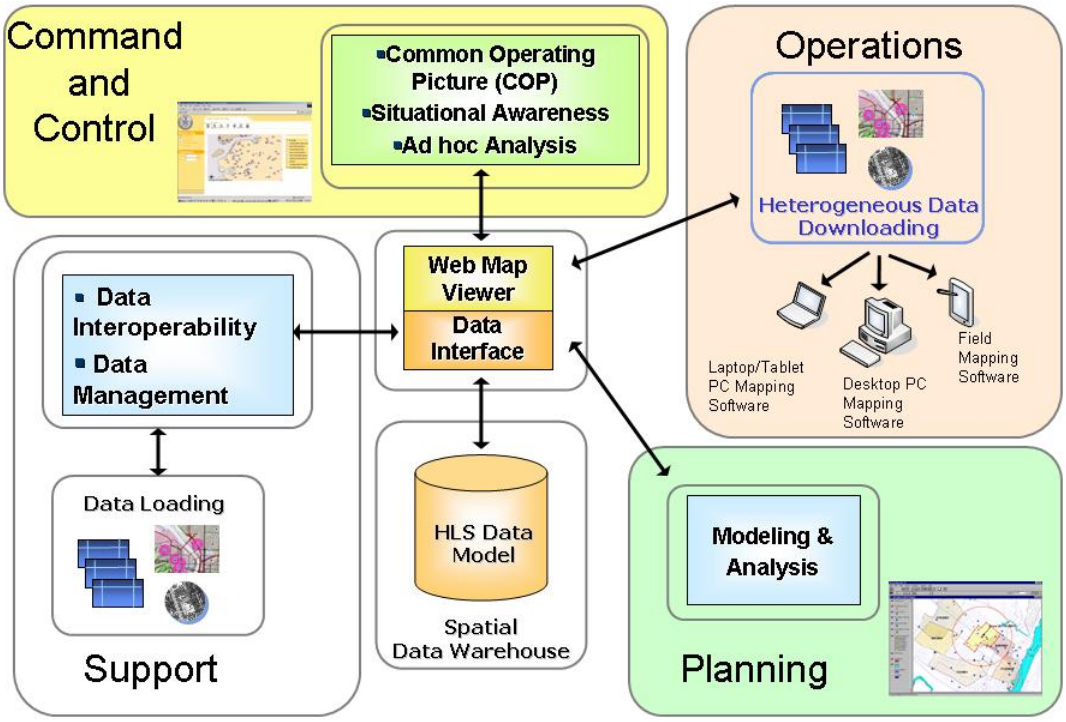
Executive Summary

In 2004, ESRI performed a Pilot Project under Project Homeland for NGA and USGS, in cooperation with the San Francisco BAR-GC, to initiate the development of long-term data sharing capability between local, regional, state, and federal government agencies, which together have primary responsibilities for emergency planning and response in and around the bay area. The significant outcome was a working prototype of a geographic data management and response system that incorporates the data from various bay area municipalities, integrated with state and federal data, for emergency management. This prototype has provided a working blueprint, or template, that can now be delivered to other regional consortiums or states seeking to implement their HLS and HLD GIS infrastructure. This section of the document outlines, at a high-level, the project vision and concepts, while implementation details are provided in the accompanying appendixes.

Vision

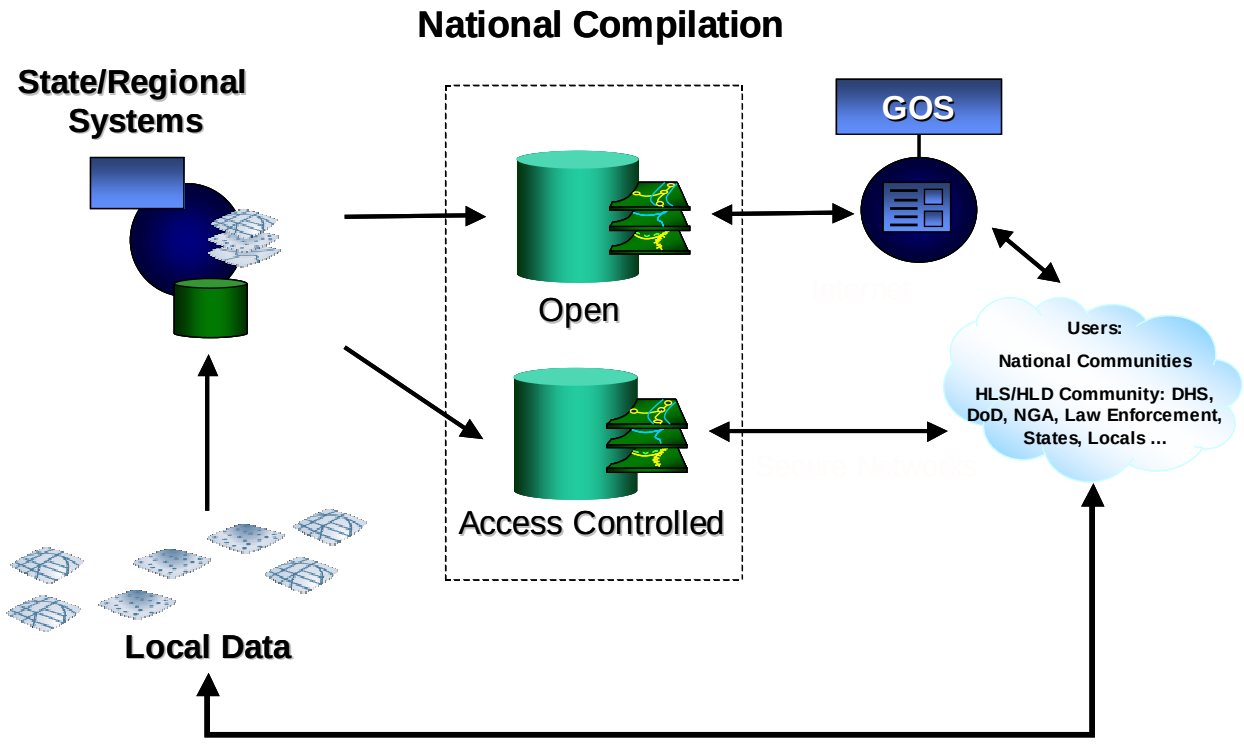
The guiding vision of this approach is to develop state and regional HLS data management system implementations within a national GIS framework. State and regional server implementations will serve as a first step in building a state's HLS infrastructure by providing a consistent view of spatial data atop an HLS data model, a centralized GIS repository, a Web-based map viewer, and a data exchange framework (see figure 4-1), to support first responder activities and operations.

Figure 4-1
Conceptual View of State/Regional HLS Data Management System



Data shared and collected at local levels is synthesized and replicated to other regional or state nodes and eventually to national servers to get incorporated into a national GIS by USGS, as shown in figure 4-2. As part of the process, USGS will broker the sensitive data to the federal HLS/HLD communities, and public data will be served up through the Geospatial One-Stop (GOS) Portal for public access.

Figure 4-2
National HLS Data Exchange and Compilation Framework



4.2 BAR-GC HSDS Data Model Summary

Data Model

The HLS data model has been designed as a hybrid data-sharing and application-driven information model defined to support a variety of HLS prevention, response, and data sharing functions. Before continuing, a distinction should be made between a data-sharing and an application-driven model.

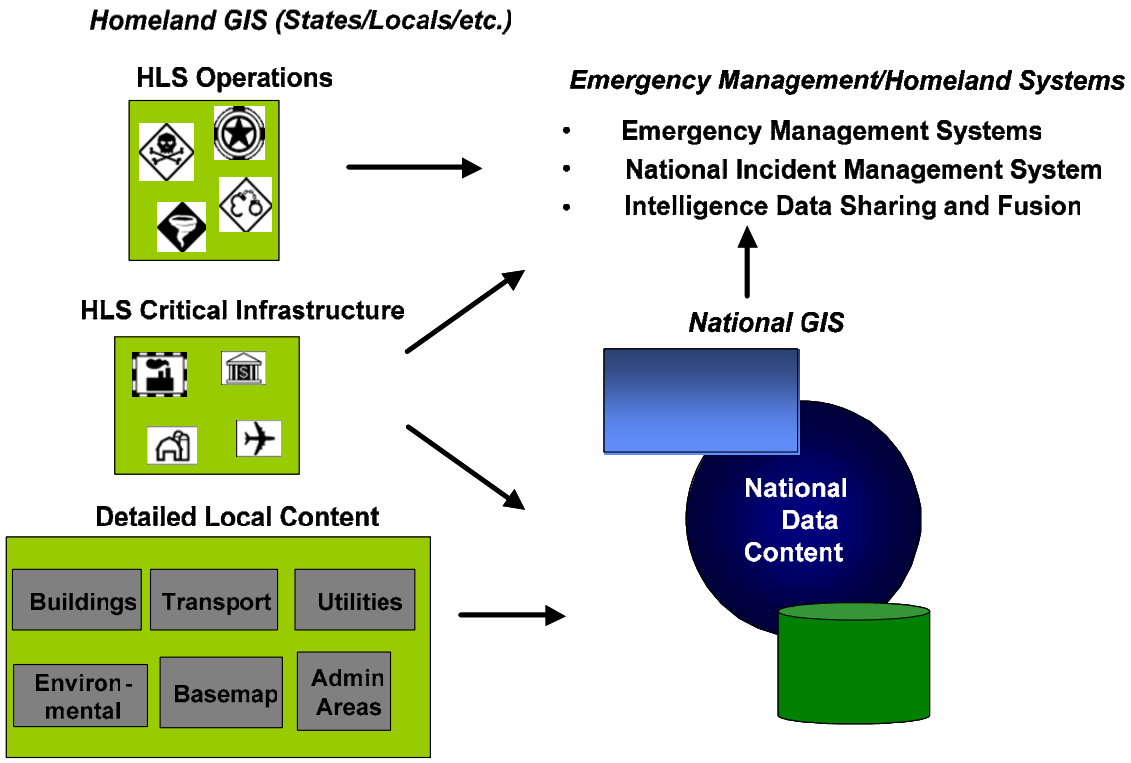
- **Data-sharing model—**A data-sharing model is focused on the application and use of information, rather than a specific organization's work flow.
- **Application-driven information model—**An application-driven model is designed to serve a set of useful applications.

As a hybrid model, the HLS data model supports the data and operation needs of emergency responders and provides a data synthesizing framework at a regional and national scale. The HLS data model can be divided in three broad data categories.

- Detailed Local Content
- HLS Critical Infrastructure
- HLS Operations

Figure 4-3 shows the relationship between these information categories and the respective national datasets and specific local/state/federal homeland and emergency management systems.

Figure 4-3
Conceptual View of HLS Data Model



The detailed local content represents the traditional framework layers plus a few additions. Specifically, this includes buildings and emergency management content along with utilities. These layers represent local best practices for information content, and the content probably goes beyond what is available in most jurisdictions. The recommendation is that this detailed information should be collected primarily where critical infrastructure locations have been

established. The goal is that the data model provide a reasonable container for most detailed data that organizations would collect and manage according to local best practices.

There has been some investigation into adding this type of content into more recent framework layers under the name "Structures." This represents the evolution of ideas started with the Homeland Security Infrastructure Program (HSIP) and continued through the Homeland Infrastructure Foundation Level Data (HIFLD) working group. Recently a Federal Geographic Data Committee (FGDC) Homeland Security Working Group team was established to provide further HSIP-style guidance based on a number of project and organizational activities since the original HSIP documentation was developed.

At a high level, the HLS operations data includes incidents, natural events, key infrastructure, and emergency operations content. This is primarily information that is specific to emergency management and is dynamic in nature. As a result, this is important at a state/regional level but is not currently included as part of the national data-sharing plans.

The HLS critical infrastructure layers were not part of the BAR-GC data model, but they are being included for the pending USGS Colorado pilot. These features represent 31 types of points and lines for such things as police stations, hospitals, fire stations, schools, and large gathering areas. In addition to the basic geometry/location and identity for these key locations, the intent is also to provide emergency point of contact information for these locations. These 31 classifications with minimal attributes are the current recommended best practice for the minimum HLS critical infrastructure data that should be shared at a national level. The goal should be to create an integrated set of national layers out of these high-priority locations. It is likely that access to these datasets will be restricted or access-controlled due to the sensitive nature of the infrastructure they represent.

Further detail on the current data model is provided in Appendix A—Geodatabase Design for Bay Area Regional GIS Council Pilot Project and Appendix B—Homeland Security Data Model Diagrams.

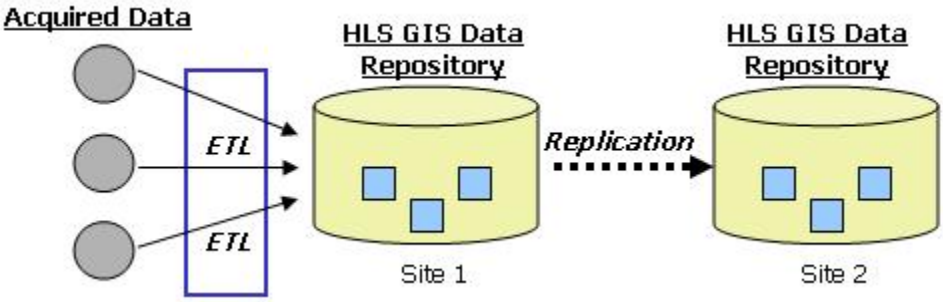
4.3 Data Management Summary

Data Management

Data management is the blanket term used to describe the procedures utilized to collect GIS data from data providers, load data into the HLS data model, and replicate the data to regional or national servers. At a regional scale, data replication allows the HLS community to create a network of redundant data servers, each capable of providing critical geospatial information for the emergency response community. At a national scale, data replication is the data push mechanism to synthesize data at a national level to construct a National GIS and to disseminate the data to the federal HLS/HLD communities. In this context, data management has three components and is illustrated in figure 4-4.

- **Data Acquisition—**Activities associated with data collection.
- **Data Extract, Transform, and Load—**Activities associated with data manipulation.
- **Data Replication—**Activities associated with data exchange.

Figure 4-4
Conceptual View of Data Management Activities



Acquiring data at the local level is usually achieved by working with existing regional partnerships and consortiums and leveraging their existing data-sharing relationships to construct a data-sharing agreement for the purposes of local homeland security and emergency response. Once this is achieved, a password-protected/encrypted site can be established as a staging site to encourage data providers to share data.

Once data is acquired, the task of loading the data into the HLS data model ensues. This is accomplished with COTS tools such as ArcGIS Data Interoperability extension, which maps attributes from source data to the target HLS data model layer. Once the attribute mapping is performed, the tool loads the data into the data model and quality control procedures are run for quality assurance.

Data is replicated between servers to create a redundant network of geographically distributed data servers. This is especially crucial in the HLS community where neither single data servers nor Internet connectivity can be relied upon as a basis for accessing data. A network of redundant data servers provides fail-over backup—for example, when a server goes down in a catastrophe in one area, a server in a different location can be utilized. Data replication can be achieved out of the box with RDBMS technology or ArcSDE® when dealing with disparate RDBMS platforms.

® ArcSDE is a trademark, registered trademark, or service mark of ESRI in the United States, the European Community, or certain other Jurisdictions.

More detailed information is available in Appendix C—Bay Area Regional GIS Council Pilot Project Data Migration Report and Appendix D—Bay Area Regional HSDS ETL and Data Replication Framework.

4.4 Organizational Issues Summary

At the beginning of this project, BAR-GC was an ad hoc group of GIS managers and technical personnel representing many of the jurisdictions and agencies working in the San Francisco Bay Area. As such, BAR-GC could not legally enter into any data-sharing agreements. The Association of Bay Area Governments (ABAG) has officially designated BAR-GC as an ABAG task force. This will resolve the issue of its legal standing as well as create a much stronger organizational platform for BAR-GC to coordinate from.

Outside the GIS community, little is known of BAR-GC efforts. Outreach efforts have identified the need to develop communication materials for distribution among the emergency management and public safety communities.

Working through these organizational issues, risk assessments, site surveys, and memorandums of understanding were developed. These resources are available in Appendix E—Bay Area Regional GIS Council Memorandum of Understanding, Appendix F—Bay Area Regional GIS Council Risk Assessment/Management Plan, Appendix G—Site Suitability Matrix, and Appendix H—Bay Area Regional GIS Council HSDS Host Site Data Center Survey.

4.5 System Security Summary

System Security

Security is a broad and multifaceted topic that encompasses a variety of goals and components. One conventional breakdown of this topic uses the mnemonic "CIA" to delineate three aspects of information security.

- **Confidentiality—**Assurance that information and services are only available to the appropriate users and that privacy and intellectual property rights are protected
- **Integrity—**Assurance that the information is authentic and complete and has been protected from unauthorized additions, deletions, or edits
- **Availability—**Assurance that the information is available when and where needed, without interruption

This framework should be used for analyzing security goals and criteria for any HLS data management system implementation.

The tools and techniques used to reach security goals are also numerous. Normally, these are accomplished by the following resources:

- **Technology—**Hardware and software mechanisms
- **Procedure—**Standards and norms for users
- **Human Resources—**Training, awareness, and attitudes
- **Organization—**Structures, practices, and policies for intra- and interorganizational cooperation

For security implementation guidelines refer to Appendix I—Bay Area Regional HSDS Security Framework.

4.6 System Architecture Summary

System Architecture

The system architecture for a local implementation, one node, of an HLS data management system consists of a Web/application server, a data server, a storage server, and an external hard drive. Each provides the following:

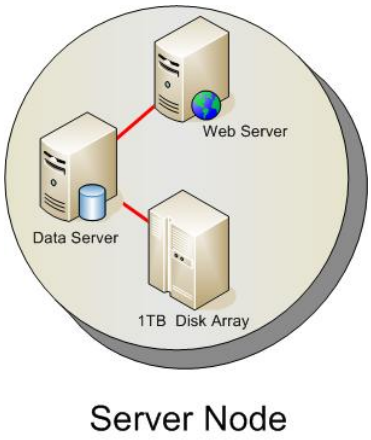
- **Web/Application Server—**Provides for FTP server for data sharing, Internet map server for serving maps, and application server for the Web-based map viewer.

- **Data Server—**Hosts the GIS data repository.

- **Storage Server—**Provides at least 1 terabyte of storage for the data server. For state or regional implementations, increase storage accordingly to be able to absorb data from local nodes.

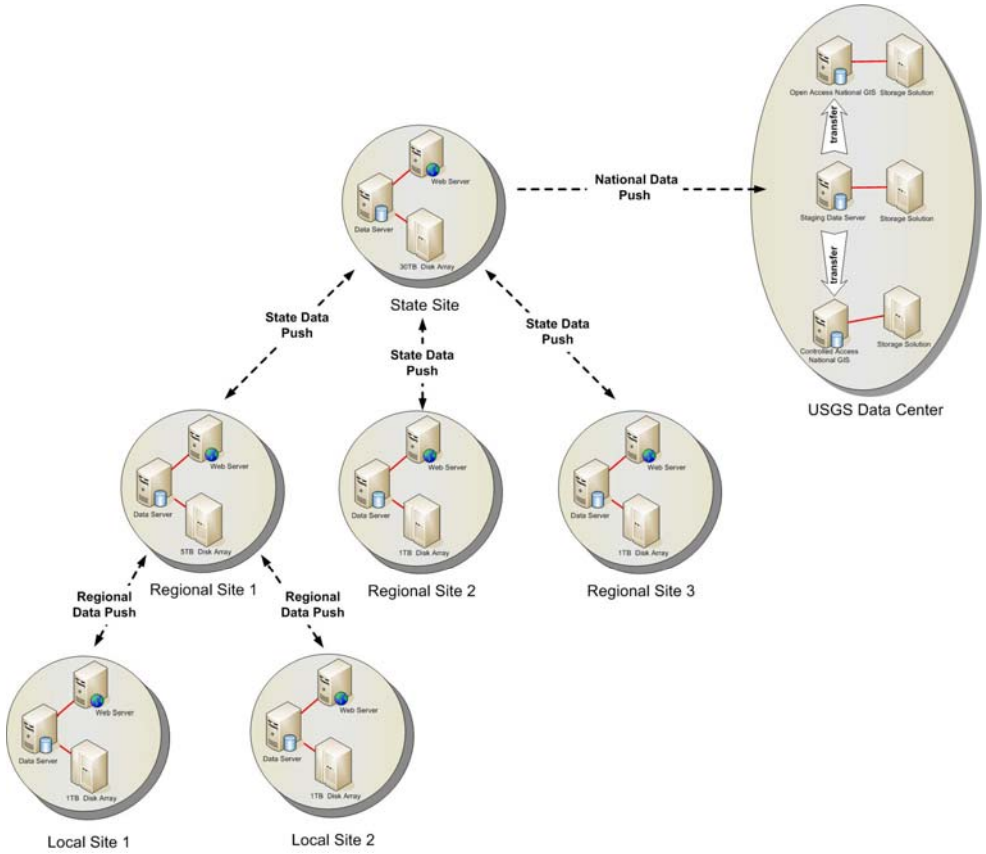
- **External Drive—**Provides a fast method of exchanging raster data.

Figure 4-5
System Architecture for One Node of an HLS Data Management System



At a regional or state level, server nodes can be added on and linked together to provide a network of redundant data servers, each capable of serving their local first response community in emergency situations. To achieve redundancy and data sharing, each server is configured to replicate data to a regional central node.

Figure 4-6
System Architecture for a State Network of HLS Data Management System Nodes



The detailed Reference Architecture is in Section 4 of Appendix J—*The National Map*/Palanterra™ Synchronization System Architecture Design.

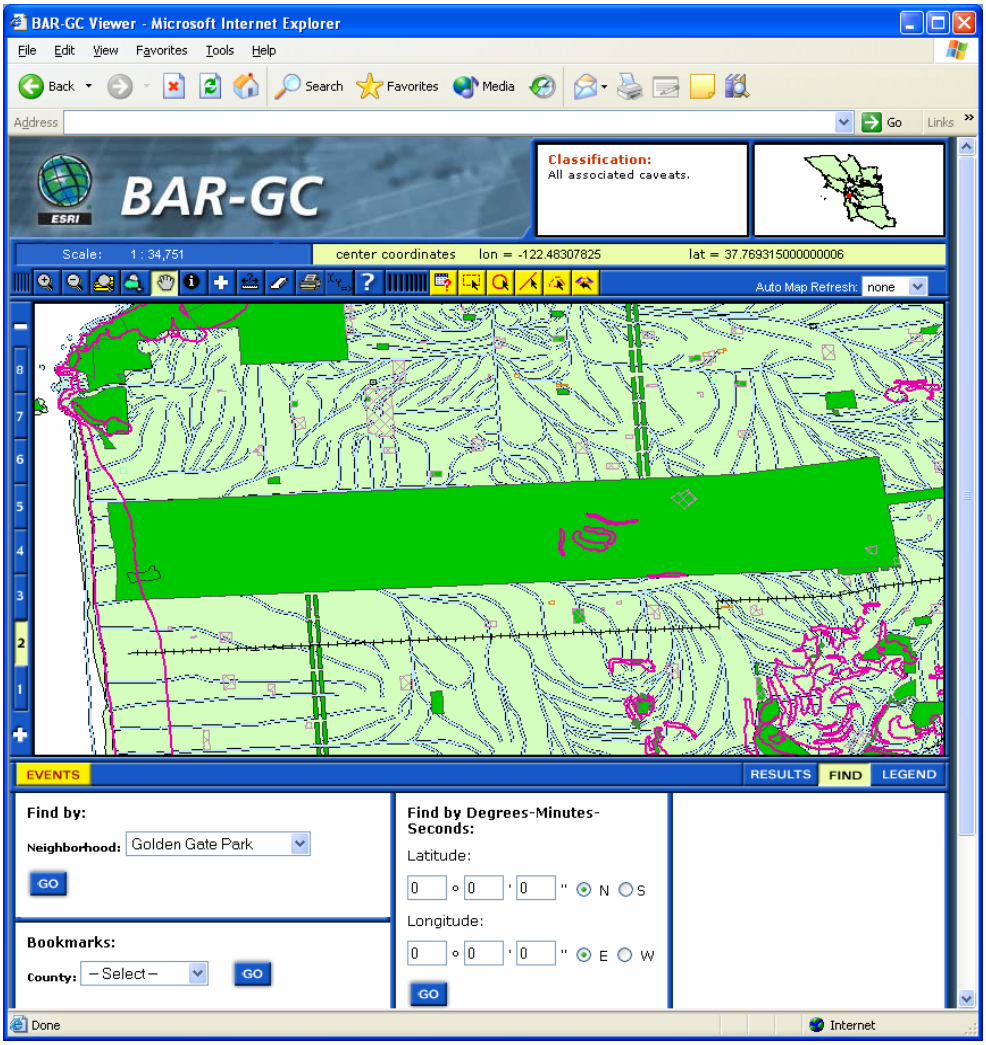
4.7 Operational Use/Viewer Summary

Sample Application

Once an HLS data management system has been developed at one location or a regional network of servers implemented, applications can be developed to take advantage of the jurisdictional data consolidation. One such application is a Web-based map viewer that provides an operating picture of the region. Such an application can be useful for providing a base environment for situational awareness at a local, regional, or national scale. The application can be used for ad hoc spatial analysis, data download for remote users, and information access. Furthermore, the map viewer can be enhanced as a common operating

picture by capturing tracking feeds from GPS-enabled first responders out in the field or by integrating data feeds from an emergency operations center.

Figure 4-7
Sample Map Viewer



Other applications that can be developed atop the HLS data management system include a critical infrastructure inventory system, field data collection systems, map production operations, and advanced spatial analysis systems.

More detail is provided in the BAR-GC Viewer Release Notes, which is Appendix K—Bay Area Regional GIS Council Viewer Release Notes.

5.0 Lessons Learned Summary

5.1 Data Model

The Project Homeland data modeling effort determined that it was better to focus on a real data environment, versus trying to identify the optimum data model through abstract methods. So the BAR-GC data was used to test the model as it stood in late September 2004. This was an expansion in scope of the project and was well received by BAR-GC. As the system starts to grow as an operational system, the data model will need to evolve as new needs are identified.

5.2 Data Management

The process of gathering and processing data into the data model was far more difficult than it was expected to be. It quickly became apparent that it was time consuming to determine what all the BAR-GC data was in order to map it to the target Homeland Security Model. Topological issues also created some surprise problems, requiring some robust and processor-intensive solutions.

5.3 Organizational Issues

At the beginning of this project, BAR-GC was an ad hoc group of GIS managers and technical personnel representing many of the jurisdictions and agencies working in the San Francisco Bay Area. As such, BAR-GC could not legally enter into any data sharing agreements. The Association of Bay Area Governments has officially designated BAR-GC as an ABAG task force. This will resolve the issue of it's legal standing as well as create a much stronger organizational platform for BAR-GC to coordinate from.

Outside the GIS community, little is known of BAR-GC efforts. Outreach efforts have identified the need to develop communication materials for distribution among the emergency management and public safety communities.

5.4 System Security

Much of the work for implementing system security for BAR-GC remains to be done. The technical means for controlling access are in place, but the organizational structure to determine specifically who will have access to what remains to be developed. Setting up the technical structure was not as straightforward as hoped, and differences between locations posed unique problems. Unlike many data situations, the data going into the BAR-HSDS is

meant to move around, both across the Internet and via hard media. Many GIS professionals are proficient in moving data but are unfamiliar with simple hardware maintenance procedures.

5.5 System Architecture

Unlike most systems that are intended to provide redundancy, the BAR-GC opted to have two–four servers around the region meet the goal. This presented unique challenges for systems folk, because they needed to work with other servers in new ways. Institutional protocols in addition to technical solutions needed to be developed to maintain order in the process. Incremental updates have proven elusive due to a lack of consistent and complete unique IDs for the original datasets coming from the local jurisdictions.

5.6 Operational Use/Sample Applications

There are a number of powerful features in the BAR-GC viewer. Some, such as the Data Download, work well and can be relatively simple. The cartographic look and feel of the viewer, although seemingly simple, has actually turned out to be particularly challenging. The inclusion of an event tracking system, especially one that integrates with existing systems, appears to be an important next step, with the emphasis on using the BAR-HSDS as a response data viewing and transfer point.

5.7 Conclusion

At the national level, USGS is currently developing a program for managing HLS data from state and regional systems and incorporating it into the national GIS. USGS will be embarking on a Colorado pilot project to make advances on the lessons learned in the BAR-GC project and will be passing on those implementation guidelines in the form of an Implementation Template. In the mean time, to move forward with the implementation of state and regional HLS data management systems within this national GIS framework, the appendixes in this document provide details as to approach, configuration, and other relevant technical matters encountered in the BAR-GC pilot project to serve as a first step in bridging the implementation information gap.

5.8 Future Work Recommendations

- Universal Task List analysis and lookup to BAR-GC findings
- Expanded catalog of information products
- Exercises and stress testing through disaster scenarios

- Continued training
- Data operations and maintenance plan, budgeting

Appendix A

Geodatabase Design for Bay Area Regional GIS Council Pilot Project

NGA Homeland Security Enterprise GIS Support Project

Geodatabase Design for BAR-GC Pilot Project

ESRI Database Services

Prepared by: ESRI
Version: Final
Last Modification Date: December 31, 2004

Prepared for:

Brad Lucas, Contracting Officer's Representative
National Geospatial-Intelligence Agency
4600 Sangamore Road
Bethesda, Maryland 20816-5003
Tel.: 301-287-6572

Prepared by:

ESRI
380 New York Street
Redlands, California 92373-8100

Table of Contents

Section **Page**

Section **Page**

Appendixes

Change History		
Date	**Description**	**Person(s) Responsible**
Dec. 20, 2004	Initial Draft	Matt Lott/Jan Mares
Dec. 31, 2004	Final Deliverable	Charlie Wells

1.0 About This Document

1.1 Purpose

This document describes the homeland security (HLS) data model developed for the Bay Area Regional GIS Council (BAR-GC) pilot project as part of the NGA Homeland Security Enterprise GIS Support Project (Project Homeland).

The BAR-GC data model represents an initial implementation of a homeland security data model that incorporates foundation data and operations data that is typically of interest to the emergency response community at the local level. The model is based on ESRI's best practices homeland security data model that was developed for local governments, with some modifications that have been implemented for the pilot project.

1.2 Intended Audience of This Document

The Geodatabase Design for the BAR-GC Pilot Project is intended primarily for use by BAR-GC member organizations and agencies that, through arrangements with BAR-GC, make use of GIS data in the Bay Area. The document is also intended for use by NGA Project Homeland management team and technical staff, management and technical staff from NGA's partner agencies, and the ESRI NGA Homeland Security Enterprise GIS Support Project team.

1.3 Document Overview

This document is organized into the following sections:

- **Section 1.0** presents an introduction to the ESRI geodatabase and the BAR-GC pilot project.
- **Section 2.0** presents an overview of the BAR-GC specific geodatabase design issues including those considered in the design process.
- **Section 3.0** presents the specific properties of the geodatabase feature data sets.
- **Section 4.0** presents specific properties of the geodatabase object classes, feature attributes, feature subtypes, and relationship classes.

- **Section 5.0** presents the list of attribute domains used in the BAR-GC pilot geodatabase.

- **Appendix A** provides an example of the HTML report of the BAR-GC pilot geodatabase design presented in this document as well as a layer listing for the geodatabase.

- **Appendix B** provides a general overview of information concerning the geodatabase design model. **Note:** For more detailed information on the geodatabase, please refer to other ESRI publications, specifically two books published by ESRI Press entitled Building a Geodatabase: Case Studies in GIS Data Modeling and Modeling Our World: The ESRI Guide to Geodatabase Design. Other external documents are listed in Section 1.3.1.2.

- **Appendix C** provides a summary of the ArcGIS HLS Model in Microsoft Word format. The Word document also provides several sources for more information on the concepts used to create the Homeland Security (HLS) geodatabase model. Also included in the appendix are two GIF files, which present the submodel concept, and their role in the model in graphic format—the ArcGIS Homeland Security Analysis Data Model.

1.4 Assumptions, Dependencies, and Constraints

The following assumption is applicable to the scope of work described in the NGA HLS Data Model Work Plan.

- The HLS data model developed for BAR-GC represents an initial implementation of a local HLS data model. The data model is expected to evolve as a result of information gained during the creation of the BAR-GC pilot database, as well as through the implementation of the model in at least one other pilot project.

1.5 Referenced Documents

This section lists other documents referenced in the BAR-GC Geodatabase Design document. These include project documents that are deliverables as well as external documents referenced by this document.

Note: The digital versions of the project documents listed in Section 1.5.1 were delivered along with this Geodatabase Design document in the geodatabase project zip file.

1.5.1 Project Documents

Document Name	Date	Document Owner
Project Homeland BAR-GC National Map Pilot Project Management Plan (Word document)	11/08/2004	ESRI
BAR-GC ETL Data Migration Report (Word document)	12/20/2004	ESRI
BAR2_RevisedMDB_reporter.htm (HTML document)	12/2004	ESRI
BAR-GC Geodatabase Data Dictionary (HTML document)	12/2004	ESRI

1.5.2 External Documents

Document Name	Date	Document Owner
Designing Geodatabases—Case Studies in GIS Data Modeling	2004	ESRI Press
GIS and Land Records—The ArcGIS Parcel Data Model	2004	ESRI Press
Modeling Our World—The ESRI Guide to Geodatabase Design	1999	ESRI Press
Arc Hydro—GIS for Water Resources	2002	ESRI Press
ArcGIS HLS Model (Word document)	2004	Steve Grise
ArcGIS Homeland Security Analysis Data Model (.gif)	2004	Steve Grise

2.0 BAR-GC Geodatabase Overview

This document represents the final BAR-GC geodatabase as refined during completion of the BAR-GC pilot project. It addresses all aspects of the BAR-GC pilot geodatabase including feature data sets, feature classes, feature attributes, and attribute domains.

ESRI worked with the Bay Area Regional Homeland Security Data Server (BAR-HSDS) Working Group to develop the BAR-GC geodatabase model based on the ESRI homeland security data model template. The data model provides the foundation for moving data from many different sources at the local level to be utilized successfully at a single source at the federal level. This effort coincides with a BAR-GC goal to develop data and attribute feature mapping between various data layers contained in the local input data sets and data and attributes contained in the HLS Best Practices Data Model.

2.1 Geodatabase Characteristics

Just as the analytical capabilities of GIS software have evolved, recent advances in GIS data modeling are facilitating more accurate representations of real-world features and behaviors in the GIS database. The representation of earthly features by vector coordinates alone (as in ESRI's coverage, shapefile, and ArcSDE layer data models) has an important shortcoming—features are aggregated into homogeneous collections of points, lines, and polygons with generic behavior. The behavior of a line representing a road is identical to the behavior of a line representing a stream.

A new object-oriented data modeling concept—the geodatabase data model—allows modelers to endow features with their natural behaviors and define specific characteristics about the relationships between features.

- **A uniform repository of geographic data—**Because a geodatabase can be implemented within an relational database management system (RDBMS), geographic data can be stored and centrally managed in one database.

- **Data entry and editing are more accurate—**Fewer mistakes are made because most of them can be prevented by intelligent validation behavior. For many users, this alone is a compelling reason to adopt the geodatabase data model.

- **Users work with more intuitive data objects—**Properly designed, a geodatabase contains data objects that correspond to the user's model of data. Instead of generic points, lines and areas, the users work with objects of interest such as landmarks, open water, and roads.

- **Features have a richer context—**With topological associations, spatial representation, and general relationships, users not only define a feature's qualities but also its context with other features. This lets users specify what happens to features when a related feature is moved, changed, or deleted. This context also lets you locate and inspect a feature that is related to another.

- **Better maps can be made—**You have more control over how features are drawn, and you can add intelligent drawing behavior. You can apply sophisticated drawing methods directly in a mapping application.

- **Features on a map display are dynamic—**When users work with features, they can respond to changes in neighboring features. You can also associate custom queries or analytic tools with features.

- **Shapes of features are better defined—**The geodatabase data model lets you define the shapes of features using straight lines, circular curves, elliptical curves, and Bézier splines.

- **Sets of features are continuous—**By their design, geodatabases can accommodate very large sets of features without tiles or other spatial partitions.

- **Many users can edit geographic data simultaneously—**The ArcSDE geodatabase data model permits work flows where many people can edit features in a local area and then reconcile any conflicts that emerge.

The principal advantage of the geodatabase data model for the BAR-GC project is that it incorporates many of the advances in geospatial data storage to create a framework of data layers. These data layers can be populated with intelligent features that mimic the interactions and behaviors of real-world objects, providing a highly sophisticated set of data for users and facilitating data integrity and maintenance for staff who will be responsible for maintaining the BAR-GC enterprise GIS data resources.

2.2 Geodatabase Organization

The base model for the BAR-GC pilot geodatabase was the homeland security best practices data model template. This model was updated to reflect the findings of the BAR-HSDS working group and any issues specific to the source data received from the BAR-GC members. In most other homeland security modeling efforts, the work has revolved around documenting a set of layers rather than detailed data/information needs. The goal of this model has been to capture these layers as feature types and organize the feature classes and attributes closely with HLS local government best practices in mind (see Table 4-1).

The notation of this model involves some basic Unified Modeling Language (UML) diagrams with icons that represent whether the feature classes (also known as database tables at an implementation level) described are points, lines, or polygons. Efforts have been made to simplify the diagrams for use by non-HLS specialists.

2.3 Geodatabase Specifications

The specifications described in this section are applicable to the BAR-GC pilot geodatabase as a whole.

2.3.1 Spatial Reference

All data in the geodatabase is in the geographic coordinate system (GCS_WGS_1984) as shown in Figure 2-1. The mapping units are in decimal degrees. This spatial reference will allow for the viewing of continuous data worldwide.

Figure 2-1
Spatial Reference

Dimension	Minimum	Precision
Homeland Security Operations		
X	-190	2000000
Y	-100	
M	0	100000
Z	0	100000
Coordinate System Description GEOGCS["GCS_WGS_1984",DATUM["D_WGS_1984",SPHEROID["WGS_1984",6378137.0,298.257223563]], PRIMEM["Greenwich",0.0],UNIT["Degree",0.0174532925199433]]		
Local Base Data		
X	-190	2000000
Y	-100	
M	0	100000
Z	0	100000
Coordinate System Description GEOGCS["GCS_WGS_1984",DATUM["D_WGS_1984",SPHEROID["WGS_1984",6378137.0,298.257223563]], PRIMEM["Greenwich",0.0],UNIT["Degree",0.0174532925199433]]		

2.3.2 Metadata

Metadata is tracked through a table called Dataset_Detail. This table contains a record for every source that was used in populating the BAR-GC geodatabase as well as a location attribute to find the metadata for each original source. No new metadata was created during the pilot Extract, Transfer, and Load (ETL) data migration process.

2.3.3 Security

Geodatabase security is managed via the RDBMS. Therefore, accessing data requires a connection to the host RDBMS. A user must enter connection information including a user name and password. This connection information, passed on to the ArcSDE server, is used to log in to the RDBMS. If the login connection is successful, the user can begin working with the geodatabase.

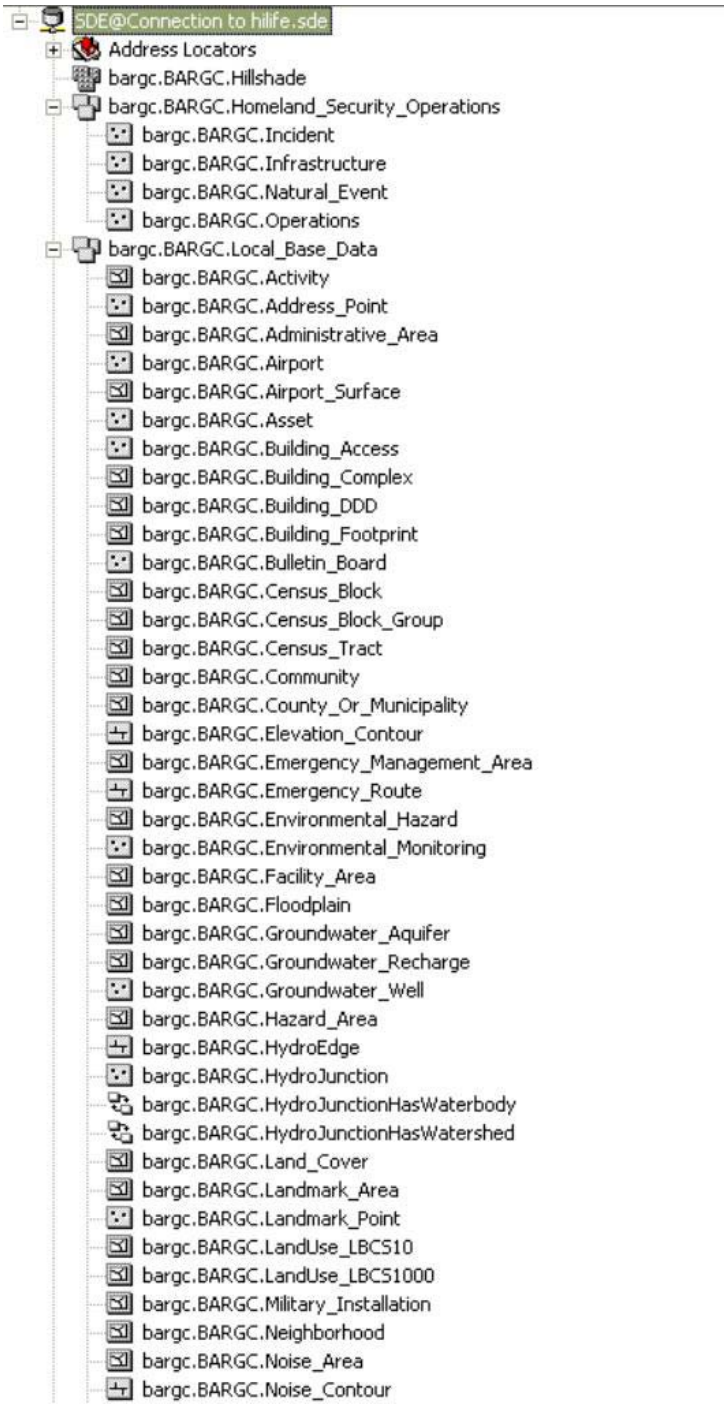
Access privileges for users of the GIS will be established by BAR-GC prior to system implementation. These privileges will be implemented in the RDBMS prior to the users going online with the system. To restrict write permissions to the database, the ArcSDE administrator user can set the permissions to the default version of the database to establish protection, while still allowing users to view the data and perform necessary functions on the data. ArcGIS 9.x is COM-compliant, and ArcSDE supports versioning and long transactions, permitting multiuser editing of the database and procedural management of database conflicts.

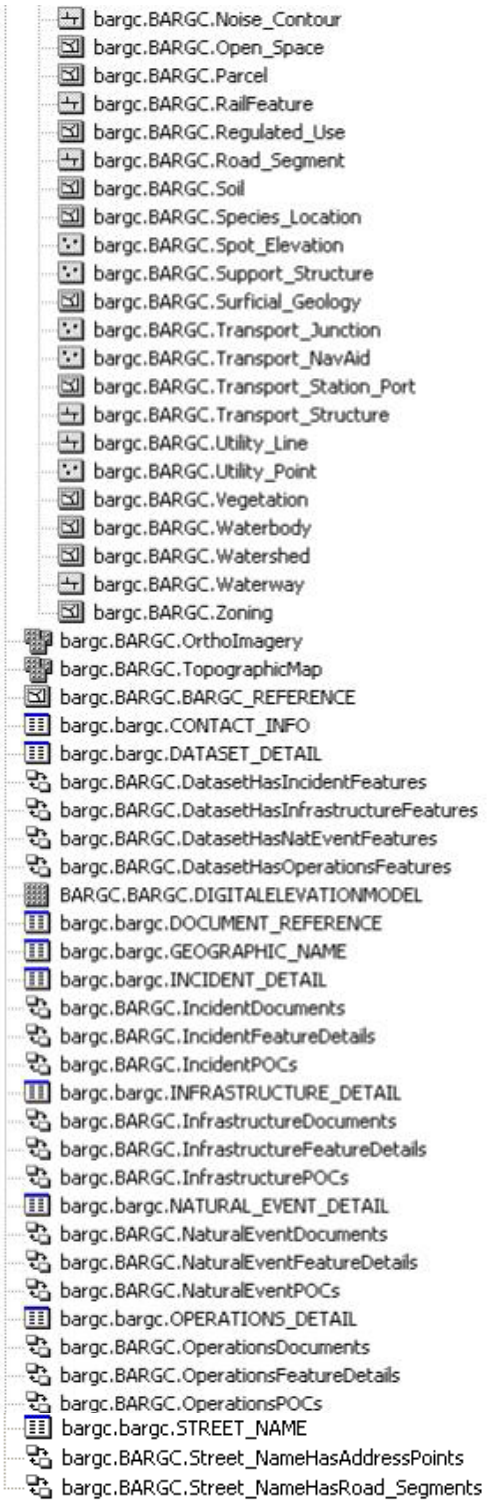
2.4 Feature/Object Class Descriptions

For the BAR-GC model, the feature/object class descriptions can be reviewed in the BAR-GC Geodatabase Data Dictionary HTML document provided in Appendix A. For each feature class, the specifications presented include attributes, subtypes, relationships, annotation, and domains. Attributes for each feature class are also shown in the diagram, and they reflect the attribute name, a description of the attribute, and any attribute domains associated with the attribute. The diagram also reflects whether the feature class utilizes subtypes and the fields related to each of those subtypes. Relationships between a specific feature class and its related feature/object class are shown with connecting line work.

2.5 ArcCatalog View of the Database

The following graphic illustrates the ArcCatalog view of the BAR-GC pilot ArcSDE geodatabase. This view shows the geometry type of each layer as well as its organization within a specific feature data set.





2.6 Model Revisions for the BAR-GC Pilot Project

The following subsections document the changes that were made to the geodatabase during the BAR-GC pilot ETL process. It also lists future proposed changes to the BAR-GC data model should they be required.

2.6.1 General Data Model Issues

This section outlines general data model issues identified and addressed during the pilot.

- Input source attributes that could not be mapped to the target geodatabase were not included in the pilot.
    - A possible future solution is to implement traceback capability whereby the entire set of source attributes would be inserted as Geography Markup Language (GML) data into binary large object (BLOB) fields to be retrieved when needed. At present, the BLOB field is not supported in the ArcGIS Data Interoperability Extension. Therefore, more research will need to be done before this process can be implemented.
- Several input data sources can be mapped to both the HLS Operations and the Local Base Data Feature Datasets. As a proof of concept, only a select sample of input source data was mapped to both feature datasets (i.e., Airport, Bus Terminal, and State Building, to list a few). Therefore, a decision must be made to finalize whether or not these input source data types are mapped to one or both feature datasets. One option is that the HLS Operations Feature Dataset only contains mission-specific data, which represents a subset of these features, which can also be contained in the Local Base Data Feature Dataset, such as bridges, government installations, incident locations, etc.
    - North American Industry Classification System (NAICS) codes were not used in the pilot mapping example.
- Existing input source data feature relationships and domain values were evaluated for incorporation into the BAR-GC geodatabase. Fields mapped to the model also had the related tables and domain values mapped where appropriate. If a target domain and/or relate table was not needed because the related input source attribute was not being mapped, it will remain in source and not appear in target.
- Metadata documents such as PDFs and Word documents are not included in the ArcSDE geodatabase. The Dataset_Detail table provides a link to locate these ancillary input source data files.

- The elevation attributes in Elevation_Contours and Spot_Elevation feature classes were populated as-is from the original input source.

2.6.2 Changes Made to BAR-GC Data Model During ETL Data Migration Process

This section outlines the specific changes made to the geodatabase model during the pilot.

- Changed entire model to reflect the GCS_WGS_1984 projection
- Added polygon Feature Class "Vegetation" to the model
- Added polygon Feature Class "Landmark_Area" to the model
- Added a domain for Address_Point tied to Address_Type, ADDRESS_TYPE (Parcel, Primary Building, Building Entrance, Supplemental Building, Other)
- Added a domain for Airport_Surface tied to Feature_Type, FTYPE_AIRPORT_SURFACE (Runway, Taxiway, Apron/Hardstand, Other)
- Added a domain SOURCE_DATA_SECURITY to control access and use constraints
- Added a domain for Landmark_Area tied to Feature_Type, FTYPE_LANDMARK_AREA (Cemetery, Park, Named Place, Reserve, Golf Course, Other)
- Expanded domain FTYPE_HAZARD_AREA to accommodate FEMA flood zones
- Expanded domain OWNER_CLASS to accommodate county-owned parcels
- Expanded domain FTYPE_MILITARY_INSTALLATION to accommodate Air Force and Coast Guard bases
- Expanded domain FTYPE_AIRPORT to accommodate military airfields
- Changed tagged value LENGTH = 255, Parcel_Address, Tax_Address attributes of Parcel to accommodate full address from source
- Changed tagged value LENGTH = 50, in Operations_ID attribute in FC Contact_Info to accommodate Permanent_ID

- Changed tagged value LENGTH = 50, Classification attributes of Soil to accommodate full coding from source

- Changed tagged value LENGTH = 50, Zone_Code, attributes of Zoning to accommodate full coding from source. Dropped Zone_Code domain to limited for the various sources

- Changed tagged value LENGTH = 50, Zone_Code, attributes of Parcel to accommodate full coding from source. Dropped Zone_Code domain to limited for the various sources

- Changed tagged value LENGTH = 255, Name, attribute of Building_Footprint to accommodate full names from source

- Combined Access and Use_Constraints fields in Dataset_Detail table to make a field AccesandUse_Constraints. This field was tagged to the Source_Data_Security Domain

- Changed Name attribute of Street_Name table to Primary_Name; added Alternate_Name1, Alternate_Name2, Alternate_Name3. Set tagged value Length = 150 for all. Dropped the attribute Primary_Or_Alternate

2.6.3 Proposed Changes to Best Practices Data Model for Future Pilots

This section outlines the proposed changes for future pilot projects.

- Change tagged value LENGTH = 255, Name, attribute of Building_Complex to accommodate full names from source.

- Change tagged value LENGTH = 255, Name, attribute of Landmark_Area to accommodate full names from source.

- Change tagged value LENGTH = 255, Landmark_Name, attribute of Landmark_Point to accommodate full names from source.

- Change domain for SEC_CLASSIFICATION to be short_integer, SEC_CLASSIFICATION (0 = Unknown, 1 = TS, 2 = S, 3 = Confidential, 4 = Restricted, 5 = UNCLASSIFIED, 999 = Other)

3.0 Feature Data Sets

This section describes the feature data sets that are used for the BAR-GC geodatabase.

3.1 Homeland_Security_Operations Feature Data Set

HLS Operations users want to tag operations, infrastructure, incidents, and activities onto an underlying basemap. One way to think of this is that most operations people are looking for digital pushpins to put on top of a map—a replacement for the traditional wall map and pushpins. The subtle data design and GIS specialist considerations are not of much interest, but good, accurate base data is of great interest. The ideas suggested in this data model are that we combine the Federal Geographic Data Committee (FGDC) symbology standards with the Homeland Security Working Group Symbology Standard (www.fgdc.gov/HSWG) and implement this as an ArcGIS feature data set.

3.2 Local_Base_Data Feature Data Set

The Local_Base_Data feature data set contains many feature classes that represent the "background" information that would typically be found on a homeland security preparedness and response map. It includes such features as roads, streams, buildings, and topography. For documentation purposes, this description has been broken into submodels that will be described in more detail below. For a graphic description of each submodel and some of its complex relationships, see Appendix C.

The first submodel is Emergency Management and Buildings. The most complex part of the Emergency Management submodel involves buildings. The level of detail that local organizations manage depends on funding and design decisions. The most sophisticated view provides enough detail for emergency management personnel to effectively respond to local emergencies. This involves building footprints, building entrances/access points, and area descriptions of complexes and campuses. It also includes a link to addresses for the buildings. It may also include 3D representations of buildings. The intent in this part of the model is that you can choose any and all of BuildingAccess, BuildingFootprint, and BuildingComplex classes for your design and use a common list of feature types for all three feature classes.

The Transportation submodel is a large part of the model. The intent is to capture basic infrastructure information for water, air, road, transit, and rail networks. The model also includes address points and address ranges (along street segments) to provide a simple data structure for geocoding and address matching.

The Environmental submodel includes a variety of classes that relate to environmental monitoring and response. Most of these are oriented more toward natural disasters and recovery efforts rather than environmental monitoring for homeland security purposes. However, many feature class that are of great significance when preparing for homeland security are included, such as the locations and descriptions of hazardous materials storage and natural hazards. Also, land use and land cover features, for example, can be used to assess damage after an event.

The Utilities submodel includes a high-level model for utility infrastructure. This is a simplistic model that will be adequate for many emergency planning and response activities but is not enough for infrastructure managers. More research is required to get the right balance between information needs and data sensitivity concerns in this area.

The Governmental Units/Boundaries submodel includes the basic administrative areas for local government. While the Administrative_Areas class shows one feature class with many feature types, most implementations will actually organize this into multiple feature classes based on data management responsibilities and other more technical considerations. Basic census and community/neighborhoods and cadastral uses and restrictions are also included.

The BaseMap submodel contains common feature classes and attributes for basemapping purposes. It does not attempt to provide a multiscale representation because it can be assumed that, at a local level, the most detailed scale (under 1:24,000) is what local agencies manage. Content includes cartographic products as backdrop imagery, orthophotos, hydro, cadastral, landmarks, and geographic names content. Military installation footprints are also included, with the expectation that other data from defense organizations will fall into other classes in the model (utilities, environmental, etc.).

3.3 Stand-Alone Object Classes

For the BAR-GC pilot, a number of stand-alone object classes have been developed. An object class is a table in the Geodatabase that stores non-spatial data. It is a matrix of rows that represent objects/features and columns that contain feature attributes and information. An example of one BAR-GC stand-alone object class is the Street_Name object class. When joined to the appropriate feature classes within the Geodatabase it will provide the primary and alternate names for road features in the Geodatabase. A listing of the stand-alone object classes currently contained in the BAR-GC pilot geodatabase is reflected at the end of the following table, Table 4-1. Also listed in this section of the table are several tables that contain raster information.

4.0 Feature Classes and Object Classes

This section provides a listing of the Feature and Object classes (Table 4-1) that exist in the BAR-GC geodatabase. For a complete diagram of the feature and object classes that includes the subtypes, domains, and relationships, please view the HTML file BAR-GC Data Dictionary that accompanies the digital delivery of this document.

Table 4-1
Feature Data Sets and Feature Classes

Feature Data Set	Feature Class	Description
Homeland Security Operations		
	Incident	Location pushpin for incidents
	Infrastructure	Location pushpin for infrastructure features
	Natural_Event	Location pushpin for a natural event
	Operations	Location pushpin for an operation
Local Base Data		
	Activity	Location of an activity
	Address_Point	Address points
	Administrative_Area	Administrative areas
	Airport	Airports
	Airport_Surface	Airport runways, taxiways, etc.
	Asset	Location of an asset
	Building_Access	Location of entrance and exit points
	Building_Complex	The larger grouping of individual buildings
	Building_DDD	Three-dimensional building features
	Building_Footprint	The outline of buildings or roofs
	Bulletin_Board	A push-pin feature class for miscellaneous information
	Census_Block	Census blocks
	Census_Block_Group	Census block groups
	Census_Tract	Census tracks
	Community	Community boundaries
	County_Or_Municipality	County or city boundaries
	Elevation_Contour	Elevation contours
	Emergency_Management_Area	Emergency management areas
	Emergency_Route	Emergency routes
	Environmental_Hazard	Environmental hazard areas

Feature Data Set	Feature Class	Description
	Environmental_Monitoring	Environmental monitoring sites
	Facility_Area	Facility areas
	Floodplain	Floodplains
	Groundwater_Aquifer	Groundwater aquifers
	Groundwater_Recharge	Groundwater recharge
	Groundwater_Well	Groundwater wells
	Hazard_Area	Hazard areas
	HydroEdge	Shorelines, rivers, streams
	HydroJunction	Location where hydro features meet at a point
	Land_Cover	Land cover types
	Landmark_Area	Landmark areas
	Landmark_Point	Landmark points
	LandUse_LBCS10	Land Use Classifications LBCS10
	LandUse_LBCS1000	Land Use Classifications LBCS1000
	Military_Installation	Military Installation boundaries
	Neighborhood	Neighborhood boundaries
	Noise_Area	Noise sensitive areas
	Noise_Contour	Noise level contours
	Open_Space	Open space areas
	Parcel	Cadastral parcels
	RailFeature	Railroad features
	Regulated_Use	Regulated use areas
	Road_Segment	Roads
	Soil	Soil classifications
	Species_Location	Species boundaries
	Spot_Elevation	Spot elevations
	Support_Structure	Utility support features
	Surficial_Geology	Geologic classifications
	Transport_Junction	Transportation junctions
	Transport_NavAid	Transportation navigation aids
	Transport_Station_Port	Transportation ports
	Transport_Structure	Transportation structures
	Utility_Line	Utility line features
	Utility_Point	Utility point features
	Vegetation	Vegetation coverage types
	Waterbody	Water bodies

Feature Data Set	Feature Class	Description
	Watershed	Watersheds
	Waterway	Waterways
	Zoning	Zoning classification areas
Stand-Alone Object Class(es)		
	Contact_Info	Contact information for operations
	Dataset_Detail	Source and metadata information
	Document_Reference	Document reference for operations
	Geographic_Name	Geographic place names table
	Incident_Detail	Details table for incidents
	Infrastructure_Detail	Details table for infrastructure
	Natural_Event_Detail	Details table for natural events
	Operations_Detail	Details table for operations
	Street_Name	Primary and alternate names for road features
	DigitalElevationModel	Digital elevation models
	Hillshade	Hillshades
	OrthoImagery	Orthophotography
	TopographicMap	Topographic map scans
	BARGC_REFERENCE	Reference layer for the BAR-GC area

5.0 Attribute Domains

This section presents specifications for geodatabase attribute domains. Attribute domains in ArcGIS geodatabases are of two types: coded value and range. For the BAR-GC pilot, only coded value domains were utilized.

Coded value domains specify a valid set of values for an attribute. The coded value domain includes both the actual value that is stored in the database and a more intuitive description of what the coded value represents. Table 5-1 presents a listing of the domains used for the BAR-GC pilot geodatabase. For a complete diagram of the domains and their associated attributes, review the HTML file BAR-GC Data Dictionary that accompanies the digital delivery of this document.

Table 5-1
BAR-GC Geodatabase Domain List

Domain Name	Domain Type
Access	Coded Value Domain
Address_Type	Coded Value Domain
Air_Incident_Type	Coded Value Domain
Air_Surface	Coded Value Domain
CARDINALITY	Coded Value Domain
Civil_Incident_Type	Coded Value Domain
COMMODITY_TYPE	Coded Value Domain
COUNTY_DOMAIN	Coded Value Domain
COVER_TYPE	Coded Value Domain
Criminal_Incident_Type	Coded Value Domain
DAffiliationType	Coded Value Domain
Damage_Infrastructure	Coded Value Domain
Damage_Operations	Coded Value Domain
EMERGENCY_MANAGEMENT_AREA	Coded Value Domain
Fire_Incident_Type	Coded Value Domain
FTYPE_ACTIVITY	Coded Value Domain
FTYPE_ADMINISTRATIVE_AREA	Coded Value Domain
FTYPE_AIRPORT	Coded Value Domain
FTYPE_AIRPORT_SURFACE	Coded Value Domain
FTYPE_ASSET	Coded Value Domain
FTYPE_BUILDING	Coded Value Domain
FTYPE_BULLETIN_BOARD	Coded Value Domain

Domain Name	Domain Type
FTYPE_COMMODITY_TYPES	Coded Value Domain
FTYPE_EMERGENCY_ROUTE	Coded Value Domain
FTYPE_ENVIRONMENTAL_HAZARD	Coded Value Domain
FTYPE_ENVIRONMENTAL_MONITORING	Coded Value Domain
FTYPE_FACILITY_AREA	Coded Value Domain
FTYPE_HAZARD_AREA	Coded Value Domain
FTYPE_LANDMARK_AREA	Coded Value Domain
FTYPE_LANDMARK_POINT	Coded Value Domain
FTYPE_MILITARY_INSTALLATION	Coded Value Domain
FTYPE_NOISE_AREA	Coded Value Domain
FTYPE_RAIL_FEATURE	Coded Value Domain
FTYPE_REGULATED_USE	Coded Value Domain
FTYPE_TRANSPORT_JUNCTION	Coded Value Domain
FTYPE_TRANSPORT_STATION_PORT	Coded Value Domain
FTYPE_TRANSPORT_STRUCTURE	Coded Value Domain
FTYPE_TransportNavAid	Coded Value Domain
FTYPE_UTILITY_LINE	Coded Value Domain
FTYPE_UTILITY_POINT	Coded Value Domain
FTYPE_VEGETATION	Coded Value Domain
FTYPE_WATERWAY	Coded Value Domain
Geologic	Coded Value Domain
Hazardous_Material_Incident	Coded Value Domain
Hydro-Meteorological	Coded Value Domain
Inf_Agriculture_and_Food	Coded Value Domain
Inf_Banking_Finance	Coded Value Domain
Inf_Commercial	Coded Value Domain
Inf_Educational_Facilities	Coded Value Domain
Inf_Energy_Facilities	Coded Value Domain
Inf_Government_Site	Coded Value Domain
Inf_Military	Coded Value Domain
Inf_Postal_Service	Coded Value Domain
Inf_Public_Venues	Coded Value Domain
Inf_Special_Needs	Coded Value Domain
Inf_Telecommunications	Coded Value Domain
Inf_Transportation	Coded Value Domain
Inf_Water_Supply	Coded Value Domain
Infestation	Coded Value Domain

Domain Name	Domain Type
LBCS_TYPE	Coded Value Domain
Marine_Incident_Type	Coded Value Domain
Op_Emergency_Medical	Coded Value Domain
Op_Emergency_Operation	Coded Value Domain
Op_Law_Enforcement	Coded Value Domain
Op_Sensor_Operation	Coded Value Domain
OWNER_CLASS	Coded Value Domain
Rail_Incident_Type	Coded Value Domain
Rail_Type	Coded Value Domain
RAIL_USAGE	Coded Value Domain
RCLASSIFICATION_ROAD_SEGMENT	Coded Value Domain
Sec_Classification	Coded Value Domain
Source_Data_Security	Coded Value Domain
SPECIES_TYPE	Coded Value Domain
STATUS	Coded Value Domain
Vehicle_Incident_Type	Coded Value Domain
ZONE_CODE	Coded Value Domain

Appendix A

BAR-GC Layer List and Schema Report

Appendix A—BAR-GC Layer List and Schema Report

The following provides a list of the layers that are included in the final delivery of the BAR-GC HLS pilot geodatabase. The geometry type of each layer is also reflected. While this list identifies a logical grouping for layers, the geodatabase itself is made up of two feature data sets: Homeland_Security_Operations and Local_Base_Data. For a more detailed listing of the final BAR-GC pilot geodatabase, see the accompanying digital document, BAR2_RevisedMDB_reporter.htm.

Homeland Security_BAR-GC GeoDB Model

HLS Data Categories	HLS Foundation Layers	Feature Type
Homeland_Security_Operations	**ObjectClass Name**	**Geometry**
	Incident	Point
	Infrastructure	Point
	Natural_Event	Point
	Operations	Point
Local_Base_Data		
Environmental	**ObjectClass Name**	**Geometry**
	Environmental_Hazard	Polygon
	Environmental_Monitoring	Point
	Floodplain	Polygon
	Groundwater_Aquifer	Polygon
	Groundwater_Recharge	Polygon
	Groundwater_Well	Point
	Hazard_Area	Polygon
	Land_Cover	Polygon
	LandUse_LBCS1000	Polygon
	LandUse_LCSC10	Polygon
	Noise_Area	Polygon
	Noise_Contour	Polyline
	Open_Space	Polygon
	Soil	Polygon
	Species_Location	Polygon
	Surficial_Geology	Polygon
	Vegetation	Polygon

HLS Data Categories	HLS Foundation Layers	Feature Type
Transportation	**ObjectClass Name**	**Geometry**
	Activity	Polygon
	Address_Point	Point
	Airport	Point
	Airport_Surface	Polygon
	RailFeature	Polyline
	Road_Segment	Polyline
	Transport_Junction	Point
	Transport_NavAid	Point
	Transport_Station_Port	Polygon
	Transport_Structure	Polyline
	Waterway	Polyline
Basemap	**ObjectClass Name**	**Geometry**
	Elevation_Contour	Polyline
	HydroEdge (Hydro Framework)	Polyline
	HydroJunction (Hydro Framework)	Point
	Landmark_Point	Point
	Landmark_Area	Polygon
	Military_Installation	Polygon
	Parcel (Cadastral)	Polygon
	Spot_Elevation	Point
	Waterbody (Hydro Framework)	Polygon
	Watershed (Hydro Framework)	Polygon
Utilities	**ObjectClass Name**	**Geometry**
	Facility_Area	Polygon
	Support_Structure	Point
	Utility_Line	Polyline
	Utility_Point	Point
Government Units and Boundaries	**ObjectClass Name**	**Geometry**
	Administrative_Area	Polygon
	Census_Block	Polygon
	Census_Block_Group	Polygon
	Census_Tract	Polygon

HLS Data Categories	HLS Foundation Layers	Feature Type
	Community	Polygon
	County_Or_Municipality	Polygon
	Neighborhood	Polygon
	Regulated_Use (Cadastral)	Polygon
	Zoning (Cadastral)	Polygon
Emergency Management and Buildings	**ObjectClass Name**	**Geometry**
	Asset	Point
	Building_Access	Point
	Building_Complex	Polygon
	Building_DDD	Polygon
	Building_Footprint	Polygon
	Bulletin_Board	Point
	Emergency_Management_Area	Polygon
	Emergency_Route	Polyline
Stand-Alone Object Class(es)	**ObjectClass Name**	**Geometry**
	Contact_Info	Table
	Dataset_Detail	Table
	Document_Reference	Table
	Geographic_Name	Table
	Incident_Detail	Table
	Infrastructure_Detail	Table
	Natural_Event_Detail	Table
	Operations_Detail	Table
	Street_Name	Table
	Hillshade	Raster Catalog
	OrthoImagery	Raster Catalog
	TopographicMap	Raster Catalog
	DigitalElevationModel	Raster Data Set
	BARGC_REFERENCE	Polygon

Appendix B

General ESRI Geodatabase Concepts

Appendix B—General ESRI Geodatabase Concepts

The geodatabase model defines a generic model for geographic information. This generic model can be used to define and work with a wide variety of different user or application specific models. By defining and implementing a wide variety of behavior on a generic geographic model, the geodatabase provides a robust platform for the definition of a variety of user data models.

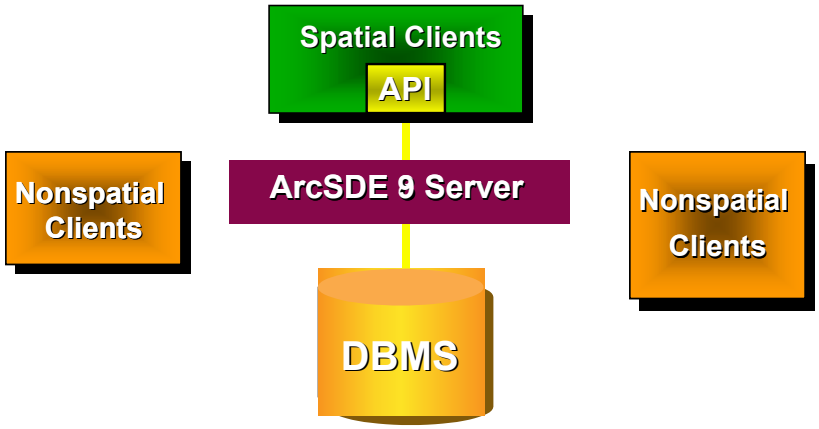
The geodatabase model supports an object-oriented vector data model. In a model, entities are represented as objects with properties, behaviors, and relationships. Support for a variety of different geographic object types is built into the system. These object types include simple objects, geographic features (objects with location), network features (objects with geometric integration with other features), annotation features, and other more specialized feature types. The model allows you to define relationships between objects as well as rules for maintaining the referential integrity between objects.

Geodatabases organize geographic data into a hierarchy of data objects. These data objects are stored in feature classes, object classes, and feature data sets. An object class is a table in the geodatabase that stores nonspatial data. A feature class is a collection of features with the same type of geometry and the same attributes.

Geodatabase and ArcSDE

The ArcInfo geodatabase model is implemented on standard relational databases with the ArcSDE application server. ArcSDE defines an open interface to database systems for ArcInfo users. It allows ArcInfo to manage geographic information on a variety of database platforms including DB2, Oracle, SQL Server, and others.

ArcSDE Architecture



In the ArcInfo system, ArcSDE 9 layers and tables are treated as geodatabase feature classes and tables such that they can be displayed, queried, and edited. However, these layers and tables participate in other functionalities such as validation rules, relationship classes, and so on.

ArcSDE 9 layers and tables, as defined here, were created with applications other than those that use the ArcInfo common object model (COM) components to manage the database. These include ArcSDE layers that were created with the ArcSDE administration tools and existing ArcSDE layers from an ArcSDE 3 database. All of these layers and tables can be displayed and queried in ArcInfo 9, and if versioned, they can also be edited.

For the layers to participate in the more advanced functions of the geodatabase, they must be registered as feature classes and object classes (tables) with the geodatabase system tables. Once they are registered with the geodatabase, they can have subtypes, default values, and validation rules and participate in relationship classes.

Geodatabase Data Sets

There are three general types of geographic data models: vector, raster, and triangulation. In the geodatabase, three types of geographic data sets implement these data models: the feature, raster, and TIN data sets.

Feature Data Sets

A feature data set is a collection of feature classes that share the same spatial reference. Feature classes that store simple features can be organized either inside or outside a feature

data set. Simple feature classes that are outside a feature data set are called stand-alone feature classes. Feature classes that store topological features must be contained within a feature data set to ensure a common spatial reference.

Spatial Reference

The spatial reference for a feature class describes its coordinate system (for example, Clarke_1866_Transverse_Mercator), its spatial domain, and its precision. The spatial domain is best described as the allowable coordinate range for *x,y* coordinates, measure (*m*) values, and *z* values. The precision describes the number of system units per one unit of measure. A spatial reference with a precision of one will store integer values, while a precision of 1,000 will store three decimal places. Once the spatial reference for a feature data set or stand-alone feature class has been set, only the coordinate system can be modified. The spatial domain is fixed.

All feature classes in a feature data set share the same spatial reference. The spatial reference is an important part of geodatabase design because its spatial domain describes the maximum spatial extent to which the data can increase. Care must be used in choosing appropriate *x,y* coordinates and *m* and *z* values. For example, if a feature data set is created with minimum *x* and *y* values of zero and precision of 1,000, none of the features in the feature data set can have *x* or *y* values that are less than zero, and all *m* values will be stored to three decimal places. Because the spatial domain for a feature class or feature data set cannot be changed, if the required *x,y* coordinates, *m* value, or *z* value ranges for your database change, the data would have to be reloaded into feature classes with a spatial reference that accommodates the new value range.

Spatial Index Grid Size

Each feature class has a spatial index that is automatically generated and maintained by the ArcInfo system. The spatial index grid is a two-dimensional grid system that spans a feature class such as the locator grid you might find on a common road map. It is used to quickly locate features in a data set that might match the criteria of a spatial search. The spatial index is calculated based on a grid size that is provided when the feature class is created.

For most data, only a single grid size is required. Because feature size is an important factor in determining an optimum grid size, data that contains features of different sizes may require additional grid sizes so larger features can be queried faster. Feature classes may have up to three grid sizes. Each grid size must be at least three times the previous grid size.

Shared Editing Within Feature Data Sets

ArcMap supports shared editing of feature classes within feature data sets. Using the concept of a topological association, editing a boundary or vertex shared by two or more features updates the shape of each of those features.

Before editing data that has topological associations, the data must be topologically integrated so that all features with parts that should be shared are shared. Integration of coincident features within a feature data set can be achieved using automatic integration in ArcMap. When a feature data set is integrated, all feature classes in the feature data set are integrated.

Feature Classes

A feature class is a collection of features with the same type of geometry: point, line, or polygon. The two categories of feature classes are simple and topological.

Simple feature classes contain points, lines, polygons, or annotation without any topological associations among them. That is, points in one feature class may be coincident with, but distinct from, the endpoints of lines in another feature class. These features can be edited independently of each other.

Topological feature classes are bound within a graph, which is an object that binds a set of feature classes that comprise an integrated topological unit. ArcInfo 8 introduces the first type of graph in a geodatabase—a geometric network.

Feature Geometries

A feature can be created with one of these types of geometries: point, multipoint, polyline, and polygon. An envelope is a geometry that describes the spatial range of feature geometries.

Points and Multipoints

Points are zero-dimensional geometries. They have an *x,y* coordinate with optional altitude (*z*), measure (*m*), and point IDs. Points are used to represent small features such as wells and survey points.

Multipoints are unordered collections of points. Multipoint features represent a set of points with a common set of attributes, such as a set of wells, that form a single unit.

Polylines

A polyline is an ordered collection of paths that can be disjointed or connected. Polylines are used to represent the geometry of all linear features. Polylines are used for roads, rivers, and contours. Simple linear features are represented by polylines with one path. Complex linear features, such as routes, are represented by polylines with many paths.

Polygon

A polygon is a collection of rings that is partially ordered by its containment relationship. A segment consists of a start and endpoint and a function defining a curve between the points.

Polygons are used to represent the geometry of all area features. Simple area features are represented by polygons with one ring. When rings are nested, they alternate between interior rings and island rings. Rings in a polygon can be disjointed, but they cannot overlap.

Envelope

An envelope represents the spatial extent of features. An envelope is a rectangle that spans the minimum and maximum coordinates of a geometry. An envelope also records the range of *z* and *m* values for a geometry. The sides of an envelope are parallel to a coordinate system.

All geometries have envelopes. Envelopes are used in ArcInfo to enable rapid display and spatial selection of features.

Subtypes

A subtype is a lightweight classification of features (or objects) within a feature (or object) class. The key motivation for using subtypes is performance. A geodatabase with one or two dozen feature classes will perform better than a geodatabase with many dozen feature classes.

Subtypes let you control specific behavior for a set of features in a feature class through attribute rules, default values, connectivity rules, and relationship rules. Whenever possible, your first preference should be to use subtypes to differentiate groups of related features.

It is necessary to split groups of related features into distinct feature classes. Following are some reasons why:

- When each group of related features requires distinct custom behavior
- When the set of feature attributes is different (all features in a feature class have the same set of attributes)
- When you require different access privileges for each group of features
- When some features are to be accessed through versions and some are not

Attribute Data Types

The geodatabase, like other data models, has the ability to store various types of attribute data. Often, different data models interpret differently the length of entry the various data types can store. The table below summarizes the geodatabase field types and their associated storage properties.

Geodatabase Field Types

Geodatabase Field Type	Item Width
Long integer	4
Text	1–255
Date	8
Float	4
Double	8
Short integer	1–4
Long integer	5–9
Double	10–16
Float	1–9
Double	10–16

Attribute Domains

Attribute domains are used to constrain the values allowed in any particular attribute for a table, feature class, or subtype. Each feature class or table has a set of attribute domains that

applies to different attributes and/or subtypes. These attribute domains can be shared across feature classes and tables in a geodatabase.

There are two different types of attribute domains: range and coded value. Each domain has a name, a description, and a specific attribute type to which it can apply.

A range domain specifies a valid range of values for a numeric attribute. A coded value domain can apply to any type of attribute—text, numeric, date, and so on. Coded value domains specify a valid set of values for an attribute. The coded value domain includes both the actual value that is stored in the database and a more user-friendly description of what that value actually means.

When editing feature classes and tables, you can enforce attribute domain rules by validating individual or sets of objects. Attribute domains do not have a property that allows or disallows null values in an associated field. When a table or feature class is created in a geodatabase, each field has a property that indicates whether or not null values are permissible values. The database itself will not permit null values to be inserted into columns that do not support them. Therefore, all domains treat null values as valid values.

Split and Merge Policies

When editing data, a single feature is split into two features, or two separate features are combined, or merged, into a single feature. For example, a land parcel may be split into two separate land parcels because of rezoning. Similar zoning changes may require two adjacent parcels to be merged into a single parcel.

While the result of these types of editing operations on the feature geometry is easily predictable, the effects on the attribute values are not. The behavior of an attribute value when a feature is split is controlled by its split policy. When two features are merged, an attribute's value is controlled by its merge policy.

Each attribute domain has both a split policy and a merge policy. When a feature is split or merged, the ArcInfo system looks to these policies to determine what values the resulting features have for a particular attribute.

An attribute for any given table, feature class, or subtype can have one of three split policies that controls the value of an attribute in the output object.

- Default value: The attribute of the two resulting features takes on the default value for the attribute of the given feature class or subtype.

- Duplicate: The attribute of the two resulting features takes on a copy of the original object's attribute value.

- Geometry ratio: The attribute of resulting features is a ratio of the original feature's value. The ratio is based on the ratio in which the original geometry is divided. If the geometry is divided equally, each new feature attribute gets half of the value of the original object attribute. Geometry ratio policies only apply to domains for numeric field types.

Annotation Feature Classes

Annotation in the geodatabase is stored in special feature classes called annotation classes. Unlike points, lines, and polygons, which are stored as ESRI simple features, annotation is stored as ESRI annotation features. As with other composite relationships, the origin feature controls the destination feature. Therefore, when the origin feature is moved or rotated, the linked annotation moves or rotates with it.

Like other feature classes in the geodatabase, all features in an annotation class have a geographic location, and attributes can be inside either a feature data set or a stand-alone annotation class. Each annotation feature has its own symbology including font, color, and so on. The annotation need not be text but can include shapes such as boxes and arrows.

There are two kinds of annotation in the geodatabase: feature linked and nonfeature linked. Nonfeature-linked annotation is geographically placed text strings that are not associated with features in the geodatabase. An example of nonfeature-linked annotation is the text on a map for a mountain range. No specific feature represents the mountain range, but it is an area you would want to mark.

Feature-linked annotation is associated with a specific feature in another feature class in the geodatabase. The text in feature-linked annotation reflects the value of a field or fields from the feature that it is linked to. The annotation feature class participates in a composite relationship with the feature class that it is annotating. The annotation feature class is the destination class in the relationship, while the feature class it is annotating is the origin class. This means the feature controls the location and lifetime of the annotation.

Object Classes

An object class is a table in a geodatabase with which you can associate behavior. Object classes keep descriptive information about objects that are related to geographic features but are not features on a map.

An object class represents a set of objects that share the same type of object such as a house, lake, or customer. The behavior of an object class is implemented with a behavior class that is stored in a DLL in conformance with the COM architecture.

ArcInfo provides a hierarchy of object classes ready for use. These are object, feature, simple junction feature, complex junction feature, simple edge feature, and complex edge feature. The ArcInfo object classes include a number of predefined fields such as object identifiers and geometries. These fields define the required properties of the objects in these classes. The data modeler can add additional custom fields.

Each of these object classes implements a set of interfaces. Each interface contains a set of related methods for actions such as storing, editing, drawing, querying, and validating objects.

Relationship Classes

A relationship class is a table that stores relationships between features or objects in two feature classes or tables. Relationships model dependencies between objects. Relationships can exist between spatial objects (features in feature classes), nonspatial objects (rows in a table), or spatial and nonspatial objects. While spatial objects are stored in the geodatabase in feature classes and nonspatial objects are stored in tables, relationships are stored in relationship classes.

Like any association, relationships have particular characteristics. One important characteristic is the notion of cardinality. Cardinality describes how many objects of Type A are related to an object of Type B. In general, relationships can have one-to-one, one-to-many, many-to-one, and many-to-many cardinalities. Certain types of relationships support certain cardinalities, and you can control cardinalities for any relationship class when you define relationship rules.

A relationship between two objects is maintained through attribute values for key fields. Relationship classes can also have attributes. Any relationship class that has attributes must be stored as a table in the database and have a pair of foreign keys, each referencing the origin and destination classes of the relationship class. In this case, each relationship is stored as a row in the relationship classes table.

Topology Class

A topology class is a geodatabase feature with which you have the ability to enforce planar topological associations between features within a feature class or between feature classes. A topology class stores various rules that define the class and the features that are involved in these rules.

A topology, like other geodatabase classes, has properties that define it. These characteristics include name, cluster tolerance, feature classes involved and their relative priority ranking, rules between feature classes, errors, and exceptions.

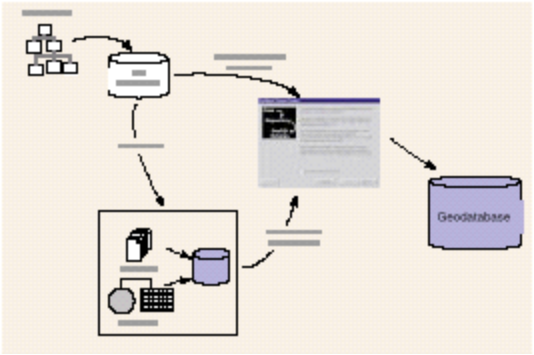
Geodatabase Modeling With UML

Geodatabase design can be greatly facilitated by the use of computer-aided software engineering (CASE) tools, such as UML, and the CASE tools subsystem of ArcInfo.

UML can be used to design schema for ESRI simple and network objects as well as to create custom objects. Most aspects of geodatabase creation can be modeled using UML. For example, UML will support the design of feature classes and tables including their attribute fields, subtypes, attribute domains, relationship classes, connectivity rules, and so on. Custom objects that inherit behavior from existing ESRI data objects can be designed. You can use C++ to override and add behaviors to the behavior inherited from the ESRI objects.

Two general strategies exist for using UML and CASE tools to design and create a geodatabase. The first strategy involves using UML to define the schema for the geodatabase, generating that schema, and then populating the schema with data.

Geodatabase Design Strategies

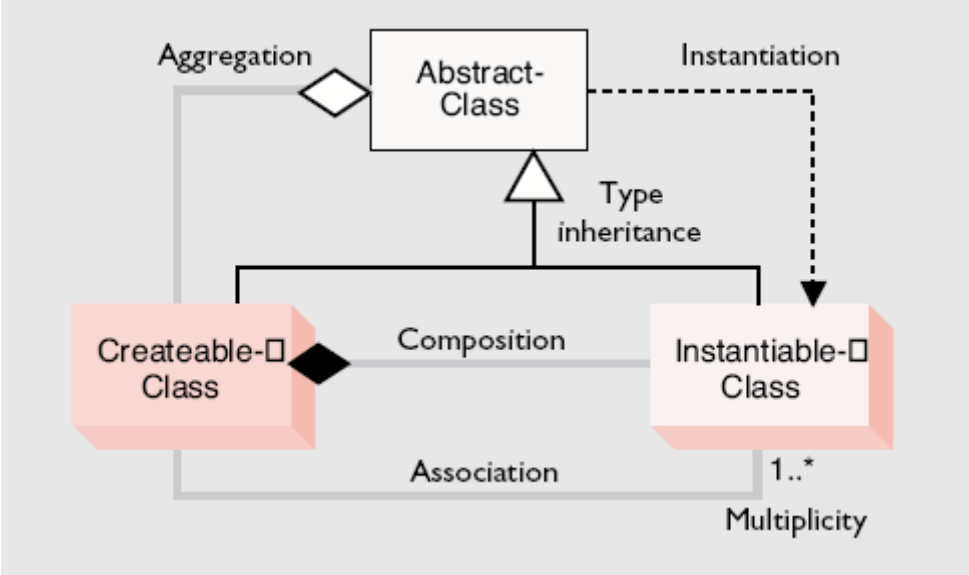


There are two strategies for using UML and CASE tools to create schema for your geodatabase. You can either generate the whole schema using the UML model, or you can import data and then apply your UML design to existing data in your geodatabase.

The second strategy takes the opposite approach. It involves creating schema by importing existing data into a geodatabase, building geometric networks, and then using CASE tools to apply the UML model to the existing data.

A combination of these two strategies can be used if your UML model describes a larger schema than defined when the existing data was imported. Once the schema has been created using one or both of these strategies, it can be modified by modifying your UML model and reapplying the model to the geodatabase schema using the Schema Creation wizard.

Object Model Overview



Object model diagrams are an important supplement to the information you receive in object browsers. The development environment (Visual Basic, etc.) lists all of the many classes and members but does not hint at the structure of those classes.

Appendix C

ArcGIS HLS Best Practices Data Model

Appendix C—ArcGIS HLS Best Practices Data Model

This appendix provides additional reference documentation related to the HLS best practices data model.

Note: Along with this Geodatabase Design document, digital versions of the following documents were also included in the original ZIP file delivery to BAR-GC.

- ArcGIS Local HLS Model (Word document)
- HLS Analysis Model Foundation (GIF image)
- HLS Analysis Model Operations (GIF image)

Appendix B

Homeland Security Data Model Diagrams

Homeland Security Data Model

Level 1 Critical Infrastructure Features with Level 2 Attributes (in Blue)

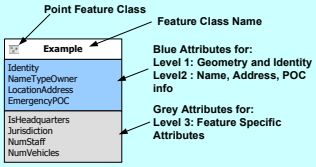
Diagram Key

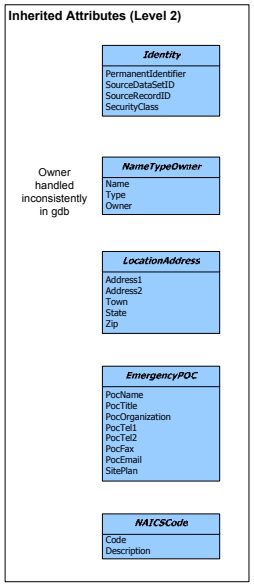
Design team recommendation is to capture the features listed in this diagram as the priority Critical Infrastructure Layers. This is a subset of the original HSIP layers and is intended as a reasonable, simple subset of HSIP.

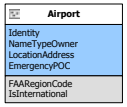
This list is based on experience with Boston, Wash DC, and BARGC pilot experience and reflects current professional recommendations for State systems.

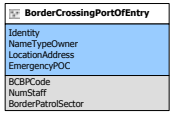
In the diagrams, the feature-specific or grey attribute boxes are considered "level 3" and are not mandatory.

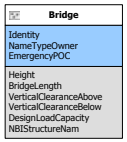
Additional features and attributes can be added from the longer list of ~90 layers as desired by implementation teams.



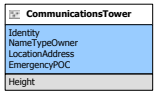


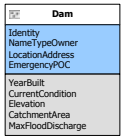


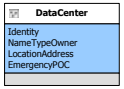


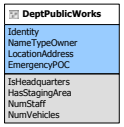


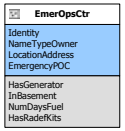
Bridge Point or Line?

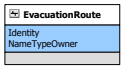




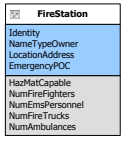


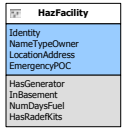




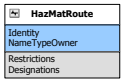


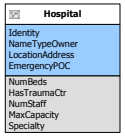
Needs POC?

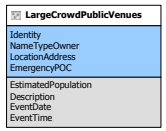




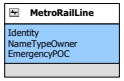
Change to HazMatFacility?

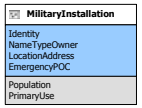


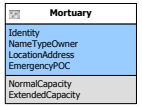


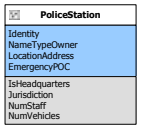


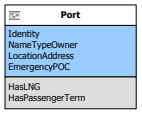
Date and Time?

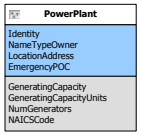


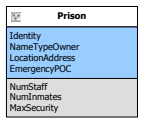




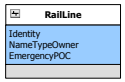


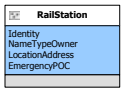


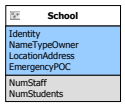


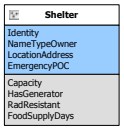


SecurityLevel?

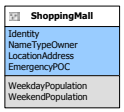


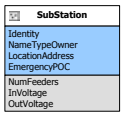


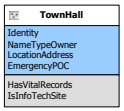




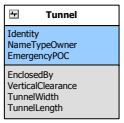
"Haz" in gdb

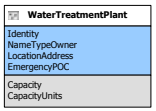






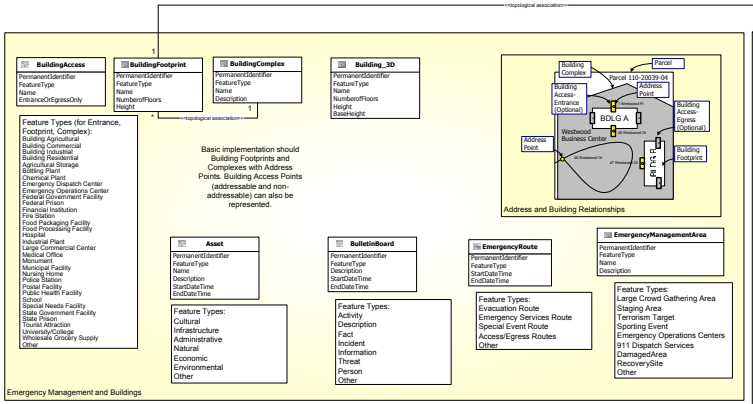
Not city hall? Seems like this is an odd choice...

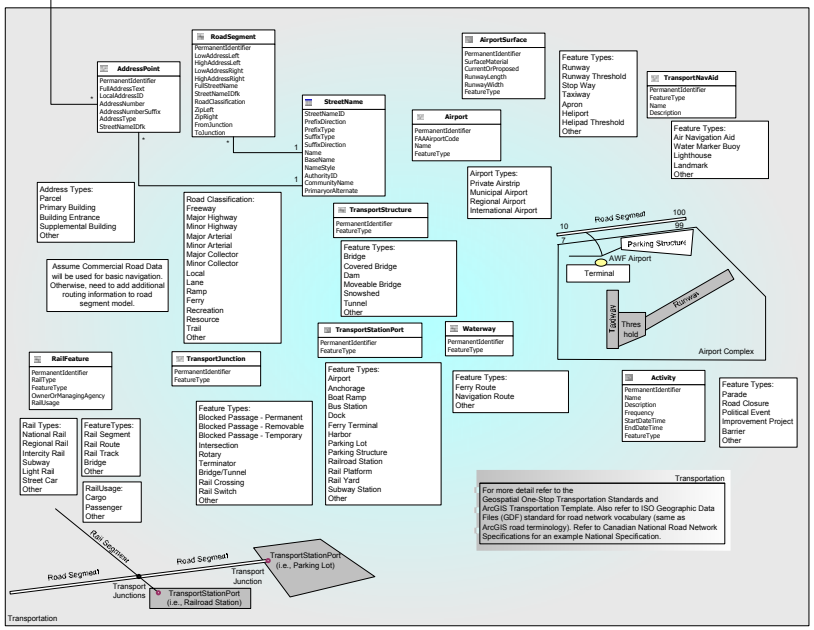


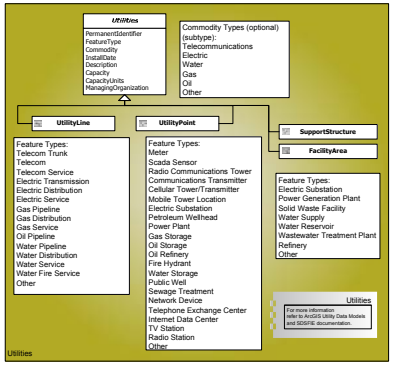


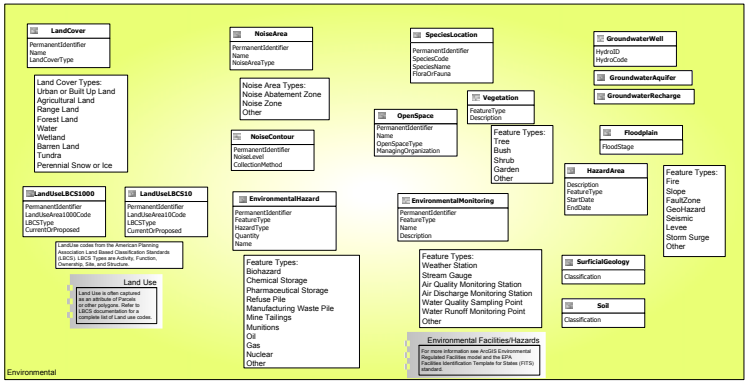
Homeland Security Data Model

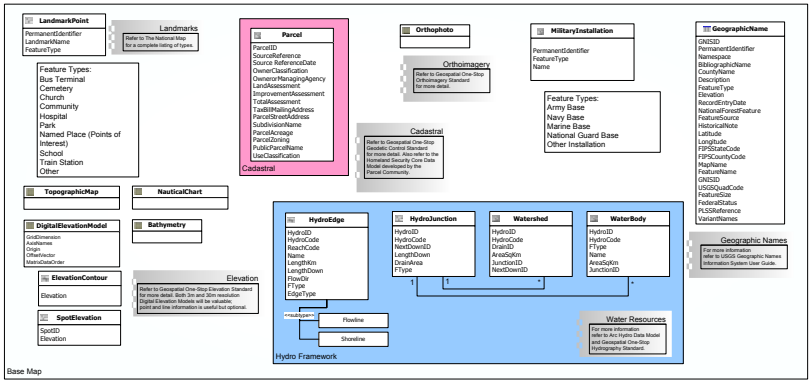
A Publication/Data Exchange Model for Local Foundation Data

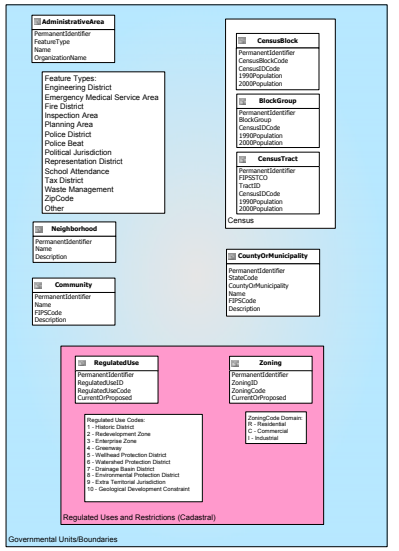








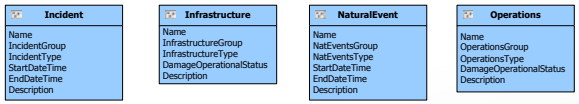




Selected References	
ESRI Data Model Templates	http://support.esri.com/datamodels
ANSI Geospatial Standards	http://www.geo-one-stop.gov/Standards/index.html
SDSFIE Standards	https://tbc.wes.army.mil/products/TSSDS-TSFMS/tssds/html/
Canadian Road Specification	http://www.geobase.ca
The National Map Viewer	http://nmviewogc.cr.usgs.gov/viewer.htm
USGS Geographic Names Documentation	http://geonames.usgs.gov/gnis_users_guide_descripdbs.html
HSIP Report (Government Access Only)	http://www.fgdc.gov/i-Team/library/reports.html
USDA NRCS Soils Data/Documentation	http://soildatamart.nrcs.usda.gov/
EPA Regulated Facilities, Chemical, and Industrial Classification	http://oaspub.epa.gov/edr/
Applied Geographics Inc. NOAA Boston Preparedness Pilot	http://www.appgeo.com/clients/NOAA_HomelandSecurity/
National Cadastral Website	http://www.nationalcad.org

Homeland Security Data Model

HLS Operations Data Model Content



Feature Class	Subtype	Domain
Incident		
	1) Civil Disturbance Incident	
		2) Civil Demonstrations
		3) Civil Displaced Population
		4) Civil Rioting
	5) Criminal Activity Incident	
		6) Bomb Threat
		7) Bomb
		8) Bomb Explosion
		9) Looting
		10) Poisoning
		11) Shooting
	12) Fire Incident	
		13) Hot Spot
		14) Non-Residential Fire
		15) Origin
		16) Residential Fire
		17) School Fire
		18) Smoke
		19) Special Needs Fire
		20) Wild Fire
	21) Hazardous Material Incident	
		22) Chemical Agents
		23) Corrosive Material
		24) Hazardous When Wet
		25) Explosive
		26) Flammable Gas
		27) Flammable Liquid
		28) Flammable Solid
		29) Non-Flammable Gas
		30) Organic Peroxides
		31) Oxidizers
		32) Radioactive Material
		33) Spontaneously Combustible
		34) Toxic Gas
		35) Toxic and Infectious
		36) Unexploded Ordnance
	37) Air Incident	
		38) Air Accident
		39) Air Hijacking
	40) Marine Incident	
		41) Marine Accident
		42) Marine Hijacking
	43) Rail Incident	
		44) Rail Accident
		45) Rail Hijacking
	46) Vehicle Incident	
		47) Vehicle Accident
		48) Vehicle Hijacking

Feature Class	Subtype	Domain
Infrastructure		
	1) Agriculture and Food Infrastructure	
		2) Agricultural Laboratories
		3) Animal Feedlots
		4) Commercial Food Distribution Center
		5) Farms Ranches
		6) Food Production Center
		7) Food Retail
		8) Grain Storage
	9) Banking Finance and Insurance Infrastructure	
		10) ATMs
		11) Banks
		12) Bullion Storage
		13) Federal Reserve Banks
		14) Financial Exchanges
		15) Financial Services Other
	16) Commercial Infrastructure	
		17) Chemical Plant
		18) Firearm Manufacturers
		19) Firearm Retailers
		20) Hazardous Material Production
		21) Hazardous Material Storage
		22) Industrial Site
		23) Landfill
		24) Pharmaceutical Manufacturer
		25) Superfund Sites NPL
		26) Toxic Release Inventory
	27) Educational Facilities Infrastructure	
		28) College University
		29) Schools
	30) Energy Facilities Infrastructure	
		31) Generation Stations
		32) Natural Gas Facilities
		33) Nuclear Facilities
		34) Petroleum Facilities
		35) Propane Facilities
	36) Government Site Infrastructure	36) Government Site Infrastructure
	37) Military Infrastructure	
		38) Military Armory
		39) Military Base
	40) Postal Service Infrastructure	
		41) Postal Distribution Center
		42) Post Office
	43) Public Venues Infrastructure	
		44) Enclosed Facility
		45) Open Facility
		46) Recreational Area
		47) Religious Institution
	48) Special Needs Infrastructure	
		49) Adult Day Care
		50) Child Day Care
		51) Elder Care
	52) Telecommunications Infrastructure	
		53) Telecommunications Facility
		54) Telecommunications Tower
	55) Transportation Infrastructure	
		56) Air Traffic Control Facility
		57) Airport
		58) Bridge
		59) Bus Station
		60) Ferry Terminal
		61) Helicopter Landing Site
		62) Lock
		63) Maintenance Facility
		64) Port
		65) Rail Station
		66) Rest Stop
		67) Ship Anchorage
		68) Toll Facility
		69) Traffic Control Point
		70) Traffic Inspection Facility
		71) Tunnel

Feature Class	Subtype	Domain
Natural Event	1) Geologic	
		2) After Shock
		3) Avalanche
		4) Earth Quake Epicenter
		5) Landslide
		6) Subsidence
		7) Volcanic Eruption
		8) Volcanic Threat
	9) Hydro-Meteorological	
		10) Drizzle
		11) Drought
		12) Flood
		13) Fog
		14) Hail
		15) Inversion
		16) Rain
		17) Sand Dust Storm
		18) Snow
		19) Thunder Storm
		20) Tornado
		21) Tropical Cyclone
		22) Tsunami
	23) Infestation	
		24) Bird Infestation
		25) Insect Infestation
		26) Microbial Infestation
		27) Reptile Infestation
		28) Rodent Infestation

Domain
Damage Infrastructure
3) Level 1 (Normal)
4) Level 2 (at capacity or closed)
5) Level 3 (partial damage)
6) Level 4 (total damage)

Domain
Damage Operations
3) Level 1 (Normal)
4) Level 2 (at capacity or closed)
5) Level 3 (partial damage)
6) Level 4 (total damage)

Feature Class	Subtype	Domain
Operations		
	1) Emergency Medical Operation	
		2) EMT Station Locations
		3) Ambulance
		4) Medical Evacuation Helicopter Station
		5) Health Department Facility
		6) Hospital
		7) Hospital Ship
		8) Medical Facilities Out Patient
		9) Morgue
		10) Pharmacies
		11) Triage
	12) Emergency Operation	
		13) Emergency Collection Evacuation Point
		14) Emergency Incident Command Center
		15) Emergency Operations Center
		16) Emergency Public Information Center
		17) Emergency Shelters
		18) Emergency Staging Areas
		19) Emergency Teams
		20) Emergency Water Distribution Center
		21) Emergency Food Distribution Centers
		22) Fire Suppression Operation
		23) Fire Hydrant
		24) Other Water Supply Location
		25) Fire Station
	26) Law Enforcement Operation	
		27) ATF
		28) Border Patrol
		29) Customs Service
		30) DEA
		31) DOJ
		32) FBI
		33) Police
		34) Prison
		35) Secret Service
		36) TSA
		37) US Coast Guard
		38) US Marshals Service
	39) Sensor Operation	
		40) Biological Sensor
		41) Chemical Sensor
		42) Intrusion Sensor
		43) Nuclear Sensor
		44) Radiological Sensor

Appendix C

Bay Area Regional GIS Council Pilot Project Data Migration Report

NGA Homeland Security Enterprise GIS Support Project

BAR-GC Pilot Project Data Migration Report

Final

ESRI Database Services
December 2004

Prepared for:

Brad Lucas, Contracting Officer's Representative
National Geospatial-Intelligence Agency
4600 Sangamore Road
Bethesda, Maryland 20816-5003
Tel.: 301-287-6572

Prepared by: ESRI
Version: Final
Last Modification Date: December 31, 2004

Table of Contents

Section **Page**

Section **Page**

Section **Page**

Change History		
Date	**Description**	**Person(s) Responsible**
Dec. 20, 2004	Initial Draft	Matt Lott/Jan Mares
Dec. 31, 2004	Final	Charlie Wells

1.0 Introduction

1.1 Purpose of This Document

The purpose of this document is to describe the processes, procedures, and tools, associated with the creation of the homeland security (HLS) database for the Bay Area Regional GIS Council (BAR-GC) pilot project, and to provide findings and recommendations resulting from the pilot project that are applicable to the HLS data requirements of NGA and its federal partners.

1.2 Intended Audience of This Document

This document is intended for the NGA Homeland Security Enterprise GIS Support Project team, NGA partners, and management and technical staff associated with BAR-GC member organizations.

1.3 Document Overview

This document is divided into the following sections:

- **Section 1.0—Introduction:** This section provides an introduction to the BAR-GC Pilot Project Data Migration Report.
- **Section 2.0—BAR-GC Pilot Project Data Migration Task Summary:** This section gives a brief summary of the overall Extract, Transform, and Load (ETL) processes used during the pilot project.
- **Section 3.0—Administrative Issues:** This section details the administrative considerations related to the pilot data migration process.
- **Section 4.0—BAR-GC Pilot Data Model:** This section reflects the data model that was populated during the data migration process.
- **Section 5.0—High-Level BAR-GC Pilot Project Data Migration Work Flow:** This section describes the data migration methodology that was implemented during the BAR-GC Pilot Project.
- **Section 6.0—Findings and Recommendations:** This section describes ESRI's findings and recommendations from the BAR-GC Pilot.

- **Section 7.0—Pilot Deliverables:** This section lists the deliverables of the BAR-GC Pilot.

- **Appendix A—BAR-GC HLS Data Model Layer List:** This appendix provides a listing of the layers that are in the final delivery of the BAR-GC homeland security (HLS) Pilot geodatabase.

- **Appendix B—SourceList Database Column Definitions:** This appendix provides definitions for the columns contained in the SourceList Access database.

- **Appendix C—TargetsExcelMapping File Column Definitions:** This appendix provides definitions for the columns contained in the TargetsExcelMapping file.

- **Appendix D—Listing of Source Data Received:** This appendix provides a list of all source data received from the BAR-GC as well as the geometry and format of each source data received.

1.4 Assumptions, Dependencies, and Constraints

The following assumptions and dependencies are applicable to this document.

- This document is assumed to be for reporting purposes only. Although it contains a detailed process for the data migration task it should not be considered a step-by-step training manual.

- The digital documents referenced in this document are included in the ZIP file accompanying this report. The entire ZIP file should be transferred along with any digital copies of the document.

1.5 Referenced Documents

This section identifies documents that are referenced in this BAR-GC Data Migration Pilot Project Report. These include project documents that are deliverables as well as external documents referenced by this report.

1.5.1 Project Documents

Document Name	Date	Document Owner
Project Homeland BAR-GC National Map Pilot Project Management Plan (Word document)	11/08/2004	ESRI
TargetsExcelMapping.xls (Excel file)	12/2004	ESRI
10-7-04_BAR-GC_SourceList.mdb (Access file)	12/2004	ESRI
Geodatabase Design for BAR-GC Pilot Project. (Word document)	12/20/2004	ESRI
BAR2_RevisedMDB_reporter.htm (HTML document)	12/2004	ESRI
BAR-GC ETL Process Presentation (PowerPoint document)	12/2004	ESRI
BAR-GC Geodatabase Data Dictionary (HTML document)	12/2004	ESRI

1.5.2 External Documents

Document Name	Date	Document Owner
System Design Strategies White Paper (PDF document)	7/04	Dave Peters
ArcGIS Data Interoperability Tutorial (PDF document)	10/8/2004	ESRI

Online information and training materials for ESRI's ArcGIS Data Interoperability extension are available at http://www.safe.com/esri/arcgis-data-interoperability/course/index.htm.

2.0 BAR-GC Pilot Project Data Migration Task Overview

This section provides an overview of the BAR-GC Pilot Project extract, transfer, and load (ETL) data migration tasks.

2.1 Project Overview

Historically, when the National Geospatial-Intelligence Agency (NGA) needed to support a local event or incident around the country, an arduous data gathering process was undertaken for each area. This involved NGA requesting data from the United States Geological Survey (USGS), which would task one of its local liaisons to work with the local jurisdictions to collect the information. The information was collected on a onetime basis and was transferred through the most expedient method, usually CDs, Data Brick, or file transfer protocol (FTP).

When an unplanned event occurred, such as the recent death of President Reagan, the same process was followed but within a compressed time frame. The BAR-GC Pilot Project was designed as a first step in the development of long-term data sharing agreements between local jurisdictions and the USGS, which acts as a broker for the NGA, and other federal homeland security (HLS) and homeland defense (HLD) agencies. The project will also establish the technological means for local jurisdictions to move the data to USGS.

The BAR-GC Data Migration Pilot Project provides the technical means of moving the heterogeneous local data sets to the local HLS Best Practices Data Model that can be loaded by USGS into *The National Map.* The pilot was accomplished using ESRI COTS technology including ArcGIS 9, the Data Interoperability Extension, and several data loading productivity tools within ESRI's Production Line Tool Sets (PLTS). The goal of the pilot was to use this combination of tools to load BAR-GC-provided data into the BAR-GC ArcSDE geodatabase model and document process methodologies, findings, and recommendations.

3.0 Administrative Considerations

As is widely recognized by NGA, its partners, and ESRI's Project Homeland team, technical issues do not necessarily represent the most significant obstacle to the acquisition of data by the federal government from local providers. Administrative and legal considerations can be at least as significant as technical considerations in an environment where data, some of it potentially sensitive, is exchanged between different levels of government. This section describes how such administrative considerations were approached in the BAR-GC pilot project.

The implementation of data sharing arrangements has been an ongoing process in the BAR-GC pilot project. BAR-GC's participation as the first pilot project in Project Homeland was at least partially predicated on a consensus agreement that local data resources would be shared among the BAR-GC members as well as with the federal government, subject to appropriate conditions to be documented in a series of memoranda of understanding (MOU).

Three distinct aspects of data sharing were encountered:

- Initial data acquisition by ESRI to exercise the data model and develop the extract, transform, and load (ETL) methodology
- Data sharing among BAR-GC member organizations once the assembled database is made available on the BAR-GC servers
- Data sharing with NGA and its federal partners involved in homeland security activities

These topics are discussed in the following sections.

3.1 Initial Data Acquisition by ESRI

Soon after being approached by NGA and USGS to participate in Project Homeland, the BAR-GC membership began a process of developing an MOU that would formalize each member's willingness to share its data with other members and with the federal government.

In anticipation that the MOU process might take some time, and in an effort to expedite the pilot project, ESRI proposed to acquire data from each of the participating members, pending completion of the MOUs, for use in exercising the data model and implementing the ETL methodology. At each agency's discretion, ESRI offered to provide written assurances that the data would remain in ESRI's possession until each agency entered into a formal MOU allowing further dissemination of the data.

This approach was proposed to the BAR-GC leadership, and, in turn, to the member organizations.

The response by the data providers was variable. Several BAR-GC entities requested letters from ESRI, while others did not. In no case was this initial data acquisition process more complicated than ESRI having to provide a brief letter stating the uses that would be made of the data.

3.2 Data Sharing Among BAR-GC Members

A key aspect of the BAR-GC pilot project is that the data provided by individual local governments will be mapped and loaded into a regional homeland security database for use by BAR-GC members and emergency response agencies in the Bay Area. The regional database will be deployed on a pair of servers situated in two locations in the Bay Area. Once deployed, the data will be available to all participating member organizations.

Data sharing among BAR-GC member organizations will be governed by the MOUs that are being put in place. Although there is a large degree of cooperation among the members on a normal basis, the MOUs are important as a means of establishing data use guidelines, data security levels, constraints and restrictions on data use, and data disclaimers.

The MOU content is being established jointly by the member organizations with the assistance of a consultant (GIS Consultants). The data sharing arrangements for sharing among BAR-GC members will clearly be tailored to the requirements of the local organizations.

3.3 Data Sharing with NGA and Federal Partners

The primary interest of NGA and its federal partners in the BAR-GC pilot project is as a first step toward establishing a process for acquiring current data from local providers in a timely and efficient manner. NGA and its partners require access rights to the data that will allow them to use the data in a variety of environments, including on classified networks. It is essential, therefore, that NGA is granted formal approval by the data providers to use the data as needed without making commitments to the providers regarding specific uses of the data or *quid pro quo* arrangements concerning data being returned from NGA to the providers.

The MOU being developed by the BAR-GC membership is intended to address the issues associated with sharing data with NGA and its federal partners. As such, the MOU content must be agreeable to both the BAR-GC providers and to NGA.

More generally, the policy issues associated with these data sharing agreements between locals and the federal government must be further evaluated. The facilitation role played by USGS in establishing data sharing arrangements between the USGS and local providers is an essential part of the process. The USGS has encountered and dealt with many of the issues involved through its Mapping Partnerships with states. Subsequent pilots will provide additional opportunities to explore the issues involved with data sharing between local providers and the federal government.

4.0 BAR-GC Pilot Data Model

The data model used for the BAR-GC pilot effort represents an enhanced version of the local HLS Best Practices data model. Below is a brief summary of the development of the BAR-GC data model. For the complete design schema of the as-built BAR-GC data model, please reference the accompanying design document entitled Geodatabase Design for BAR-GC Pilot Project.

4.1 HLS Best Practices Data Model

The local HLS Best Practices data model was the starting point for the BAR-GC geodatabase design. The basic approach for this HLS Best Practices model was to start with ESRI's current local government data model template and add additional homeland security content. The most significant additions are in the areas of emergency management, transportation, environmental, and utilities content. In most other homeland security modeling efforts, the work has revolved around documenting a set of "layers," rather than detailed data information needs. The goal of the HLS Best Practices model is to capture these layers as feature types and organize the feature classes and attributes in alignment with local government best practices.

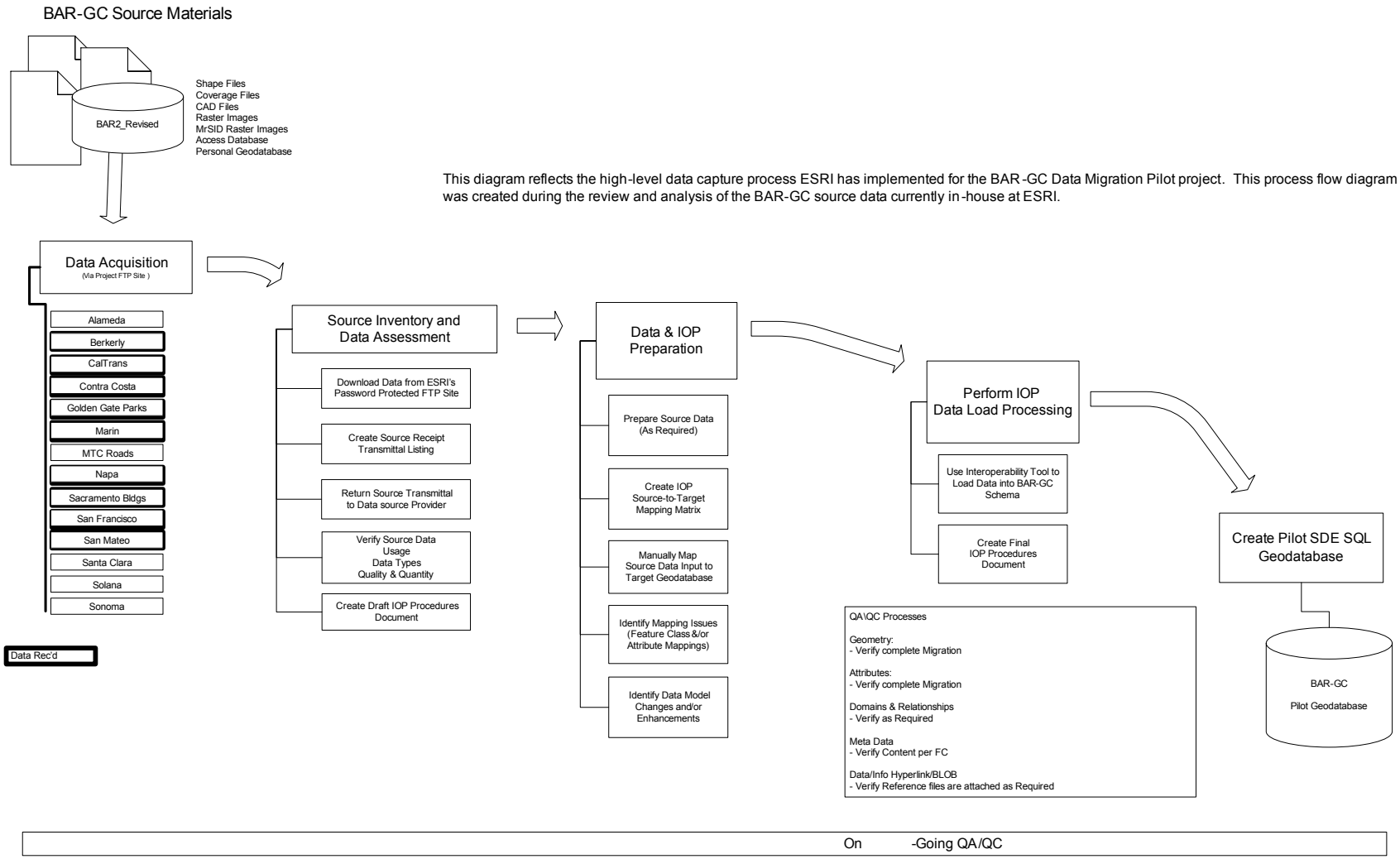
4.2 BAR-GC Data Model

ESRI worked with the Bay Area Regional Homeland Security Data Server (BAR-HSDS) working group to develop a BAR-GC pilot data model based on the ESRI HLS Best Practices Model template. For a listing of the feature datasets and feature classes, please see Appendix A. The pilot database model will provide the foundation for storing data from many different input sources from local providers, at the regional/state level, and will also be utilized as a single source at the federal level. This effort will coincide with a BAR-GC goal of developing crosswalks between various data providers and data layers.

5.0 High-Level BAR-GC Pilot Project Data Migration Work Flow

The ETL process used to convert BAR-GC source data into the HLS data model was developed using an iterative approach throughout the lifecycle of the BAR-GC pilot project. The following diagram, Figure 5-1, reflects the ETL process methodology at the conclusion of the pilot project. The specific steps in the work flow are described in the sections following the diagram.

Figure 5-1



5.1 Data Acquisition

Members of the BAR-GC community used a password-protected FTP site to transfer data to ESRI. Since data was being provided by different groups at different times, this FTP site was checked by the ESRI project manager on a daily basis during the data acquisition period, and new data postings were downloaded to ESRI production servers as received. A receipt transmittal was created, and a digital copy was e-mailed to the data provider. This transmittal served as a record of the data received as well as a quality control (QC) checklist to ensure all data received had been accounted for during the data load process.

5.2 Data Inventory and Assessment

After the data was downloaded to the ESRI server, information associated with new data source layers was entered in a SourceList Access Database (see Appendix B). The populated fields in this database aided in the mapping of input source data attributes to targets within the pilot geodatabase. It was updated during the pilot project to reflect changes made during the mapping process as source data was further examined in ArcCatalog and ArcMap. The Access database also automatically generated the Source_DatasetIDs as layers were added to the spreadsheet. These Unique ID's and associated information are loaded in the ArcSDE geodatabase table Dataset_Detail where they can be used to "join" the table with individual feature classes to track each feature back to its original input source.

5.3 Data Preparation

The following processes were used for organizing and preprocessing the input source data.

5.3.1 Set Up Directory Structure

Individual directory folders were created for each set of input source data received from the BAR-GC participants. The folders were then used as storage locations for source data, the final generated target data, and the custom data export tools (Data Interoperability Extension mapping files) created via the ETL process. The following format was used to organize data:

First, a data source folder was created and named using the date the data was received and the name of the data it refers to: <date mm-dd-yy>_<source name>, for example: **10-7-04_Marin.** This folder also contains all the support folders for the source data and for the target data.

Under this folder, another folder was created called "Original." All the original feature types are copied into this folder.

Also created under the data source folder was the target folder. It is named "Processed." Since multiple data sets were being processed at once, a personal geodatabase for each provider was used to house the features in the model before final data load into the final BAR-GC pilot ArcSDE. To do this, copy "BAR2_RevisedMDB.mdb" and its contents from the Pilot_Data_Models folder to the data source folder and rename it using the following scheme: BAR2_RevisedMDB-<source name>.mdb, for example: BAR2_RevisedMDB-Marin.mdb.

5.3.2 Convert Input Source to Shapefile

The majority of source data received from the BAR-GC participants was in an ESRI shapefile format. To ensure the most successful load of this input source data into the BAR-GC ArcSDE geodatabase, ESRI converted all remaining input source to the ESRI shapefile format prior to processing. Using the ESRI shapefile format produced the most stable processing results when using the ArcGIS Data Interoperability Extension.

5.3.3 Review and Repair Geometry Errors

ESRI used the Check Geometry and Repair Geometry ArcToolbox tools as well as additional PLTS tools on all input source data. The Check and Repair Geometry tools were run in a batch process on the BAR-GC input data sets to ensure proper data structure and geometry prior to the final IOP processing into the pilot geodatabase. Review Section 6.5.2.3 for a description of the ESRI PLTS tools.

5.3.4 Review Projection

The projections of all incoming source data were reviewed and defined based on its metadata information. For the pilot ArcSDE geodatabase, GCS_WGS_1984 is the standard projection used for all layers. All input source data was reprojected to match this projection as necessary.

5.4 Input Source to Geodatabase Feature Class Target Mapping

The information that was entered into the Access database was used to identify which target feature classes would be utilized during the data migration pilot. During data processing, the Access database is updated, as necessary, to reflect any reevaluation of source data previously received from the BAR-GC. During this step, the corresponding target pages from the Excel spreadsheet, TargetsExcelMapping (see Appendix C), were identified and printed. These pages contain the attributes of target layers and any associated relate tables and domain values. Attributes and values from the source data were examined in ArcCatalog and ArcMap to determine the best mapping scheme to load the source attributes into the

target attributes. Linking the source attributes to the target attributes was first completed using the Excel mapping files, so that during QC, changes could be easily made. The completed source-to-target mapping printouts were later used as reference to create the custom data export tool within the workbench functionality of ESRI's ArcGIS Data Interoperability extension.

5.5 Data Interoperability Extension Processing

The following subtasks describe the use of the ArcGIS Data Interoperability extension to perform the mapping to, and population of, the BAR-GC pilot geodatabase data sets and feature classes.

5.5.1 Convert Target Layers to Shapefile

Before creating the Custom Data Export Tools for each ETL, the target feature classes were converted to the ESRI shapefile format. This conversion allowed the analyst to import the target schema as a destination type in the ArcGIS Data Interoperability Extension without being concerned with the problems found when importing the schema of a Geodatabase. For a more in depth look at these findings see section 6.5.2.2.

5.5.2 Create Custom Data Export Tools

The completed Excel mapping files were put through the QC process and were then used as a reference for creating the new custom data export tools/mapping files using the Data Interoperability Extension in ArcCatalog. The Custom Export Tools were then renamed to match the description of the target feature classes.

In many cases, "transformers" were required to generate the proper attribution for linking input source data to the targets. Commonly used transformers were the Universal Unique Identifier (UUID) Generator, Attribute Filter, Concatenator, and the Attribute Creator. In other cases, adding fields to the input source data and populating them manually in ArcMap helped simplify the schema in the Interoperability tool. This worked very well for source data layers that have more than one target or when a data source layer needed to be split into multiple targets. Fields were created in ArcCatalog and edited in ArcMap. If it was determined that adding new fields would help the migration of data, then this task was completed before the creation of a custom data export tool.

5.5.3 Run the Custom Data Export Tool

Once the Custom Data Export tool mapping file was completed, it was run on the input source data to populate the target feature classes. The results were verified by checking the target layers' geometry and attribution. The attribute fields were examined to ensure they

were populated correctly and that all records were accounted for. The geometry was spot-checked visually for consistency with the other layers. If errors were detected, the tool was modified and rerun after the target had been cleaned of the bad data.

5.5.4 Input Source Data Post-Processing

To verify that a successful load of input source data into the BAR-GC ArcSDE geodatabase, the Check Geometry and Repair Geometry ArcToolbox tools were run again on all data sets. These checks alerted the GIS technician to remaining problems that may be encountered before loading final data into the pilot ArcSDE geodatabase and allowed the problems to be investigated and/or documented as needed.

5.6 ArcSDE Geodatabase Loading, QC, and Delivery

The following tasks must be completed before proceeding with the final geodatabase load:

- The database server is operational with the ArcSDE software.
- An ArcSDE instance has been created and started.
- The ArcSDE instance has been populated with the empty geodatabase schema.

5.6.1 Load the Vector Data Into the ArcSDE Geodatabase

The ArcCatalog Simple Data Loader was used to load vector data into ArcSDE.

5.6.2 Load the Tabular Data Into the ArcSDE Geodatabase

The ArcCatalog Simple Data Loader was used to load the tabular data into ArcSDE.

5.6.3 Load Any Raster Data Into the ArcSDE Geodatabase

The ArcCatalog raster loading tools were used to load raster data into the appropriate raster catalog in ArcSDE.

5.6.4 QC and Finalize the Load

ArcCatalog and ArcMap were used to compare the loaded features with the Excel mapping spreadsheet and source receipt transmittals to ensure all features loaded as required. After final QC, the BAR-GC pilot ArcSDE geodatabase was compressed and analyzed using ArcCatalog to ensure it performs as efficiently as possible when queried and displayed by ArcGIS.

5.6.5 ArcSDE Data Export

All data layers were exported using the ArcSDE Export command for delivery to the BAR-GC HSDS sites.

5.6.6 ArcSDE Data Import

All data layers were imported using the ArcSDE Import command for implementation on the BAR-GC HSDS sites.

6.0 Findings and Recommendations

The following findings and recommendations were identified by ESRI as a result of the data analysis and ETL process for the BAR-GC pilot project. They reflect the best practices recommendations for future ETL data migration processing.

6.1 Administrative Findings

The following sections describe the findings as related to the data sharing agreements and missing metadata.

6.1.1 Data Providers

Source data was not received from all of the BAR-GC members. See Appendix D for a list of the data sets that were received. Many of the critical infrastructure features were not delivered for the pilot data migration task, either because they are not maintained by the member organizations or because they were not provided. Without these important layers in the database, the usability of the data in response to a homeland security incident may be limited.

6.1.2 Metadata

If the source data providers did not deliver metadata to ESRI along with the data, ESRI analysts had a very difficult time deciphering the attribute mappings.

6.1.3 Administrative Recommendations

ESRI recommends that all future data sharing agreements identify "critical" layers that are required from each data provider. The sharing of these critical layers could be made a requirement in order to access the final data warehouse.

ESRI recommends a long-range plan to develop an input source to the target template that can be distributed to local data providers to be populated and delivered along with their original input source data. This template may include a mapping diagram of the geodatabase schema and attribute-level metadata requirements, for this will greatly expedite the ETL process.

6.2 Data Model Findings

The following sections describe the changes made to, and proposed changes to, the BAR-GC data model. This information is also contained in Section 2.5 of the BAR-GC Geodatabase Design document. It is included here in the event that the document is not accessible.

6.2.1 General Data Model Findings

This section outlines the data model findings.

- For the pilot, the data security attribute of all features was populated with a default value of "Public Domain." Further review of data security alternatives is ongoing as part of Project Homeland and the results of this review will contribute to a proposed approach to data security later in the project.

- The Data Interoperability UUID Generator was used to assign a unique ID to all geodatabase features. The Global Unique Identifier (GUID) field was not populated during this pilot. The UUID and/or GUID population process will require further analysis before being implemented as part of the ETL process.

- Attributes that could not be mapped to targets were not processed during the pilot migration. Future alternatives include storing the entire input source data inserted as Geography Markup Language into a binary large object (BLOB) field to be retrieved and viewed when and if required. This will make available the complete content of the original data including attributes that are not migrated into the HLS model. The BLOB field is currently not supported in the Data Interoperability Extension. Further research and analysis must occur before this process is implemented.

- During the pilot it was realized that several input data source types could be mapped to both the HLS Operations and the Local Base Data Feature Data Sets. As a proof of concept for the pilot analysis, only a select sample of input source data was mapped to both feature datasets (i.e., airport, bus terminal, and state buildings, to list a few). Therefore, a decision must be made to finalize whether or not these input source data types are mapped to one or both feature datasets. One option is that the HLS Operations Feature Dataset only contains mission-specific data, which represents a subset of these features, which can also be contained in the Local Base Data Feature Dataset, such as bridges, government installations, incident locations, etc.

    - North American Industry Classification System (NAICS) codes were not used in Pilot mapping example, as none of the incoming source data had an existing code that could be readily mapped to the NAICS code.

- Existing input source data valid value and subtype domain table relationships were evaluated for inclusion into the BAR-GC model. Where appropriate, these existing relationships were maintained after the features were mapped into the new geodatabase structure. If a domain and/or relate table was not needed because of attribution not being mapped, or similar domains did not exit, they were not mapped to the target geodatabase, they will only remain in the source and will not appear in the target. Future considerations to populate the BLOB field with the entire input source data content may address this migration issue.

- Raster catalogs and data sets were created using ArcCatalog, since these cannot be modeled using CASE tools.

- Metadata documents related to the input source data received, such as PDFs and Word documents, are not included in the ArcSDE geodatabase. The Dataset_Detail table within the pilot geodatabase provides a link to locate these ancillary files.

- The elevation attribute in Elevation_Contours and Spot_Elevation feature classes was populated as is from the original input source without any unit conversion.

- The Contact_Info relate table was not populated when no usable information was provided.

6.2.2 Changes to BAR-GC Data Model During Pilot Project

This section outlines the changes made to the data model during the course of the BAR-GC pilot project.

- Changed the entire model to reflect a GCS_WGS_1984 projection.

- Added the polygon feature class "Vegetation" to the data model schema.

- Added the polygon feature class "Landmark_Area" to the data model schema.

- Added a domain for Address_Point tied to Address_Type, ADDRESS_TYPE (Parcel, Primary Building, Building Entrance, Supplemental Building, Other).

- Added a domain for Airport_Surface tied to Feature_Type, FTYPE_AIRPORT_SURFACE (Runway, Taxiway, Apron\Hardstand, Other).

- Added a domain SOURCE_DATA_SECURITY to control access and use constraints.

- Added a domain for Landmark_Area tied to the Feature_Type, FTYPE_LANDMARK_AREA (Cemetery, Park, Named Place, Reserve, Golf Course, Other).

- Expanded domain FTYPE_HAZARD_AREA to accommodate Federal Emergency Management Agency flood zones.

- Expanded domain OWNER_CLASS to accommodate county-owned parcels.

- Expanded domain FTYPE_MILITARY_INSTALLATION to accommodate air force and coast guard bases.

- Expanded domain FTYPE_AIRPORT to accommodate military airfields.

- Changed tagged value LENGTH = 255, Parcel_Address, Tax_Address attributes of Parcel to accommodate full address from source.

- Changed tagged value LENGTH = 50, of Operations_ID attribute in Feature ClassContact_Info to accommodate Permanent_ID.

- Changed tagged value LENGTH = 50, Classification attributes of Soil to accommodate full coding from source.

- Changed tagged value LENGTH = 50, Zone_Code, attributes of Zoning to accommodate full coding from source. Dropped Zone_Code domain to limited for the various sources.

- Changed tagged value LENGTH = 50, Zone_Code, attributes of Parcel to accommodate full coding from source. Dropped Zone_Code domain to limited for the various sources.

- Changed tagged value LENGTH = 255, Name, attribute of Building_Footprint to accommodate full names from source.

- Combined Access and Use_Constraints fields in Dataset_Detail table to make a field AccessandUse_Constraints. This field was tagged to the Source_Data_Security Domain.

- Changed Name attribute of Street_Name table to Primary_Name, added Alternate_Name1, Alternate_Name2, Alternate_Name3. Set tagged value Length = 150 for all. Dropped the attribute Primary_Or_Alternate.

6.2.3 Data Model Recommendations

This section outlines the recommended data model changes for future pilot projects. These changes were not incorporated into the BAR-GC pilot due to the need to freeze the state of the pilot data model.

- Change tagged value LENGTH = 255, Name, attribute of Building_Complex to accommodate full names from source.
- Change tagged value LENGTH = 255, Name, attribute of Landmark_Area to accommodate full names from source.
- Change tagged value LENGTH = 255, Landmark_Name, attribute of Landmark_Point to accommodate full names from source.
- Change domain for SEC_CLASSIFICATION to be short_integer, SEC_CLASSIFICATION (0—Unknown, 1—TS, 2—S, 3—Confidential, 4—Restricted, 5—UNCLASSIFIED, 999—Other).
- Correct misspelling of "Pharmaceutical" in Infrastructure Domain tables.
- ESRI recommends adding a polyline feature class to contain route features that typically overlay roads, such as bus routes, truck routes, and bike routes.

6.3 Source Data Findings

The following sections provide the source data findings from the BAR-GC pilot project.

6.3.1 Original Input Source Data

- If duplicate feature classes were delivered from a source provider, only the most recently dated or more completely populated source data was used and mapped.
- For a list of feature classes that were not mapped during the pilot, please see the Comments field contained in the Access SourceList database.

6.3.1.1 Napa County

- Provided a shapefile named Census_2000 that was actually parcel information. In referencing the Napa County GIS metadata on the Web, it appears that this was a typo in the data uploaded to the FTP, so the actual Census_2000 file is missing.

- Two shapefiles were provided for runways, one from Caltrans and one from Napa County. Both of these shapefiles were polyline geometry. To accommodate the model, which specifies runways as polygons, these linear runways were buffered to create polygons.

- Airport classes had to be determined from the airport name.

6.3.1.2 Caltrans

- As noted above, runways provided by Caltrans were converted from lines to polygons through a buffering process.

- State and local bridges (st_br_reg.shp and loc_br_reg.shp) were mapped to Infrastructure and Transport_Junction.

- The rail and lt_rail_reg shapefiles from CalTrans required extra analysis and mapping to properly identify and populate the geodatabase target feature classes.

- Contact_Info relate table was not populated when no usable information existed.

- There was no target feature class for truck routes.

- bus_term_reg was mapped to Infrastructure, Landmark_Point, respectively.

6.3.1.3 Berkeley

- Many of the feature classes received had little attribution that could be mapped into the new geodatabase environment. Metadata received with the Berkeley data indicates that these were created to support the creation of aesthetically pleasing maps, not feature description.

- Feature Class U_baybridges_hypso_20040625 seems to have duplicate features that are slightly offset from each other.

- Feature Class U_baybridges_hydro_dlg_24k could not be loaded. It also contains both elevation contours and hydro lines. Attribution was not available to separate with tools.

6.3.1.4 San Mateo County

- Only the first matched record of related Parcel Attribute tables was mapped to the ArcSDE geodatabase.

6.3.1.5 Contra Costa County

- No metadata was delivered with the shapefiles.

6.3.1.6 San Francisco City and County

- planning_neighborhoods.shp was not loaded. This identical data was also contained in planning_neighborhoods_tiybi.shp.

6.3.1.7 Marin County

- The Mineral_Resource_Pres_Site.shp shapefile is corrupt.
- The Airports.shp shapefile is corrupt.
- Faults.shp, Debris_Flow_Source.shp, sewer_maintenance.shp, precipitation.shp, bus_route.shp, and bikeway.shp were not mapped to the model as no target feature class was available in the geodatabase.
- school.shp contains bad geometry.

6.3.1.8 GG Parks

- prsf_cadbase.shp was not mapped; it appears to be the source of shapefiles also delivered as separate layers.

6.3.1.9 Miscellaneous Data

- The st_bldngs_sp_orig.shp (Sacramento buildings) shapefile was loaded as an example of how point features could be mapped to multiple targets within both the Operations and Local Base data.

6.4 Data Migration Findings

The following sections provide the data migration findings from the BAR-GC pilot project.

6.4.1 Data Loading

During the pilot processing, it was found that using shapefiles was the most effective data format for all levels of ETL processing. At present, complex data (polygons with donut holes) from geodatabases often produced incomplete results when run through the custom Interoperability tool. In those cases, processing from a source shapefile to a target shapefile

was a necessary work-around. The idea of splitting the data layers into different layer types did not seem efficient, so each data set was processed in the same data format it was received (i.e., shapefiles).

Trying to use only transformers to prepare source data to link with target attributes proved to be too complex and cumbersome. A more productive method was to customize the source data by inserting new fields using ArcMap. This works well when dealing with data layers that split into multiple targets. A large amount of effort was spent on organizing data into the new fields, rather than making a complex linking scheme with multiple transformers.

6.4.2 Data Quality Control

Data assessment, loading, and migration quality control was an integral part of the ETL data migration pilot process. Typically, the features that could not be loaded into the ArcSDE geodatabase did not have clean topology and, therefore, could not be loaded. The Check and Repair Geometry tools in ArcToolbox were used to correct most of these problems; however, some of the geometry could only be fixed manually, which was not possible due to time and budget constraints of the BAR-GC pilot. Since the use of these tools and their effect on the data quality may change the number of features due to removing nulls and/or simplifying features, a more in-depth QC was done only when there were significant discrepancies in the number of features that were loaded versus the number of features received in the original source data sets.

6.4.3 Data Migration Recommendations

- It is recommended that future ETL processing be performed using shapefiles before loading the data into the ArcSDE geodatabase.

- ArcMap should be used to add/alter data fields as necessary to organize data before using the ArcGIS Data Interoperability Extension. It is, therefore, possible that this type of data enhancement would lead to more generic export tools that could be used repeatedly for multiple data sources. With this in mind, perhaps a more automated system could be developed that would migrate a preprepared source directly into ArcSDE.

- ESRI recommends that a data cleanup activity be implemented to ensure that the local providers submit data that has clean geometry.

6.5 Hardware and Software Findings

The following sections describe the hardware and software findings from the BAR-GC pilot project.

6.5.1 Hardware

The machine that was used for the bulk of the data processing during the BAR-GC pilot had the specifications described below. This machine worked well during the pilot and was able to process the large amounts of vector and raster input source data within acceptable time frames.

Processor: Intel Pentium 4 @ 3.0 GHz
RAM: 1 GB
HDD: 1 60 GB 7.2K RPM
OS: Windows XP Professional Service Pack 1.
Note: Windows XP Service Pack 2 was not loaded on this machine for the pilot project.

6.5.2 Software

Throughout the pilot, ESRI GIS technicians researched ways to improve the efficiency of ETL processing and experimented with different ESRI software products and functionality. The following findings resulted from this ongoing process.

6.5.2.1 ArcGIS

For the BAR-GC pilot, ArcGIS 9 with Service Pack 2 was used.

6.5.2.2 Data Interoperability Extension

A number of shortcomings were found with the Data Interoperability Extension. Below is a summarized list of the findings identified during the pilot. The Project Homeland team has provided this feedback to ESRI's development team for consideration as bug fixes or enhancements.

- The tool has trouble when mapping complex (donut) polygons from geodatabase to geodatabase if geometries are not clean.
- The tool is not able to properly project data sets as they are mapped, even if both input and output have a projection defined.
- Any field that is constrained by a domain in the geodatabase is not imported into the workbench session.
- Does not recognize and import BLOB fields in the workbench.
- Does not report when attributes are chopped off due to field size discrepancies.

- The tool is very processor and memory intensive. It worked best on machines that had at least a 2.0 GHz processor and 1 GB of RAM.

- There does not appear to be a way of changing the processing to something other than 500 features at a time.

- Cannot import or export mapping files created with SafeSoft's Feature Manipulation Engine tools.

- The `tool has problems with creating custom data transformations for features that participate in a geometric network`.

6.5.2.3 Production Line Tool Sets

Although the entire ETL process could have been completed using only core ArcGIS 9 and Data Interoperability Extension functionality, another ESRI ArcGIS software extension, the Production Line Tool Sets (PLTS) provided an even more efficient and productive batch process environment for several process work flows.

PLTS Foundation, comprising the core of PLTS, provides the essential functionality within each of the PLTS solutions for users who require database editing, production, managed quality control, cartographic tools, and work flow management but would like to implement their own business rules in data modeling, attribution, validation, and output. The online help feature of PLTS describes in detail how to create a customized solution.

PLTS Foundation provides tools that are universally usable while allowing users to define their own geodatabase model, map series, and symbology definitions. Business rules are implemented in valid value tables (VVT) and condition tables that extend rule-based attribute capture and validation and map symbology. Database driven cartography allows symbolization from multiple feature attributes and provides sophisticated, repeatable (consistent) map rendering. Through managed quality control, spatial and attribute errors are logged to a table for easy correction and revisiting.

There are a variety of compelling reasons for using PLTS in this environment. Single-click editing toolbars dramatically cut data editing time for GIS staff. Intelligent attribution tools provide true in-process, rule-based quality control. Quality assurance is verifiable using sampling and summary reporting tools. In addition, map series tools provide rule-based element placement, eliminating redundancy. Organizations in many industries can use these commonly required tools to increase efficiency through all stages of data production, quality control, and map production.

For the BAR-GC pilot, the PLTS Foundation extension provided several productivity tools to find and repair geometry errors as well as tools to automate the bulk loading of data sets into

the final ArcSDE geodatabase. And although core ESRI functionality could have been used to perform these tasks, the productivity enhancements provided by PLTS made the use of these tools an obvious choice.

6.5.3 Hardware and Software Recommendations

To efficiently process data sets covering a regional area the size of the BAR-GC, ESRI recommends that a machine which meets or exceeds the recommended configuration for ArcGIS 9, as specified at www.esri.com, be utilized. For a more in-depth look at systems specifications for various ESRI ArcGIS applications, see the Systems Design Strategies document available at www.esri.com/support.

ESRI recommends staying current with the ArcGIS service packs as they are released.

7.0 Pilot Deliverables

The following lists the deliverables for the BAR-GC ETL Data Migration Project.

- ArcSDE export files of geodatabase
- Data Interoperability Custom Export Tools
- BAR-GC ETL Pilot Data Migration Report (Dated 12/20/2004)
- BAR-GC Geodatabase Reporter data model report (BAR2_RevisedMDB_reporter.htm)
- BAR-GC Geodatabase Design Document (Geodatabase Design for BAR-GC Pilot Project (Dated 12/31/2004)
- BAR-GC Data Dictionary (BAR-GC Bar2_Revised Geodatabase Data Dictionary.htm)
- BAR-GC ETL PowerPoint Presentation (BAR-GC ETL Presentation.ppt)
- SourcetoTarget Excel file (TargetsExcelMapping.xls)
- SourceList Access database (10-7-04_BAR-GC_SourceList.mdb)

Appendix A

BAR-GC HLS Data Model Layer List

Appendix A—BAR-GC HLS Data Model Layer List

The following appendix provides a list of the layers that are in the final delivery of the BAR-GC HLS Pilot geodatabase. Also included is the geometry type of each layer. While this list identifies a logical grouping for layers, the geodatabase itself is made up of two feature data sets, Homeland_Security_Operations and Local_Base_Data, as well as stand alone object classes.

Homeland Security_BAR-GC Geodatabase Model		
HLS Data Categories	**HLS Foundation Layers**	**Feature Type**
Homeland_Security_Operations	**ObjectClass Name**	**Geometry**
	Incident	Point
	Infrastructure	Point
	Natural_Event	Point
	Operations	Point
Local_Base_Data		
Environmental	**ObjectClass Name**	**Geometry**
	Environmental_Hazard	Polygon
	Environmental_Monitoring	Point
	Floodplain	Polygon
	Groundwater_Aquifer	Polygon
	Groundwater_Recharge	Polygon
	Groundwater_Well	Point
	Hazard_Area	Polygon
	Land_Cover	Polygon
	LandUse_LBCS1000	Polygon
	LandUse_LCSC10	Polygon
	Noise_Area	Polygon
	Noise_Contour	Polyline
	Open_Space	Polygon
	Soil	Polygon

Homeland Security_BAR-GC Geodatabase Model		
HLS Data Categories	**HLS Foundation Layers**	**Feature Type**
	Species_Location	Polygon
	Surficial_Geology	Polygon
	Vegetation	Polygon
Transportation	**ObjectClass Name**	**Geometry**
	Activity	Polygon
	Address_Point	Point
	Airport	Point
	Airport_Surface	Polygon
	RailFeature	Polyline
	Road_Segment	Polyline
	Transport_Junction	Point
	Transport_NavAid	Point
	Transport_Station_Port	Polygon
	Transport_Structure	Polyline
	Waterway	Polyline
Basemap	**ObjectClass Name**	**Geometry**
	Elevation_Contour	Polyline
	HydroEdge (Hydro Framework)	Polyline
	HydroJunction (Hydro Framework)	Point
	Landmark_Point	Point
	Landmark_Area	Polygon
	Military_Installation	Polygon
	Parcel (Cadastral)	Polygon
	Spot_Elevation	Point
	Waterbody (Hydro Framework)	Polygon
	Watershed (Hydro Framework)	Polygon
Utilities	**ObjectClass Name**	**Geometry**
	Facility_Area	Polygon

Homeland Security_BAR-GC Geodatabase Model		
HLS Data Categories	**HLS Foundation Layers**	**Feature Type**
	Support_Structure	Point
	Utility_Line	Polyline
	Utility_Point	Point
Government Units & Boundaries	**ObjectClass Name**	**Geometry**
	Administrative_Area	Polygon
	Census_Block	Polygon
	Census_Block_Group	Polygon
	Census_Tract	Polygon
	Community	Polygon
	County_Or_Municipality	Polygon
	Neighborhood	Polygon
	Regulated_Use (Cadastral)	Polygon
	Zoning (Cadastral)	Polygon
Emergency Management & Buildings	**ObjectClass Name**	**Geometry**
	Asset	Point
	Building_Access	Point
	Building_Complex	Polygon
	Building_DDD	Polygon
	Building_Footprint	Polygon
	Bulletin_Board	Point
	Emergency_Management_Area	Polygon
	Emergency_Route	Polyline
Stand-Alone ObjectClass(s)	**ObjectClass Name**	**Geometry**
	Contact_Info	Table
	Dataset_Detail	Table
	Document_Reference	Table
	Geographic_Name	Table
	Incident_Detail	Table
	Infrastructure_Detail	Table

Homeland Security_BAR-GC Geodatabase Model		
HLS Data Categories	**HLS Foundation Layers**	**Feature Type**
	Natural_Event_Detail	Table
	Operations_Detail	Table
	Street_Name	Table
	Hillshade	Raster Catalog
	OrthoImagery	Raster Catalog
	TopographicMap	Raster Catalog
	DigitalElevationModel	Raster Data set
	BARGC_REFERENCE	Polygon

Appendix B

SourceList Access Database Column Definitions

Appendix B—SourceList Access Database Column Definitions

The following appendix provides definitions for the columns contained in the SourceList Access Database. The populated fields in this database aid in the mapping of input source data and attributes to their targets in the geodatabase. It can also be updated to reflect changes made during the mapping process as source data is further examined in ArcCatalog and ArcMap. Also, Access automatically generates a Source_DatasetID number as layers are added to the spreadsheet. These Source_DatasetID numbers are populated in the geodatabase and can be used to track data back to their original source, once mapped into the BAR-GC HSDS geodatabase. The Access database was provided in digital format along with the delivery of this report. Microsoft Access is required in order to access/open this file.

Source Dataset ID

Unique ID numbers automatically generated by Microsoft Access, used to track and identify feature data sets.

Provider

The name of the county, municipality, or government agency providing data.

Location

The location where provider data is digitally stored on the server.

Filename

The name of the original feature data as delivered by the provider.

Geometry

The geometry type of the feature. Polygon refers to area type data, Polyline refers to line features, and Point refers to point features.

Data Type

The form in which the original data was delivered.

Feature_Count

The total number of features.

Data_Security

The level of access and use security assigned to the data.

Metadata

The name of the metadata file delivered by the provider, e.g., XML, HTML, or PDF.

Description

A plain translation description of the feature data set.

Target (1, 2, 3, 4)

The name of the proposed target layer to which the source data will be migrated.

Comments

A column where comments or notations are collected that reflect actions or clarifications.

Appendix C

TargetsExcelMapping File Column Definitions

Appendix C—TargetsExcelMapping File Column Definitions

This following appendix provides an example of the definitions and descriptions of the information populated within the TargetsExcelMapping spreadsheet. This spreadsheet was used to perform an initial hard-copy mapping of the data, where the ESRI GIS technician investigates the required attribute mappings as well as identifies the type of transformers that will be needed to accomplish the source to target geodatabase mapping.

	A	B	C
1	**Source**	**Transformer**	**Target Field Name**
2			Shape
3			OBJECTID
4	Output from UUID Generator	UUID Generator	Permanent_Identifier
5	PD	Attribute Creator	Data_Security
6	Juris-ID		Source_FeatureID
7	Get from Access DB	Attribute Creator	Source_DatasetID
8	Area_Name		Name
9	County Sheriff Jurisdiction	Attribute Creator	Organization_Name
10	Police District	Attribute Creator - See Domain	**Feature_Type**
11			Shape_Length
12			Shape_Area
13			
14		**Domain Members**	**FTYPE_ADMINISTRATIVE_AREA**
15		**Name**	**Value**
16		Engineering District	Engineering District
17		Emergency Medical Service Area	Emergency Medical Service Area
18		Fire District	Fire District
19		Inspection Area	Inspection Area
20		Planning Area	Planning Area
21		Police District	Police District
22		Police Beat	Police Beat
23		Political Jurisdiction	Political Jurisdiction
24		Representation District	Representation District
25		School Attendance	School Attendance
26		Tax District	Tax District
27		Waste Management	Waste Management
28		Zip Code	Zip Code
29		Other	Other
30			

Source

The attribute fields from the source that will map to the target attribute fields on the right.

Transformer

The name of the transformer that will be used to manipulate the source attributes before insertion into the target attributes.

Target Field Names

The attribute fields from the target layers that will hold the manipulated source attributes.

Domain Members Name

The listing of domains as it appears in the table view of ArcMap and ArcCatalog.

(Domain Name, i.e., FTYPE_ADMINISTRATIVE_AREA) Value

The actual values that are populated for each domain. This value is inserted into the **Bold** attribute field in the target attribute list, i.e., Feature_Type.

Appendix D

Source Data Receipt Listing

Appendix D—Source Data Receipt Listing

The following table provides a list of the source data received from the BAR-GC participants and includes the geometry type and data format in which it was received. Individual transmittals can also be acquired from each of the data providers or requested from ESRI.

Provider	Filename	Geometry	DataType
Berkeley	Note: The Berkeley feature class data listed was received within three Personal Geodatabases. They were named: BAR_GC_HSDS_contrib.mdb, BAR_HSDS_pcl_pg.mdb, and City_Berkeley_CA_DFIRM.mdb.		
	U_ALA_area_li_20040106	Polyline	Feature Class
	U_ALA_area_pg_20040106	Polygon	Feature Class
	U_ALA_land_li_20040106	Polyline	Feature Class
	U_ALA_land_pg_20040113	Polygon	Feature Class
	U_ALB_area_li_20040102	Polyline	Feature Class
	U_ALB_area_pg_20040102	Polygon	Feature Class
	U_ALB_land_li_20040102	Polyline	Feature Class
	U_ALB_land_pg_20040102	Polygon	Feature Class
	U_baybridge_pcl_20040803	Polygon	Feature Class
	U_baybridge_pcl_20041006	Polygon	Feature Class
	U_baybridges_hydro_dlg_24k	Polyline	Feature Class
	U_baybridges_hypso_20040625	Polyline	Feature Class
	U_berk_centerline_20030819	Polyline	Feature Class
	U_CA_water	Polygon	Feature Class
	U_CAtr2k_land_li_2003	Polyline	Feature Class
	U_catr2k_land_pg_2003	Polygon	Feature Class
	U_CoB_area_li_20040217	Polyline	Feature Class
	U_CoB_area_pg_20040217	Polygon	Feature Class
	U_CoB_land_li_20040217	Polyline	Feature Class
	U_CoB_land_pg_20040217	Polygon	Feature Class
	U_CoB_parcel_li_20040217	Polyline	Feature Class
	U_CoB_parcel_pg_20040217	Polygon	Feature Class
	U_CoB_reclam_li_20040217	Polyline	Feature Class
	U_CoB_reclam_pg_20040217	Polygon	Feature Class
	U_EMY_area_li_20040102	Polyline	Feature Class
	U_EMY_area_pg_20040113	Polygon	Feature Class
	U_EMY_land_li_20040102	Polyline	Feature Class

Provider	Filename	Geometry	DataType
	U_EMY_land_pg_20040113	Polygon	Feature Class
	U_FEMA_Q3_pg_200405	Polygon	Feature Class
	U_morpho_20040805	Polyline	Feature Class
	U_morpho_20040805	Polyline	Feature Class
	U_nat_wtrshed_pg_20040419	Polygon	Feature Class
	U_OAK_area_li_20040105	Polyline	Feature Class
	U_OAK_area_pg_20040106	Polygon	Feature Class
	U_OAK_land_li_20040113	Polyline	Feature Class
	U_OAK_land_pg_20040113	Polygon	Feature Class
	U_PMT_land_li_20040102	Polyline	Feature Class
	U_PMT_land_pg_20040102	Polygon	Feature Class
	U_SnMateo_pcl_20040623	Polygon	Feature Class
	U_watrshd_li_20040805	Polyline	Feature Class
Caltrans			
	airp_bnd_reg.shp	Polygon	Shape
	airp_rw_reg.shp	Polyline	Shape
	airport_reg.shp	Point	Shape
	bus_term_reg.shp	Point	Shape
	cvef_reg.shp	Point	Shape
	fr_facl_reg.shp	Point	Shape
	loc_br_reg.shp	Point	Shape
	lt_rail_reg.shp	Polyline	Shape
	maint_fac_reg.shp	Point	Shape
	port_bnd_reg.shp	Polygon	Shape
	port_reg.shp	Point	Shape
	rail.shp	Polyline	Shape
	rest_reg.shp	Point	Shape
	st_br_reg.shp	Point	Shape
	tnkr_trm_reg.shp	Point	Shape
	trknet_reg.shp	Polyline	Shape
Census TIGER 2000			
	Counties_bayarea	Polygon	Geodatabase
Contra Costa			
	Centerline_10_03.shp	Polyline	Shape
	CLI_PWD_Gauges.shp	Point	Shape
	CLI_PWD_Rainfall_full.shp	Polygon	Shape
	coco_creeks_3_25_04.shp	Polyline	Shape
	Contours.shp	Polyline	Shape
	ContraCostaShoreline.shp	Polyline	Shape

Provider	Filename	Geometry	DataType
	fishmap.shp	Polyline	Shape
	PARC_APN.shp	Point	Shape
	PARC_PY.shp	Polygon	Shape
	restoration_projects.shp	Point	Shape
	watershed_shape.shp	Polygon	Shape
GG Parks			
	prsf_admin.shp	Polygon	Shape
	prsf_cadbase.shp	Polyline	Shape
	prsf_centerlines.shp	Polyline	Shape
	prsf_structures.shp	Polygon	Shape
	prsf_trails.shp	Polyline	Shape
Marin			
	Airports.shp	Point	Shape
	alquistpriolo.shp	Polygon	Shape
	bikeway.shp	Polyline	Shape
	bus_route.shp	Polyline	Shape
	Census_Blk_2000.shp	Polygon	Shape
	Census_Blk_Grp_2000.shp	Polygon	Shape
	Census_Trct_2000.shp	Polygon	Shape
	cityhall.shp	Point	Shape
	comm_college_dist.shp	Polygon	Shape
	county_svc_area.shp	Polygon	Shape
	dam_inundation.shp	Polygon	Shape
	Debris_Flow_Source.shp	Point	Shape
	elem_school_dist.shp	Polygon	Shape
	Faults.shp	Polyline	Shape
	Ferry_GGT.shp	Polyline	Shape
	fire_dist.shp	Polygon	Shape
	firestation.shp	Point	Shape
	Floodplain.shp	Polygon	Shape
	Geology_East.shp	Polygon	Shape
	Geology_West.shp	Polygon	Shape
	groundwater.shp	Polygon	Shape
	high_school_dist.shp	Polygon	Shape
	landslides.shp	Polygon	Shape
	LawEnforcement.shp	Point	Shape
	Library.shp	Point	Shape
	library_dist.shp	Polygon	Shape
	Liquefaction.shp	Polygon	Shape

Provider	Filename	Geometry	DataType
	MedicalFacility.shp	Point	Shape
	Mineral_Resource_Pres_Site.shp	Polygon	Shape
	Park_and_Ride.shp	Point	Shape
	postoffice.shp	Point	Shape
	precipitation.shp	Polyline	Shape
	public_utility_dist.shp	Polygon	Shape
	recreation_dist.shp	Polygon	Shape
	RefuseFranchises.shp	Polygon	Shape
	sanitary_dist.shp	Polygon	Shape
	SCA.shp	Polygon	Shape
	school.shp	Point	Shape
	school_trustee_area.shp	Polygon	Shape
	sewer_maintenance.shp	Polygon	Shape
	ShorelineTrusteeArea.shp	Polygon	Shape
	Soil_Expansive.shp	Polygon	Shape
	sra.shp	Polygon	Shape
	Supervisor_Dist.shp	Polygon	Shape
	TransitHubs.shp	Point	Shape
	Vegetation_2000.shp	Polygon	Shape
	water_dist.shp	Polygon	Shape
	wetland.shp	Polygon	Shape
	zipcode.shp	Polygon	Shape
Miscellaneous Data			
	st_bldngs_sp_orig.shp	Point	Shape
Napa County			
	airprt_napa_runway.shp	Polyline	Shape
	calveg.shp	Polygon	Shape
	calwater22.shp	Polygon	Shape
	cont_02_400.shp	Polyline	Shape
	countytra.shp	Polygon	Shape
	ctylmts.shp	Polygon	Shape
	elev_02_400.e00	GRID	Arc\Info Export
	fema_fldzn.shp	Polygon	Shape
	o38122e2.tif	N/A	TIFF Raster
	Parcel.shp	Polygon	Shape
	parcel_extract.shp	Polygon	Shape
	Q09_gdsrvimg.sid	N/A	MrSID Raster
	Q10_gdsrvimg.sid	N/A	MrSID Raster
	R09_gdsrvimg.sid	N/A	MrSID Raster

Provider	Filename	Geometry	DataType
	R10_gdsrvimg.sid	N/A	MrSID Raster
	roads_features.mdb	Polyline	Geodatabase
	spot_02_400.shp	Point	Shape
	srgosoil.shp	Polygon	Shape
	zoning.shp	Polygon	Shape
San Francisco			
	census2000_blk_nowater.shp	Polygon	Shape
	census2000_blkgrp_nowater.shp	Polygon	Shape
	census2000_tracts_nowater.shp	Polygon	Shape
	cityfeatures.shp	Polyline	Shape
	citylots.shp	Polygon	Shape
	dpt_bike_network.shp	Polyline	Shape
	dpt_bluezones.shp	Point	Shape
	dpt_signals.shp	Point	Shape
	dpt_speedlimits.shp	Polyline	Shape
	elect_bart_dists.shp	Polygon	Shape
	elect_congress_dists.shp	Polygon	Shape
	elect_pollplaces.shp	Point	Shape
	elect_precincts.shp	Polygon	Shape
	elect_stasmbly_dists.shp	Polygon	Shape
	elect_stsenate_dists.shp	Polygon	Shape
	elect_supervisor_dists.shp	Polygon	Shape
	fire_erd_districts.shp	Polygon	Shape
	fire_prev_districts.shp	Polygon	Shape
	phys_contours.shp	Polyline	Shape
	phys_seismic_hazard_zones.shp	Polygon	Shape
	phys_surveypts.shp	Point	Shape
	phys_waterbodies.shp	Polygon	Shape
	planning_drparcels.shp	Point	Shape
	planning_landuse.shp	Polygon	Shape
	planning_lessunits.shp	Point	Shape
	planning_neighborhoods.shp	Polygon	Shape
	planning_neighborhoods_tiybi.shp	Polygon	Shape
	planning_openspace.shp	Polygon	Shape
	planning_zoning.shp	Polygon	Shape
	port_jurisdiction.shp	Polygon	Shape
	realtor_neighborhoods.shp	Polygon	Shape
	schools_private.shp	Polygon	Shape
	schools_private_pts.shp	Point	Shape

Provider	Filename	Geometry	DataType
	schools_public.shp	Polygon	Shape
	schools_public_pts.shp	Point	Shape
	sfaddresses.shp	Point	Shape
	sfblocks.shp	Polygon	Shape
	sffacs.shp	Point	Shape
	sfha_properties.shp	Polygon	Shape
	sfislands.shp	Polygon	Shape
	sflnds.shp	Polygon	Shape
	sfoutline.shp	Polygon	Shape
	stclines_highways.shp	Polyline	Shape
	stclines_streets.shp	Polyline	Shape
	stintersections.shp	Point	Shape
	stnodes.shp	Point	Shape
	traffic.shp	Polyline	Shape
San Mateo County			
	ACTIVE_PARCELS_APN.shp	Polygon	Shape
	ASSEMBLY_DIST.shp	Polygon	Shape
	BREAKLINES.shp	Polyline	Shape
	CENSUS_BLKS.shp	Polygon	Shape
	CITY.shp	Polygon	Shape
	CONGRESSIONAL_DIST.shp	Polygon	Shape
	CONTOURS.shp	Polyline	Shape
	COUNTY_BOUNDARY.shp	Polygon	Shape
	EASEMENTS.shp	Polyline	Shape
	ELECTION_PRECINCTS.shp	Polygon	Shape
	G200.shp	Polygon	Shape
	G400.shp	Polygon	Shape
	G800.shp	Polygon	Shape
	LAKES.shp	Polygon	Shape
	LANDMARKS_AREA.shp	Polygon	Shape
	OCEAN_BAY.shp	Polygon	Shape
	PLANNING_ZONES.shp	Polygon	Shape
	RAILROAD.shp	Polyline	Shape
	ROW_LINES.shp	Polyline	Shape
	SENATE_DIST.shp	Polygon	Shape
	SPECIAL_DISTS.shp	Polygon	Shape
	STNODENAMES.shp	Point	Shape
	STREAMS.shp	Polyline	Shape
	STREETS.shp	Polyline	Shape

Provider	Filename	Geometry	DataType
	STREETS_ALIAS.shp	Polyline	Shape
	SUPERVISORS.shp	Polygon	Shape
	SWR_FLOW_ARROWS.shp	Polygon	Shape
	SWR_GRIDS.shp	Polygon	Shape
	SWR_MAINS.shp	Polyline	Shape
	SWR_NODES.shp	Polyline	Shape
	SWR_SRVC_LINES.shp	Polyline	Shape
	TABULAR DATA	N/A	Access Database
	ZIPCODES.shp	Polygon	Shape

Appendix D

Bay Area Regional HSDS ETL and Data Replication Framework

BAR-HSDS ETL and Data Replication Framework

Project: BAR-HSDS
Date: 01/31/2004
Notes: Added information raster data

The purpose of this document is to outline the BAR-HSDS Extract, Transform, and Load (ETL) and data replication procedures and to begin to brainstorm the data push framework in which to get data from initial data providers to remote data servers.

ETL Process Description

This section describes the ETL process undertaken by ESRI to migrate data contributed by Bay Area counties to the HLS geodatabase. Steps 1 to 11 describe the actual ETL of the data. Steps 12 to 14 describe the data loading procedures into SDE.

1. **Data Acquisition**
    a. Data will be provided by contributing local Bay Area municipalities as Data Sharing Agreements allow.
    b. Data will be usually acquired via FTP or snail mail.
    c. Data should be provided with adequate metadata.

2. **Data Inventory and Assessment**
    a. Inventory newly acquired data layers in SourceList Access Database. The populated fields in this database aid in the mapping of input source data and attributes to their targets in the geodatabase.
    b. Access Database automatically generates a Source_DatasetID number as layers are added to the spreadsheet. These Source_DatasetID numbers are populated in the geodatabase later on in the process and can be used to track data back to their original source, once mapped into the HLS geodatabase.

3. **Create Working Directories**
    a. Create folder for data named: <date mm-dd-yy>_<source name>, for example: 10-7-04_Marin.
    b. Under this directory, folders named “original” (for original data) and “processed” (for ETL’d data) will be created.

4. **Convert Source Data to Shapefile**
    a. Source data is converted to a shapefile using ArcCatalog or Data Interoperability extension and put in “processed” folder.

b. Shapefile most effective data format for all ETL levels.

5. Review and Repair Geometry Errors

a. Use the Check Geometry and Repair Geometry ArcToolbox tools on source shapefile to ensure proper data structure and geometry prior to processing.

6. Review Projection

a. Data is reprojected to HLS Data Model projection GCS_WGS_1984.

7. Shapefile Dump of HLS Geodatabase

a. Create folder called "target shapefiles."
b. Use PLTS GDB to SHP tool to export target shapefiles of HLS data model layers into newly created folder. PLTS tool may have option to export schema only. Alternatively, ArcCatalog can be used to create target shapefiles.
c. The target shapefiles will be populated with the source shapefile attributes with the data export tools in the following steps.

8. Source Shapefile to Target Shapefile Spreadsheet Mapping

a. Refer to SourceList Access Database to identify which target feature classes would be utilized during the data migration.
b. Create a TargetsExcelMapping spreadsheet to perform an initial hard-copy attribute mapping of the data. This aids in researching the required attribute mappings, as well as identifies the type of Data Interoperability extension transformers that will be needed, to accomplish the source to target geodatabase mapping.

9. Create Custom Data Export Tools

a. Create "mapping file" with aid of mapping spreadsheet.
    1. Target shapefile can be imported to populate target schema.
    2. Where necessary, create transformers for proper attribute mapping.
    3. Where necessary, add fields to source shapefile and populate to simplify target shapefile mapping.

10. Run Custom Data Export Tools

a. Run mapping file on the source shapefile to populate the target shapefile.
b. Verify results with ArcCatalog by checking the target shapefile's geometry and attribution. If errors exist, modify mapping file and rerun.

11. Assess Processed Data

a. Use the Check Geometry and Repair Geometry ArcToolbox tools on target shapefiles to ensure proper data structure and geometry.

12. Verify SDE Server

a. The database server is operational with the ArcSDE software.
b. An ArcSDE instance has been created and started.

c. The ArcSDE instance has been populated with the HLS geodatabase schema.

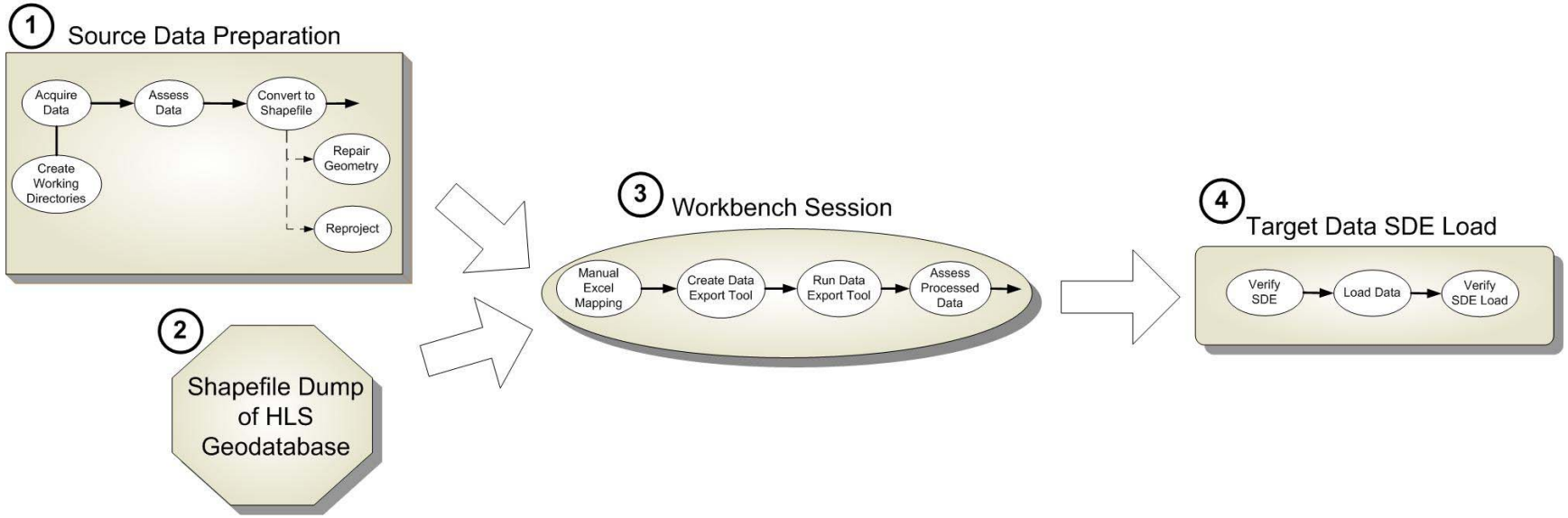
13. Load Data into SDE HLS Geodatabase

a. Use the ArcCatalog Simple Data Loader to load target shapefile into ArcSDE HLS Layer.
b. Use the ArcCatalog Simple Data Loader to load the Source_DatasetID tabular data into ArcSDE (DataSetDetail Table).
c. Use the ArcCatalog raster loading tools to load raster data into the appropriate raster catalog in ArcSDE.

14. Verify SDE Load

a. Use ArcCatalog and ArcMap to compare the loaded features with the Excel mapping spreadsheet and source receipt transmittals to ensure all features loaded as required in SDE.

ETL Process Flow



Data Transfer Framework Description

This section outlines the envisioned process for data providers in the Bay Area contributing data to the BAR-HSDS project and having that data get replicated to the various servers.

1. **ETL Incapable Data Providers** have data to share.
    a. These ETL Incapable Data Providers do not have the GIS skill set to perform the ETL but want to contribute data.
    b. These ETL Incapable Data providers do not have the software to perform the ETL.

2. Raw data is FTP'd to Data Integrator from ETL Incapable Data Providers via FTP.
    a. E-mail sent to Data Integrator affirming FTP delivery with attached zip file password, transmittal document, and attribute mapping spreadsheet.
    b. Schedule must be identified to reduce administrative burden.
    c. Data should be zipped and password protected.

3. **ETL Capable Data Provider** ETLs their data into target HLS Data Model shapefiles.
    a. A shapefile dump of the HLS Data Model must be provided to each potential ETL Capable Data Provider beforehand.
    b. Allows data providers to perform their own ETL and reduce burden on Data Integrators.
    c. Must have ArcGIS 9 and Data Interoperability Extension.

4. Raw and processed data is FTP'd to Prime Site from ETL Capable Data Providers.
    a. E-mail sent to Data Integrator affirming FTP delivery with attached zip file password and transmittal document.
    b. Schedule must be identified to reduce administrative burden.
    c. Data should be zipped and password protected.

5. **Data Integrator** receives raw data from ETL Incapable Data Providers and performs ETL.
    a. E-mail sent to ETL Incapable Data Provider acknowledging data receipt.
    b. A shapefile dump of the HLS Data Model must be provided to each potential data integrator a priori.
    c. Data Integrator can be any site with the ETL know-how and have ArcGIS 9 and Data Interoperability Extension.
    d. Uses attribute mapping spreadsheet from ETL Incapable Data Providers to create mapping file and perform ETL.

6. Raw and processed data is FTP'd to Prime Site from Data Integrator.
    a. E-mail sent to Prime Site affirming FTP delivery with attached zip file password and transmittal document (now includes information on original data and target shapefile).
    b. Schedule must be identified to reduce administrative burden.
    c. Data should be zipped and password protected.

7. **Prime Site** loads data onto server.
    a. E-mail sent to Data Integrator or ETL Capable Data Provider acknowledging data receipt.
    b. Data is logged into SourceList Access Database.
    c. Target shapefiles have Source_DatasetID added.
    d. Target shapefiles are loaded into SDE.
    e. Source_DatasetID added to SDE for each target shapefile.
    f. Raw data is loaded into appropriate FTP directory.
    g. BAR-GC Viewer is synchronized with new data.

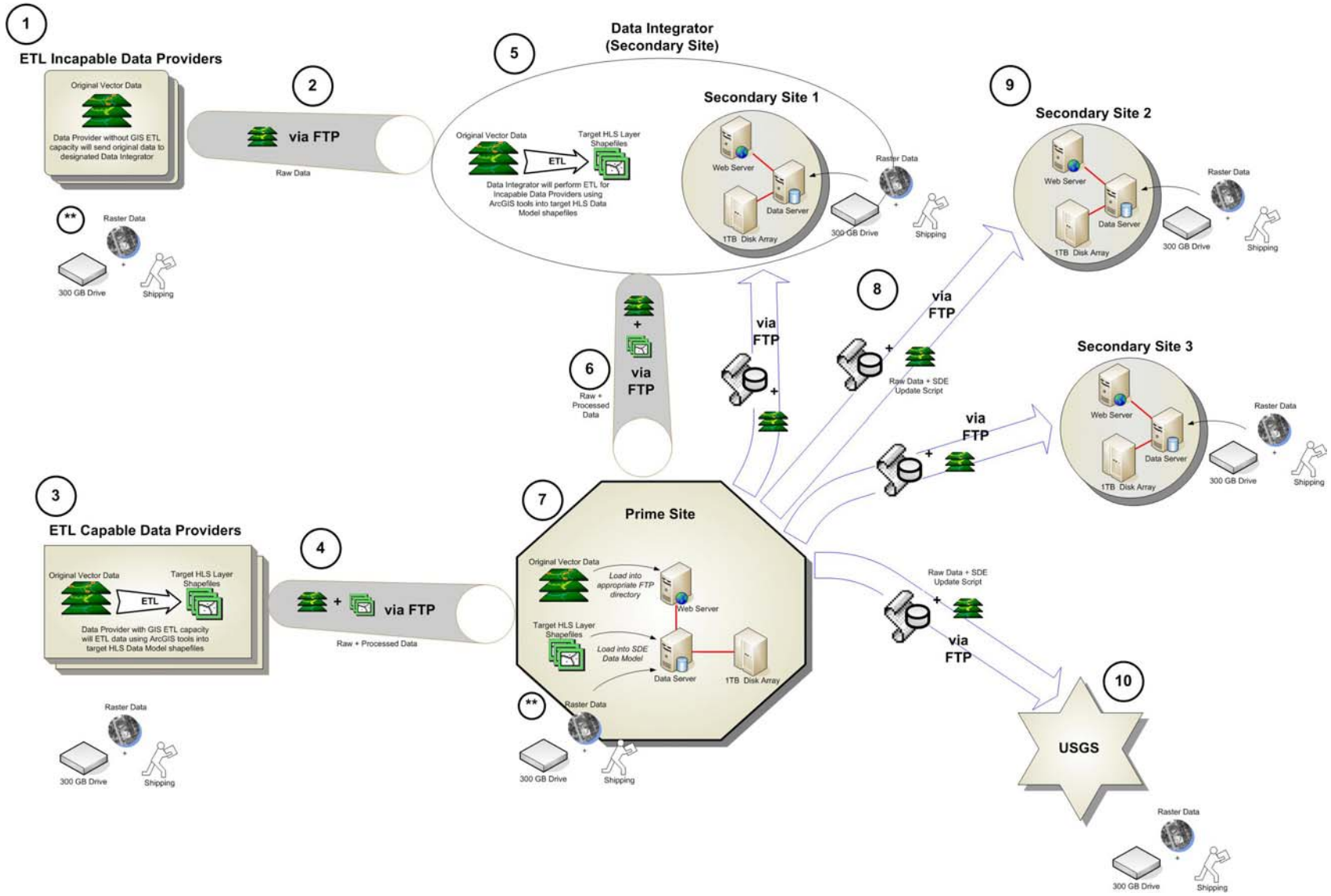
8. Data replication process initiated via FTP.
    a. Prime Site will be identified as a "master node" in the SDE scripts and other sites will be "child nodes."
    b. If "master node" goes down, a secondary site can step up as "master node" in replication process.
    c. SDE scripts will be sent in password protected zip files.
    d. Raw data will be sent in password protected zip files.
    e. E-mail sent to Secondary Sites affirming FTP delivery of SDE script with attached zip file passwords and transmittal document.
    f. Data replicated at the layer level.
    g. One way data transfer.
    h. Replication schedule must be identified to reduce administrative burden.

9. **Secondary Sites** initiate SDE scripts and import updates to SDE.
    a. E-mail sent to Prime Site acknowledging data receipt.
    b. SDE scripts are executed.
    c. Raw data is loaded into appropriate FTP directory.
    d. BAR-GC Viewer is synchronized with new data.

10. **USGS Site** initiates SDE scripts and import updates to BAR-GC data.
    a. E-mail sent to Prime Site acknowledging data receipt.
    b. SDE scripts are executed.
    c. Raw data is loaded into appropriate FTP directory.

****Raster Data:** Due to the size and infrequent updates, raster data will be transferred between sites through the use of the 300 GB Fire Wire/USB Drive that the prime site and one secondary site will have. This external drive can be transferred between data providers directly to prime site, secondary sites, and USGS via a parcel carrier and loaded into the respective raster catalog in SDE.

BAR-HSDS Distributed ETL and Data Transfer Framework



SDE Replication Scheme Description

Overview:

Replication is a two step process. The Prime site, in essence, publishes its data in the form of compressed SDE-export files at a scheduled time. The secondary sites independently check the Prime site for any new data, and if there are updates, it retrieves, uncompresses, and imports the SDE-export files. All sites have the same replication script. Arguments in a batch file that launches the script determine if the site is exporting its data, acting as a Prime site, or importing data, acting as a secondary site.

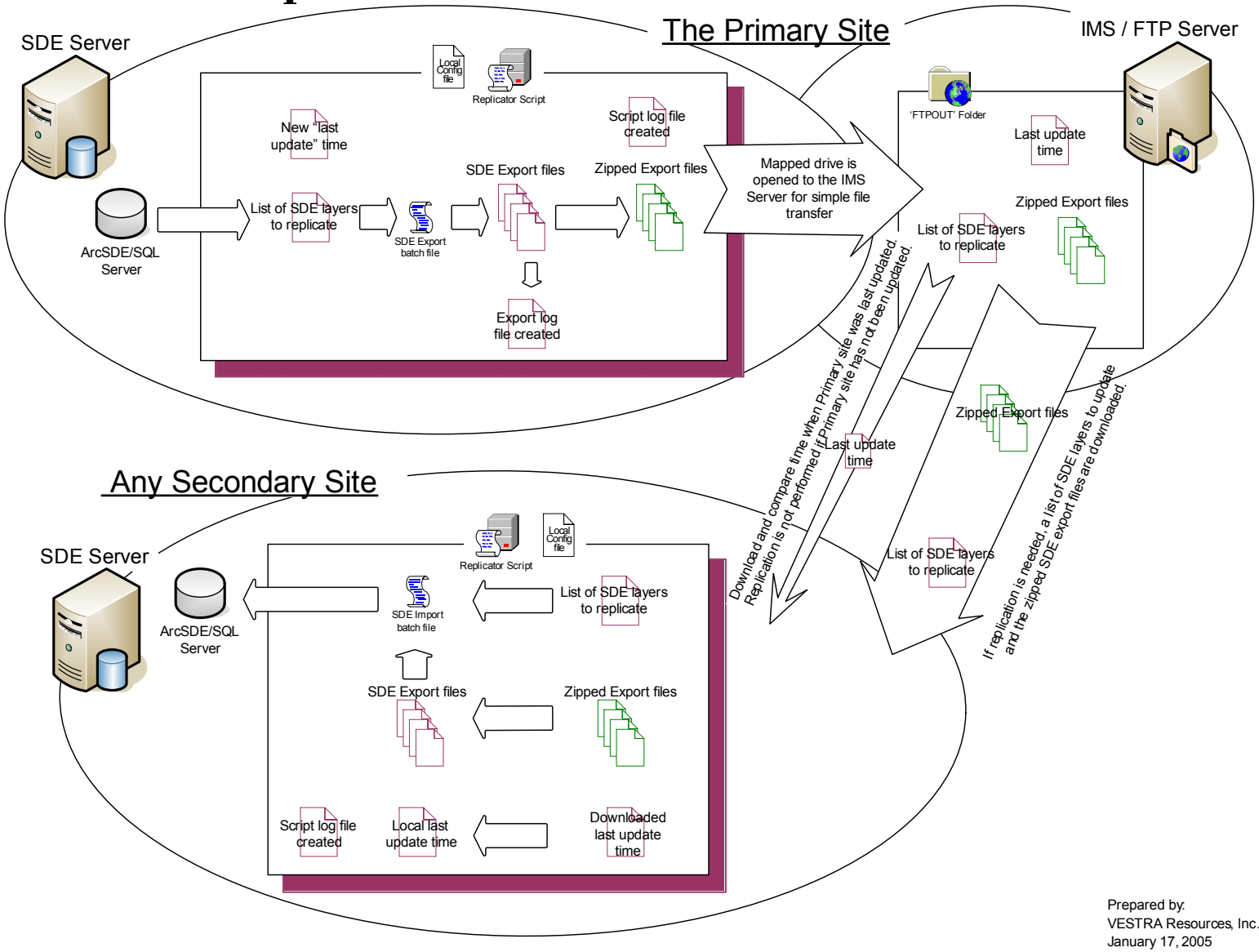
At the Prime site:

1. Data replication batch file is started manually or automatically using the Windows scheduler.
2. Log file is created.
3. Layer list file is created.
    a. Currently, this is a list of all the layers in the data repository.
4. A batch file is created, then executed to create SDE-export files for each SDE layer in the layer list file.
5. The SDE-export files are compressed.
6. A 'last updated' file is created with the date-time that the data was exported.
7. All files are built in a folder, in the ArcIMS box, that is accessible by data importer sites through FTP.

At any secondary site:

1. Data replication batch file is started manually or automatically using the Windows scheduler.
2. A connection is made to the Prime FTP site and the 'last updated' file is retrieved.
3. The 'last updated' file is compared to a local 'last updated' file, if the two are the same, then the replication script ends.
4. The layer list file is retrieved from the Prime site.
5. Using the layer list, the compressed SDE-export files are downloaded to a local folder.
6. The SDE-export files are uncompressed.
7. A batch file is created and executed that imports the SDE-export files into the local database. All the records in the current database are replaced by the records in the new export files.
8. The local 'last updated' file is replaced by the 'last updated' file from the Prime site.

BAR-HSDS Replication Scheme DRAFT



Prepared by:
VESTRA Resources, Inc.
January 17, 2005

Appendix E

Bay Area Regional GIS Council Memorandum of Understanding

BAY AREA REGIONAL GIS COUNCIL
MEMORANDUM OF UNDERSTANDING FOR DATA DISTRIBUTION, SUPPORT OF HOMELAND SECURITY REQUIREMENTS, AND DEVELOPMENT OF *THE NATIONAL MAP*

DRAFT 13 10/2004

This multi-participatory Agreement is between the Bay Area Regional GIS Council ("BAR-GC") Participating Agencies for the expressed purpose of building a regional geographic data set under the stewardship of the Bay Area Regional GIS Council, to support National Homeland Security efforts, and to participate with the US Geological Survey in the development of *The National Map* for the Nation.

WITNESSETH:

WHEREAS the BAR-GC is a regional geographic information coordinating council, formed and recognized by the State of California's GIS Council and

WHEREAS one purpose of the BAR-GC is to establish an institutional method for enabling easy, convenient, and secure access to the region's digital geographic information, and

WHEREAS participating agencies who wish to share geographic data with each other, retain all rights to said data, and

WHEREAS the BAR-GC, as the steward to this shared data, may develop a regional data repository for enabling the sharing of said data;

WHEREAS the US Geological Survey coordinates with local and regional partners to provide public domain access to the highest quality geospatial data; and

WHEREAS The US Geological Survey, as brokering agency, is working in partnership with the National Geospatial-Intelligence Agency (NGA) to support Homeland Security and Homeland Defense missions;

NOW, THEREFORE, the Participating Agencies of the BAR–GC enter into the following Agreement.

1. Statement of Purpose

Participating Agency agrees to convey to the BAR-GC, as the data steward for this MOU, a copy of the requested data themes described in Appendix YY according to the terms, conditions, and restrictions of this MOU Agreement

2. Definitions

For the purpose of clarity in this Agreement, the following terms are defined:

Geospatial Data – the digital geographic information and related attribute records, data files, and metadata

Metadata – information describing geospatial data, including the contact person in the data owner's agency, the contents of the database, the data accuracy, projection, datum, currency (date of data capture), and format of each data theme in the database

Licensee – any entity, organization or individual that has been granted access to BAR-GC data

and has agreed to the provisions of this document.

MOU –This Memorandum of Understanding

3. **Use Restrictions** – The following matrix identifies the User Type Restrictions (in the left-hand column) and Use Category Restrictions (in the top row) that the Participating Agency may specify for individual data layers, at the discretion of the Participating Agency.

USER TYPES	USE CATEGORIES (1) Internal Use Only	(2) Bitmap display via web	(3) Free distribution to third parties	(4) Free distribution to third parties via internet
(A) Emergency Service Provider	A1	A2	A3	A4
(B) Government Agencies or their delegated agents	B1	B2	B3	B4
(C) Other Public or EducationalInstitutions	C1	C2	C3	C4
(D) Data Contributors	D1	D2	D3	D4
(E) Public Domain	E1	E2	E3	E4

Explanation of User Categories

1. Internal Use Only – Geospatial Data may be used only by the Licensee or its designated agents to conduct the operations of the Licensee
2. Bitmap display via web – Geospatial Data may be displayed on the Licensee's internet website in a raster display format only, but not made available for download
3. Free distribution to third parties – Licensee may distribute data to third parties at no fee. Licensee is restricted from distribution via the internet.
4. Free distribution to third parties via Internet – Licensee may distribute data to third parties at no fee. No restriction from Internet based distribution.

Explanation of User Types

A. Emergency Service Provider – As currently structured, the Bar-GC Homeland Security Data Server (Project Homeland) partners are limited to the following:
   a. US Geological Survey, defined as broker to NGA
   b. Department of Defense (DoD)
   c. Army Corp of Engineers (USACE)
   d. Association of Bay Area Governments (ABAG)
   e. CA Office of Emergency Services
   f. All nine county participating members of the BAR-GC: Office of Emergency Services, and departments responsible for: law enforcement, fire response, and emergency medical first responders
   g. Additionally, as needed and defined in addendums to this MOU
B. Governmental Agencies or their delegated agents
C. Other Public or Educational Institutions
D. Data Contributors – Organizations or individuals that sign this MOU and contribute data to the BAR-GC data repository.
E. Public Domain – Any person or agency who requests access to the data

The User Type Restrictions and Use Category Restrictions at each level are inclusive of all less-restrictive levels. For example, a C2 equates to the following:

- Access granted to Emergency Service Providers (A), Government Agencies or Their Delegated Agents (B) and Other Public or Education Institutions (C), and
- Access granted for the above users to use the data for Internal Use (1) and Internet/Web based display (2).

4. Framework Data Layers

The data themes outlined below are the seven framework data layers, as defined by the Federal Geographic Data Committee (FGDC) that the BAR-GC intends to compile for the nine-county region. This list in no way restricts what can or cannot be contributed.

- Geodetic Control
- Orthoimagery
- Elevation
- Transportation
- Hydrography
- Governmental Units/Boundaries
- Cadastral Information

5. Ownership of Data

BAR-GC acts as the steward for the Geospatial Data in its possession. Ownership of the data remains with the Participating Agency that contributed the data to the repository. All title, ownership, and intellectual property rights that may exist or be created with the Geospatial Data shall remain with the Participating Agency.

This Agreement does not constitute the sale of any title, interest, or rights by either party other than those expressly granted by this Agreement. Any portion of the Geospatial Data, or its derivative products, that is modified or merged into another computer file by a Licensee, or is integrated with other data to form derivative products, shall continue to be subject to the provisions of the license agreements required by individual entities who contribute to the BAR-GC repository.

6. Copyright Notice

Copyright notices, if applicable, will be provided to the BAR-GC by all Participating Agencies who contribute Geospatial Data. The BAR-GC will distribute these notices along with the data and metadata to Licensees. All publications or public communications using any BAR-GC data for release to the public or to others within the Licensee's organization must include all Participating Agencies' copyright notices, if that Agency's data is used in the product.

All publications using geographic information derived from any BAR-GC data, used for release to the public or within a Licensee's organization must include the Participating Agency's copyright notice, if that Agency's derivative data is used in the product.

7. Data Update

Participating Agencies will provide updates of their respective Geospatial Data on a schedule appropriate to data theme and maintenance constraints, but no less frequently than an annual basis.

8. Metadata Maintenance

BAR-GC recognizes and endorses the tenants of the National Spatial Data Infrastructure, as promulgated by the U.S. Federal Geographic Data Committee.

- Participating Agency will provide BAR-GC with metadata for each data theme stored in the repository; such metadata shall conform to NSDI metadata standards.
- Participating Agency shall notify BAR-GC of any metadata changes, and submit updated

files. Such updated metadata shall be delivered to BAR-GC in a timely manner along with the updated data.

9. Disclaimer of Liability

BAR-GC Geospatial Data is being collected and used by Participating Agencies for the express purposes of fulfilling their respective missions.

BAR-GC, as steward for the Geospatial Data contributed by Participating Agencies, accepts the Geospatial Data "as is", with no guarantee or warranty of accuracy, currency, completeness, or fitness for any use. No oral or written information, or advice given by Participating Agency, shall be construed as a warranty.

While all due efforts will be made to assure that the data conforms to documented specifications of accuracy and completeness, neither BAR-GC nor Owner will make demands on the other if errors or omissions are found.

Neither BAR-GC nor Participating Agency shall be liable for any occurrence or activity relating to the Geospatial Data, including but not exclusive to: lost profits, the fitness of the Geospatial Data for a particular purpose, the installation of the Geospatial Data, or the analysis derived from said Geospatial Data. This disclaimer shall survive the termination of this MOU Agreement.

In no event shall Participating Agency be liable for any of BAR-GC's incidental, consequential or special damages arising out of the use, or the inability to use Participating Agency's data (including without limitation, loss of use, time or data, inconvenience, commercial loss, lost profits or savings, or the cost of computer equipment and software), or for any claim against BAR-GC by any third party.

10. Disclaimer Notice

Disclaimer notices, if applicable, will be provided to the BAR-GC by all Participating Agencies who contribute Geospatial Data. The BAR-GC will distribute these notices along with the data and metadata to Licensees. All publications or public communications using any BAR-GC data for release to the public or to others within the Licensee's organization must include all Participating Agency's disclaimer notices, if that Agency's data is used in the product.

11. Term of MOU

This MOU Agreement shall be effective from the date of signature by the Participating Agency. Agreement shall remain in force as long as its terms are not violated.

12. Termination of MOU

If a Participating Agency determines that a Licensee has violated the terms of this MOU or License Agreement (see Appendix A), Participating agency can terminate an individual MOU relationship with Licensee. Once MOU is terminated, Licensee's access to the aforementioned Participating Agency's data is removed.

Upon termination of this Agreement, for any reason, BAR-GC shall remove all files from its computer systems, return all files, documentation, and copies thereof. BAR-GC shall not be responsible for copies of Owner's data that may be in possession of Licensees. However, BAR-GC will identify all licensed users of Owner's data to the Owner.

13. Governing Law

This MOU Agreement shall be governed by the laws of the State of California and any action related to the Agreement shall be located in Alameda County, California.

14. Notice

Any notice required to be given by either party, or which either party may wish to give, shall be in writing, addressed as follows:

To BAR-GC:

BAR-GC
Contact Person
Address
City, State, Zip
Phone, E-Mail

Or to such other place as BAR-GC shall designate by written notice.

To Participating Agency:

Participating Agency's Name
Contact Person
Address
City, State, Zip
Phone, E-Mail

15. Entire Agreement

Participating Agency agree to all statements, terms and conditions stated herein this MOU Agreement.

Persons who sign this agreement hereby affirm that they are authorized to encumber their organizations with the provisions herein.

Owner		**BAR-GC**	
Organization Name:			
Date:		Date:	
Signed:		Signed:	
Name:		Name:	
Title:		Title:	

US Geological Survey			
Date:			
Signed:			
Name:			
Title:			

Appendix YY – Security Designation Form

Digital Geographic Data			
		Use Restrictions (from Matrix below)	
Layer Name	**Layer Description**	**USER TYPES (Enter A, B, C, D, E or NA*)**	**USE CATEGORIES (Enter 1, 2, 3, 4 or NA*)**
Geodetic Control	geodetic control stations		
Orthoimagery	georeferenced image from aerial photograph or other remotely sensed data		
Elevation	elevations of land surfaces and the depths below water surfaces		
Transportation			
o Street centerlines			
o Rail			
o Other transp. layers	Trails, airports/ports, bridges/tunnels, etc.		
Hydrography			
o Lakes/ponds/canals,			
o streams/rivers			
o shorelines			
Governmental Boundaries			
o Cities			
o Counties			
o Other gov. boundaries			
Cadastral Information	Parcel boundaries		
Add other data layers, as appropriate			

Digital Tabular Data			
		Use Restrictions (from Matrix below)	
Table Name	**Table Description**	**USER TYPES (Enter A, B, C, D, E or NA*)**	**USER TYPES (Enter A, B, C, D, E or NA*)**
Assessors Data	Parcel boundary attributes (owner, address, assessed value, use, etc.)		
Add other tables, as appropriate			

Attach more pages, if necessary

* **NA** means “no access”

USE CATEGORIES

USER TYPES	(1) Internal Use Only	(2) Bitmap display via web	(3) Free distribution to third parties	(4) Free distribution via internet
(A) Emergency Service Provider	A1	A2	A3	A4
(B) Government Agencies or their delegated agents	B1	B2	B3	B4
(C) Other Public or EducationalInstitutions	C1	C2	C3	C4
(D) Data Contributors	D1	D2	D3	D4
(E) Public Domain	E1	E2	E3	E4

Appendix F

Bay Area Regional GIS Council Risk Assessment/Management Plan

Bay Area Regional GIS Council

BAR-GC

Risk Assessment / Management Plan

Revised June 19, 2002

Table of Contents:

Executive Summary

Project Brief

Bay Area Regional GIS Council Plan
Project Name

Lis Klute
Project Manager

MTC, ABAG, BAAMA, CCC, City of Berkeley, Bay Area Regional GIS Council Formation Team

Risk Plan	Projected	Budgeted
Costs	$20,000	$0.00
Start Date	2/01/02	3/01/02
	5/30//02	3/4/02

History & Situation: Geographic Information System (GIS) development efforts and policy issues have generally been addressed largely County and agency by agency within the Bay Area Region. Efforts to combine and create regional core GIS layers such as parcels, roads, political boundaries have been largely unanswered. The recent creation of a State of California GIS Council (CGC) has renewed the opportunity for better coordination between state and local agencies toward the creation of regional GIS policies and data.

Issues & Opportunities: The primary issue is the lack of an established organization or local and regional agencies representing the 9 county Bay Area, dedicated to work together on GIS issues. The State has directed the Counties or agencies to form regional councils to assist the state council in reaching local jurisdictions and to act as the receptor for input to the State Council. The opportunity is to develop a Bay Area Regional (BAR) GIS Council. that can work together to provide input related to GIS policy, development of core GIS data sets, and coordinated and cooperative data sharing. The BAR will compile the business needs, functional requirements of the regional and local governments as they apply to the development, maintenance, sharing and analysis of core GIS data layers and specific strategic projects in support of the CGC.

Scope: To develop a Regional Council that will review the Californian GIS Council's reports, provide recommendations on policy alignment with local and regional needs and be a focal point for partnerships to build California's geographic information infrastructure. The goals will include building consensus on regional area definition; developing a Risk Assessment Plan for success of the region; to be inclusive on the council, of all stakeholders of geographic information system development, implementation, utilization and distribution.

SMART Objectives (specific, measurable, achievable, related to division objectives & time bound)

1. Develop a Regional Council Plan
2. Develop and receive consensus for Regional Plan by majority of sponsors by March 5, 2002.
3. Formation Plan Presentation to Regional Sub Committee Representative, March 6, 2002
4. Distribute BAR Plan to shareholders and solicit BAR membership requests by .March 18, 2002
5. Identify BAR Interest group & develop recommendations on 2 year council members. By April 4, 2002
6. Hold the first BAR meeting following the BAAMA meeting April 18, 2002

Schedule

Estimated Duration 4 mos. **Proposed Start Date: 2/4/01** **Proposed Finish Date: April 18, 2002**

Milestones	Dates
1. Initiate project	2/4/2002
2. Finish contacting potential stakeholders (Counties, regional agencies)	3/4/2002
3. Preliminary risk assessment accepted	3/5/2002
4. Deliver project plan presentation to sub-committee representative	3/6/2002
5. Create agenda for meeting	4/10/2002
6. Hold First Formation Meeting	4/18/2002

Success Criteria / Definition of Completion:
Milestones successfully completed.
Acceptance of plan presented to state CGC Regional coordinator
Council gets buy-in from stakeholders
Majority of stakeholders want to meet again after first meeting

Vision Statement: Develop a working relationship among regional agencies in the Bay Area to review the California Geographic Information System (GIS) Council's reports, provide recommendations on policy alignment with local and regional needs and be a focal point for partnerships to build California's geographic information infrastructure. The Council will help to identify, compile and coordinate the business functions of the regional and local agencies with respect to the use of GIS data. The Council will meet regularly facilitate the discussion, development and implementation of opportunities to develop GIS data and data models.

Assumptions & Constraints Assumptions:
1)BAR will have sufficient participation from 9 Counties to allow for regional representation at the State Council 2) Funding will be found to permit development of a regional council 3) Money can mitigate some of risks

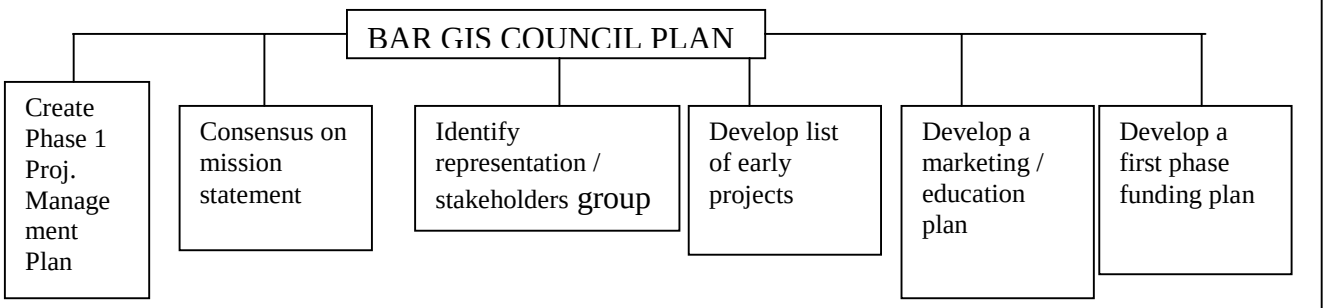
C**onstraints:** 1) Buy in from organizational leaders to support regional council. 2) Funding may be difficult during this period. 3) Most GIS coordinators/managers are already over-committed.

Major Resources:
Contra Costa County GIS coordinator, BAAMA, ABAG, MTC, State GIC Council, City of Berkeley

Risk Management Overview (3-5 Major Risks)

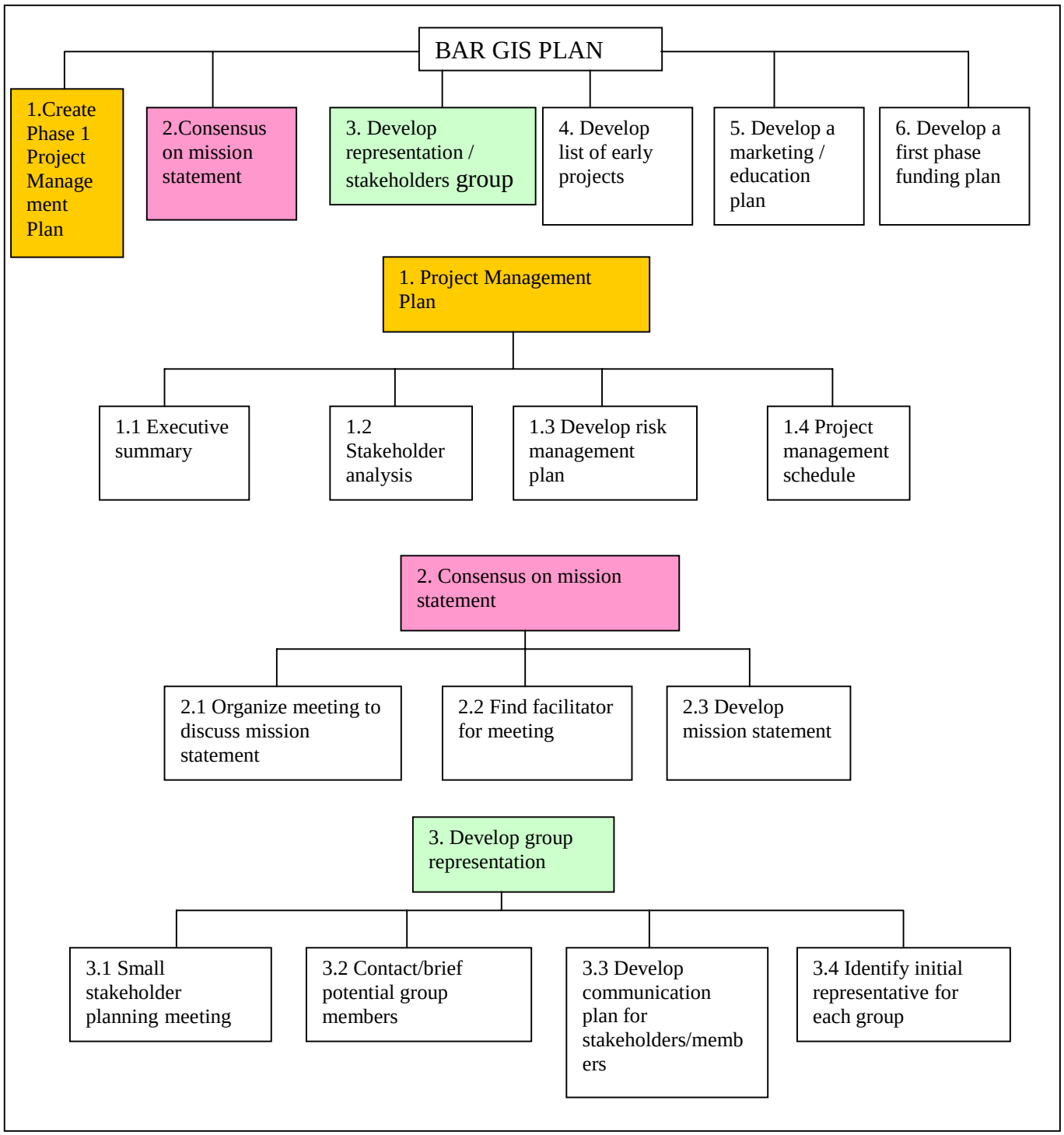
Major Risk	Prob	Impact	Avoidance
Having buy-in from enough stakeholders (Counties)	40%	High, stake holders won't participate	Good marketing, communication, organized meetings, early wins (deliverables)
Not finding a good facilitator	30%	High, stakeholders could lose interest	Identify a well qualified person to help facilitate input from the stakeholders.
Not clearly identifying potential benefits	50%	Moderate, Half-hearted participation	Close work with state/stakeholders to identify opportunities (funding, support, etc..)
Stakeholders not having enough time to participate	40%	High, success of program requires majority of entities	Providing stakeholders with a clear product(s) to show decision-makers.
Not maintaining good relationship with state GIC	30%	Success of regional group may depend on state funds	Stakeholder analysis both regional and state. Regular good communication w/ state GIC

Work Breakdown Structure (WBS) to level 2



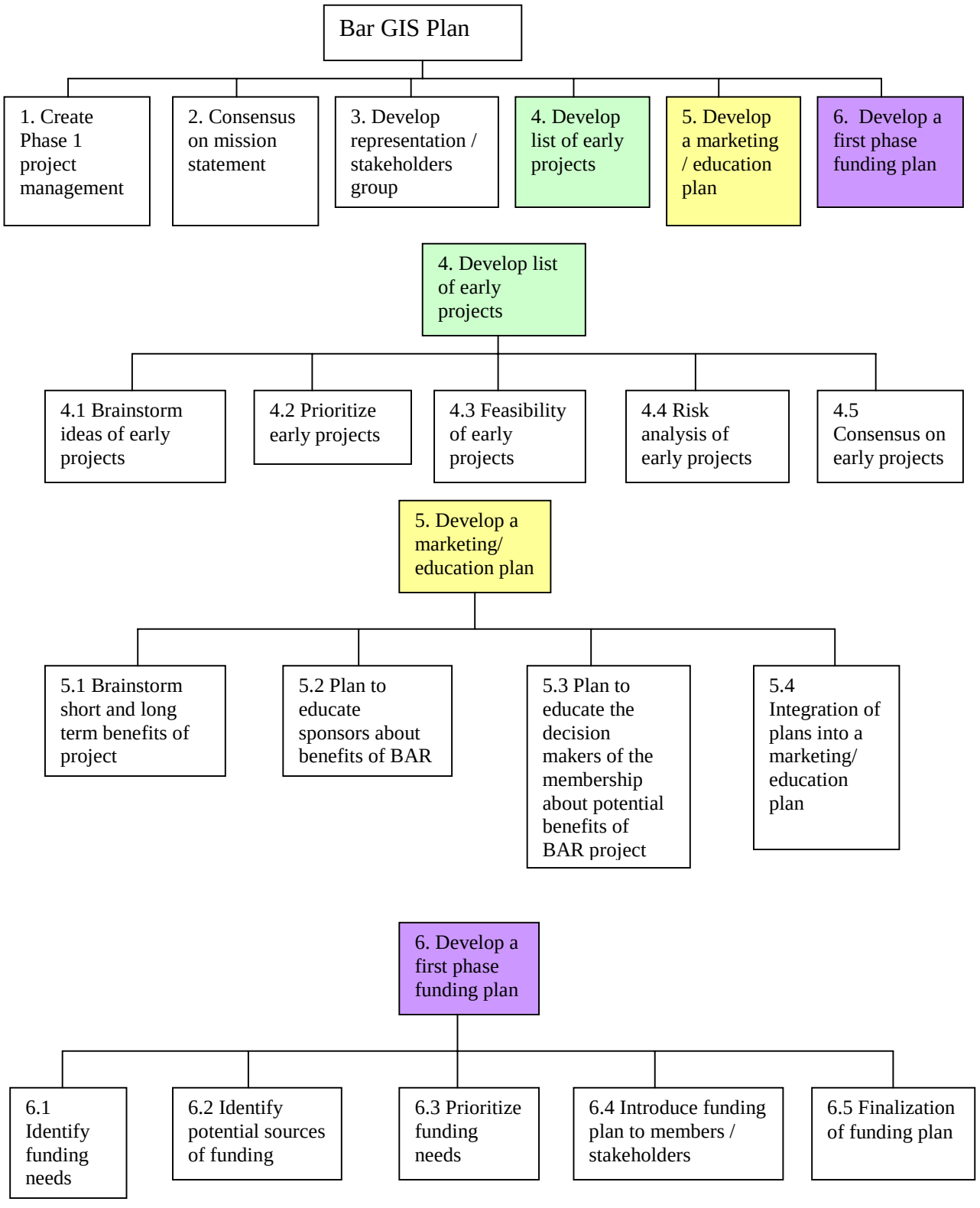
Other Special Requirements or Considerations: Most challenging part of project will be to facilitate meetings in order to maximize input from all stakeholders and install a joint feeling of ownership. In addition, it will be important to identify beneficial and marketable milestones that will justify stakeholders' participation to their respective decision-makers.

2. Work Breakdown Structure (WBS) to level 1 - 2



2. Work Breakdown Structure (continued)

Work Breakdown Structure (WBS) to level 1 – 2 (4, 5, 6)




3. List of Possible Risks:

1. Having buy-in from enough stakeholders (Counties, Cities, others) M
2. Not finding a good facilitator M
3. Not clearly identifying potential benefits H
4. Stakeholders not having enough time to participate H
5. Project champions having enough time to organize and coordinate H
6. Not maintaining good relationship with CGC M
7. Challenge of integrating multiple formats of GIS data for any given layer M
8. Challenge of developing data models M
9. Challenge of developing GIS policy H
10. Insufficient funding within state to help regional efforts L
11. Insufficient funding within state to move CGC forward L
12. Competition between local/regional agencies for money or resources M
13. Not developing trust between agencies M
14. Past history between groups M
15. Participants not willing to compromise on issues M
16. Not identifying short and long terms goals M
17. Marketing the coordination effort effectively M
18. Insufficient support by regional stakeholders to follow through with identified projects M
19. Changing priorities among participants and CGC can effect cohesiveness and focus M
20. Lack of understanding by decision makers of long and short term goals of BAR GIS Council M
21. Lack of understanding by public/clients of participants of importance of effort M
22. Ability to maintain probable website to share results L
23. Mechanics of coordinating data with existing varied data sharing policies within region H
24. Parasitic stakeholders; individuals or groups trying to "torpedo coordination effort M

Severity of Risks

H High
L Low
M Medium

Section 4. General Risk Assessment

Risk Management Overview (General Risk Assessment)

Major Risk	Prob	Impact	Avoidance
Having buy-in from enough stakeholders (Counties, Cities, others)	40%	High, stake holders won't participate	Good marketing, communication, organized meetings, early wins (deliverables)
Not finding a good facilitator	30%	High, stakeholders could lose interest	Identify a well qualified person to help facilitate input from the stakeholders.
Not clearly identifying potential benefits	50%	Moderate, Half-hearted participation	Close work with state/stakeholders to identify opportunities (funding, support, etc..)
Stakeholders having enough time to participate	40%	High, success of program requires majority of entities	Providing stakeholders with a clear product(s) to show decision-makers.
Maintaining good relationship with state GIC	30%	Success of regional group may depend on state funds for future projects.	Stakeholder analysis both regional and state. Regular good communication w/ state GIC
Project champions having enough time to organize and coordinate	25%	Moderate, effort could fall apart if primary organizers leave effort	A good marketing plan of potential benefits of BAR GIS plan to the decision makers is key. The key role that the project champions play is important to point out.
Challenge of integrating multiple formats of GIS data for any given layer	20%	Low, a plan to develop approach can likely be identified.	Communication with other regional and state efforts pursuing the BAR GIS goals plan is key. Input from GIS vendors about possibilities. Consider interoperability standards.
Challenge of developing data models	20%	Low, other data models exist; process to developing new models exists	Good facilitator to conduct needs data model process, help from technically qualified individuals; help from other groups that have worked on these models.
Challenge of developing GIS policy	25%	Moderate, some levels of GIS policy between members will need to be created	Not all issues can be solved. Respect and acknowledgment of differences important. Policies will likely involve compromise. Minimize need to impact other agencies' exiting GIS policies.
Insufficient funding within state to help regional efforts	20%	Moderate, State is experiencing large short fall, may cut-back support to the state level effort.	BAR may find local and alternative sources of funding with which to operate and carry out it's initiatives. BAR has the opportunity to provide success stories to the state on the importance of its state GIC effort.

Section 4 Cont.

Risk Management Overview (General Risk Assessment)

Risk	Prob	Impact	Avoidance
Competition between local/regional agencies for money or resources	10%	Low, might tend to dilute the cohesiveness of the BAR group.	Show membership value of doing grant together and endeavor to structure goal as to cover the most number of groups.
Not developing trust between agencies	20%	Moderate, might create difficulty in getting consensus on goals and objectives	Make sure lines of communication are open. Provide high quality information to membership. Start with small easy to achieve projects.
Past history between groups	20%	Low, past coordination/competition may slow down effectiveness of group	Emphasize "starting fresh" nature of this project; opportunities for long term and short term benefits for all. Identify small pilot projects with high likelihood of success.
Participants not willing to compromise on issues	20%	Moderate, unwillingness to compromise to meet the needs of the group will be very debilitating to success.	Make sure that issues are not from lack of understanding. Allow for ability to disagree and move off to side contentious issues.
Not identifying short and long terms goals	25%	High, this is extremely important to getting participation of members and their respective sponsors	Provide a good facilitator to group for these discussions. Plan the meeting carefully so discussion does not get side-tracked. Make sure meetings notes are collected and distributed.
Marketing the coordination effort effectively	20%	Moderate. Without effective coordination plan, membership interest in project will wane.	Develop an effective communication plan. Update state and other groups regularly on status of process, completion of projects. Present activities at conferences. Give useful feedback to membership to take to their clients.
Insufficient support by regional stakeholders to follow through with identified projects	20%	Moderate, effort may fall apart if membership does not take ownership and responsibility of accepted tasks and assignments.	Project manager needs to be proactive to facilitate memberships participation. Members should not over commit. Projects need to be clearly identified and time to perform evaluated.
Changing priorities among participants can effect cohesiveness and focus	20%	Low, Progress on short and long term goals may suffer if group cannot demonstrate some regular progress.	Develop a clear WBS and Gannt chart of projects to track progress and help gauge participation.
Lack of understanding by decision makers of long and short term goals of BAR GIS Council	15%	Low, membership may have difficulty in getting permission to participate if their decision-makers are not kept informed of progress and successes.	Provide membership with regular updates of accomplishments and milestones of projects to show their respective sponsors.
Lack of understanding by public/clients of participants of importance of effort	15%	Low, public or clients can be post advocates and antagonists of projects efforts.	Provide membership with regular updates of accomplishments and milestones of projects. Explain to public on web and various meetings BAR activities and successes.

Section 4 Cont.

Risk Management Overview (General Risk Assessment)

Risk	Prob	Impact	Avoidance
Ability to maintain probable website to share results	20%	Moderate, effort needs effective way to communicate it's results. A good web site will reduce other risks	Consider finding funding to create adequate website for this group. Minimize effort to maintain site. Regularly request membership's comments of website and its creation and design.
Mechanics of coordinating data with existing varied data sharing policies within region	20%	Low, will make it more difficult to stitch together data across boundaries or membership.	Working out a regional data sharing agreement that provides flexibility of upholding individual membership's agreement.
Insufficient funding within state to move CGC	20%	Moderate, lack of CGC support may delay development of regional project efforts.	BAR GIS needs to let state know how their support for projects is. Bar should try to get funding from other than state GCG sources.
Parasitic stakeholders. (personal, past history, etc.)	20%	Moderate, could cause the group to lose focus due to negative comments or calculated distractions	Facilitator needs to be watching out for this behavior and control it. BAR organizers need to be aware of the potential for these issues and try to win over this group of stakeholders

5. Analysis by Risk Category

These are the numbered risks put into 5 categories of risk developed in a brainstorming exercise.

Communication (1, 3, 6, 15, 17, 20, 21)

Organization (2, 5)

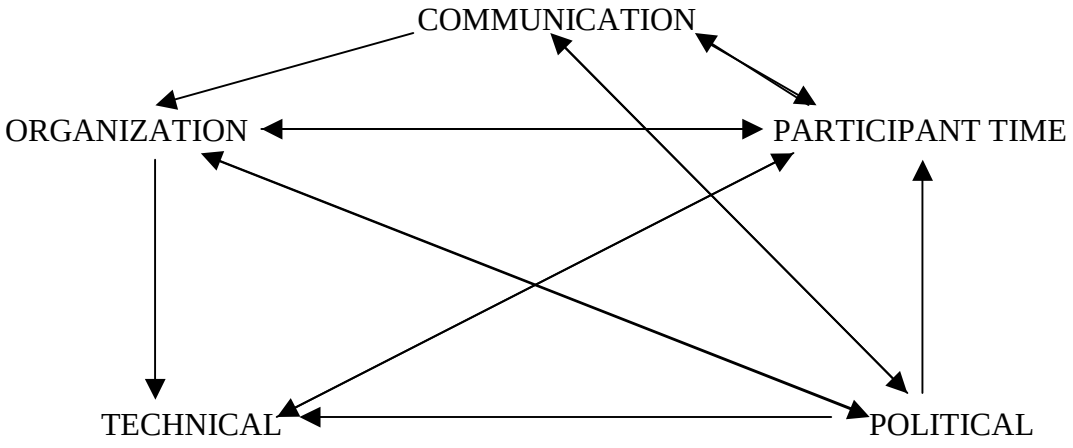
Political (9, 10, 11, 12, 13, 14, 18, 19, 23, 24)

Participant Time (4, 22)

Technical (7, 8, 16)

The following effort is a graphical method to look at which category drives or is driven by another category of risk:

Drivers: (communication, organization, participant time, technical, political)



Number of groups "Risk Category" drives:

Participant Time	3
Communication	3
Organization	3
Political	4
Technical	1

Discussion:

The purpose of this analysis was to look at the number and type of interdependencies among the various risks. The goal is to identify, that if by taking care or reducing some category of risk, that other potential risks might be likewise reduced. The 24 potential risks were grouped together into risk categories (communication, organization, participant time, technical and political). A diagram was then used to show which kinds of risks drove other activities. A summary of these probable dependencies is shown above. It seems that the political type of risks (rank of 4) drive the most number of other risks. The recommendation is that focusing on the suggested avoidances for the political risks will diminish the other major categories. The category of risk, political, also contains the highest number (10) of individual risks.

The next group of risks in number of individual risks is the communication category (6). It is also pretty clear that focusing on the #3 risk related categories (participant time, communication and organization) will also be very useful.

Method 2 Risk Analysis by Weighted Probability

This method takes a look at the possible (1) probability of the various risk and (2) the severity of risk in three general categories (High, Medium, Low). I have given a weight of 3, 2 and 1 to these severities respectively (H, M, L). I have multiplied the probability of a risk times its severity value. This was done for each of the 23 listed risks:

Example Risk #1 Having buy-in from enough stakeholders.

Moderate risk and probability of 40%
2 x .4 = 0.8

Summing up all the weighted risks per category:

Weighted Risk Probability:		Sum
Political	(0.75, 0.6, 0.6, 0.2, 0.4, 0.4, 0.4, 0.4, 0.6, 0.4)	4.75
Communication	(0.8, 1.5, 0.6, 0.4, 0.4, 0.3, 0.3)	4.3
Participant Time	(1.2, 0.4)	1.35
Organization	(0.6, 0.75)	1.35
Technical	(0.4, 0.4 0.5)	1.3

This method of risk analysis suggests that the highest amount of risk (weighted probability) comes from the category of risk characterized by "political". This category also contains the largest number of risks (10) in the 5 groups. Clearly, this effort has a lot of potential landmines. However, by developing a strategy to avoid and mitigate a number of these risks, the effort will reduce the area of greatest overall risk.

As suggested previously, this category of risk also helps drive all the other categories of risk. Following the avoidance methods listed and reviewing regularly the success of these mitigations will be very helpful in reducing this area of risk and by extension it's influence in the other areas.

Communication risk is a close second, and probably at least as important as the politics. Two of the largest single risks are #3, not clearly identifying potential benefits (1.5) and #1, having buy-in from enough stakeholders (county, cities and others (0.8). This category also contains the second largest number of risks (6) in the 5 groups. As this group meets and tries to develop into a cohesive group, the art of listening to each other will be equally as important to what is actually said. What is not said and analysis of the stakeholders' body language will be helpful in the formation of this group.

The categories of participant time and organization are basically the same in terms of weighted probability of risk. "Participants having enough time to attend these meetings and work together on solutions will be greatly facilitated if they are made to feel the meetings and activities are well organized. They must also be given marketable short and long term goals to take back to their respective organization managers and decision makers.

6. Summary of Risk Analysis:

Much of this risk identified in this report is qualitative. The highest risk is for the participating agencies to provide enough time for their representatives to be effective. In order to allow for this, benefits to the individual agencies in terms of time and money saved (or grants) will need to be developed. The participants need to be given a probable list of short and long term benefits to the participants their decision makers and the people that the agencies support. Marketing these opportunities to decision makers for GIS development, leading to enhanced business function for the agencies will be crucial. Developing agreement on identification of short term and long term goals will be important.

Finally, the coordination meetings themselves will need to be carefully planned and run in a well organized way. A project plan will need to be developed and implemented by a project manager to ensure the greatest success of this first meeting. Objectives, agendas and minutes will need to be developed and distributed as a part of the facilitation/communication plan. These risks associated with this project have been categorized and analyzed in several ways. This project will have a higher chance of success if the risk categories that drive other risks are reduced. The project manager will need to regularly reevaluate the effectiveness of the avoidance and risk mitigation methods in reducing risk. This project should have a high probability of success if the primary driver risks and highest (in terms of weighted probability) specific risk discussed are successfully managed.

Appendix G

Site Suitability Matrix

Site Suitability Matrix – BARGC HSDS

Proposed Sponsor Agency: Association of Bay Area Governments

Q. Does your site have backup procedures in place? What is the general process?

A. Yes. Full backups are performed nightly. ABAG uses Backup Exec from Veritas.

Q. Does your site have hot site disaster recovery?

A. No.

Q. What is the bandwidth from your site to the internet?

A. T3 5.0 mbits

Q. What kinds of firewall security measures are in place?

A. CISCO PIX Firewall and Intrusion Detection Device

Q. What kind of physical security measures are in place?

A. Server Room has button coded locks

Q. Is there 24/7 help desk assistance?

A. No.

Q. Are there personnel with ESRI software experience? ArcIMS?

A. Yes.

Q. What kind of RDBMS is in place?

A. Oracle 11i and Postgres for Internet based applications

Q. What kind of access can be given to personnel from other jurisdictions?

A. VPN, Citrix, minimal guest access

Q. Is the site easy to access physically in the Bay Area? What is the address?

A. Yes. Easy access from BART and the I-880 Freeway. Address: MetroCenter, 101 8th Street, Oakland Ca. 94607

Q. Can you support the software maintenance requirements? APX $6,000/yr

A. Yes.

Q. What would you see as the necessary or suggested return on investment for your organization to host a BAR-GC HSDS server site?

A. ABAG has begun the process of facilitating the coordination efforts of Bay Area Emergency Management Coordinating Council and Emergency Response Operation Centers. If ABAG is selected as a sponsor agency for the HSDS pilot study, ABAG plans to build upon the pilot study, in an effort to implement a regional system that would compliment the goals identified in the pilot study. ABAG would be happy to meet with the coordinating group for the HSDS Pilot Study to discuss this further.

Background Information

At its last regular meeting, the ABAG Executive Board approved the formation of the San Francisco Bay Area Emergency Management Coordinating Council as an ABAG task force for the purpose of promoting and strengthening a regional approach to disaster preparedness and response. The mission of the Coordinating Council is to coordinate and guide the development of regional emergency management projects and programs for the San Francisco Bay Area. Through regularly scheduled meetings, the group will embark upon a set of tasks that will generally:

- assess our current vulnerabilities and needs and define readiness goals,
- identify the gaps in current systems and inventories, and
- work together to fill the gaps and achieve readiness goals

The disaster preparedness objectives of the Coordinating Council are complimentary to the goals of the BAR-HSDS pilot study. The BAR-HSDS would be an invaluable resource for members of the Coordinating Council and as well as existing Emergency Response Operation Centers. Once the pilot study is concluded, measurable benefits and lessons learned should be identified and documented.

Appendix H

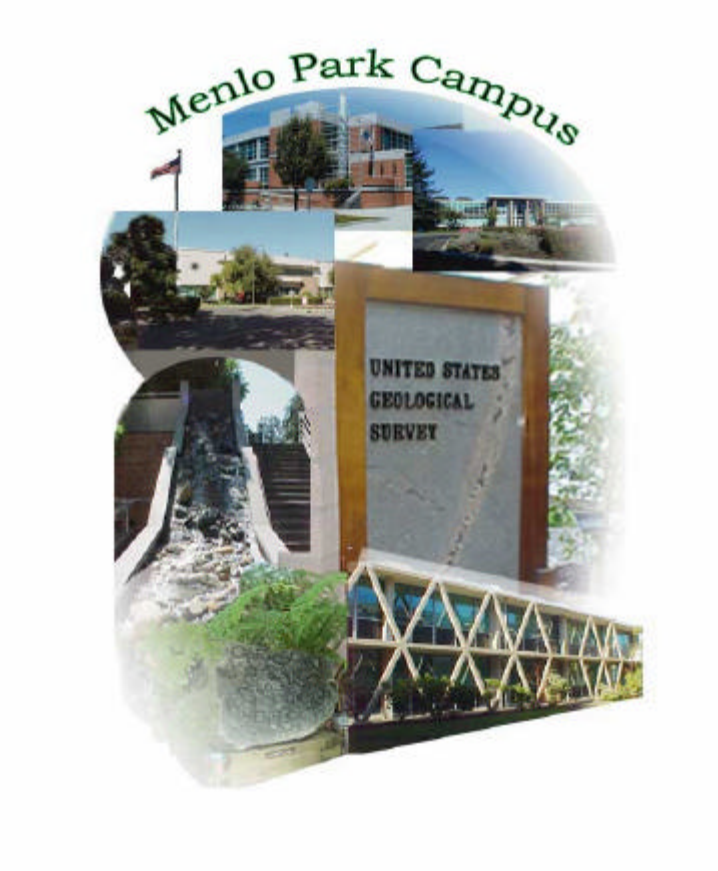
Bay Area Regional GIS Council HSDS Host Site Data Center Survey

Bay Area GIS Council (BAR-GC)

Homeland Security Data Server (HSDS)

Host Site Data Center Survey

U.S. Geological Survey, Menlo Park



January 20, 2005

Primary Information

Site Name	**U.S. Geological Survey – Menlo Park**
Address	345 Middlefield Rd
City St Zip	Menlo Park, CA 94025
Phone	(650) 329-4370
Fax	(650) 329-5528
Web	www.usgs.gov
Contact	**Carol Ostergren**
Title	California Mapping Liaison
Phone	(916) 278-9510
Email	costergren@usgs.gov
Other	3020 State University Drive East, Suite 4003 Sacramento, CA 95819
Contact	**Lore Winterman-Sturm (Systems Administration Contact)**
Title	Cartographer
Phone	(650) 329-4370
Email	lwinterm@usgs.gov
Other	
Contact	**Charlene Fischer**
Title	Alternate Systems Administration Contact
Phone	(650) 329-4370
Email	cfischer@usgs.gov
Other	**Jeff Taylor,** jtaylor@usgs.gov

Proposed Location of Server

Secured server room on the first floor of the Middlefield Road facility. The BAR-GC HSDS server rack would be located near the center of the room alongside several existing racks.

Other:

- No racks are currently anchored to the subfloor; building has been seismically retrofit to withstand an 8.0 (Richter) earthquake.

- H_2O sprinkler system currently (supplemented with manual extinguishers); move to Halatron planned in the future
- Raised floor-style server room

System Administration

- Lore Winterman-Sturm is primary contact for system-related issues: (650) 329-4370
- Charlene is the secondary contact (same contact number)

Internet Access/Bandwidth

- Multiple OC3 access to Internet (155 mbps for each OC3); 4 failover lines, including one to NASA Ames Research Center (Moffett Field) and one to EROS Data Center (Sioux Falls, SD)

Disaster Recovery/Data Protection

Surge suppression/Power backup

- Diesel generator system
- Estimated 36 hour runtime in case of total power failure
- Generator tested weekly
- Multiple battery backup system

Virus protection

- Enterprise level anti-virus maintained on system

Data Backup systems

- No system-wide backup schema; Backup is varied, and accomplished at each server and administered by project-level personnel

Security

Physical

- Entrance to facility is not monitored
- "Smartcard" access to server room
    - Only authorized personnel have cards (with picture ID)
    - Can disable cards at any time

    - Card access can be restricted to certain hours
    - Maintenance, emergency crews have smartcard access to server room
- Must complete log upon entrance/exit to server room
- Data center is located in an interior room on the first floor, with no windows

System

- Intrusion detection in place

- "Edge" routers provide first level of protection prior to DMZ

- Stringent security policies enforced; very few ports available (e.g., no Port 80 or 25 access allowed into campus)

Appendix I

Bay Area Regional HSDS Security Framework

Bay Area Regional
Homeland Security Data Server

Security Framework

DRAFT 3/11/2005

Table of Contents

Appendices

Introduction and Executive Summary

The Bay Area Regional Homeland Security Data Server (BAR-HSDS) is a distributed facility for the integration and sharing of geospatial data among organizations with homeland security and public safety concerns in the San Francisco Bay Area.

Participants include local jurisdictions and companies and certain state and federal authorities. In addition to providing a secure local clearinghouse for geospatial data, the BAR-HSDS will provide a trusted mechanism for making local geospatial data available to the national homeland-security community.

The BAR-HSDS is intended to handle unclassified material only, although some of its contents may, by their nature and/or their aggregation, fall under the broad rubric of "sensitive but unclassified" (SBU).

Elements of Security

Security encompasses a variety of goals and components. One conventional breakdown of this topic uses the mnemonic "CIA" to delineate three aspects of information security:

- **Confidentiality** – Assurance that information and services are only available to the appropriate users, and that privacy and intellectual property rights are protected;
- **Integrity** – Assurance that the information is authentic and complete and has been protected from unauthorized additions, deletions or edits; and,
- **Availability** – Assurance that the information is available when and where needed, without interruption.

We will use this as a general framework for analyzing security goals and criteria for the BAR-HSDS project.

It is critical to bear in mind that security is a design goal and an operational characteristic. It is not a commodity that readily can be "retrofit" onto a pre-existing design, nor is it one that can be implemented once and expected to persist without careful and consistent maintenance.

Resources for Security

The tools and techniques used to reach our security goals are also numerous. For this analysis we will sort those resources into four "layers";

- **Technology** – Hardware and software mechanisms;
- **Procedure** – Standards and norms for users;

- **Human Resources** – Training, awareness and attitudes; and,
- **Organization** – Structures, practices and policies for intra- and inter-organizational cooperation.

Approach to Framework Design

The design of this framework reflects the following activities undertaken by BAR-HSDS participants:

- A survey of applicable regulations, policies and recommendations at the federal, state, local and private-sector levels;
- Discussions with BAR-HSDS participants regarding their own requirements and aspirations for this project;
- Review of the initial and planned technical architecture and operating environment (hardware, software, installation sites and standards, etc.);
- Review of the initial and planned functions of the BAR-HSDS (e.g., as file repository, as geodatabase, as "back end" for fixed and mobile applications, etc.);
- Formulation of a set of policies and standards for establishment and maintenance of the BAR-HSDS security environment; and,
- Survey and cost/benefit analysis of technical options for supporting those policies and standards.

In addition, a discussion of longer-term security challenges and opportunities is provided for strategic planning purposes.

Key Recommendations

Key among the recommendations of this report are the following:

1. A Managing Entity should be designated with explicit responsibility for setting security standards and monitoring their implementation.
2. All users should be required to undertake a written agreement with the BAR-HSDS managing agency, committing each user to observe security policies and acknowledging that misuse or negligence may result in suspension or revocation of access as well as any other sanctions provided by law.
3. Each hosting agency should be required to undertake a written agreement with the BAR-HSDS managing agency, committing the hosting agency to maintain site security and safety at current or better levels, and acknowledging that failure to do so may result in that site being removed from the BAR-HSDS network.
4. All network communication of BAR-HSDS data among servers or with client computers, including geospatial data, user account data, operational commands and any other information related to the operation of the BAR-HSDS, should be encrypted at a level equivalent to 128-bit AES or TLS.
5. Details of the source, update history, copyright and any access restrictions of each data item derived from a participant-provided dataset or layer should be stored in the BAR-HSDS data structure and

provided to all data users.

6. Strong "two-factor" authentication should be implemented for system administrators and data providers. Passwords used by other users should be selected to be difficult to guess or "crack" and should be changed frequently.
7. A complete audit of user accounts should be conducted on a regular basis to verify that all account-holders are still authorized to access the system and are maintaining proper security (e.g., password changes).
8. Each host site and data provider should have a single designated BAR-HSDS Security Officer who is personally responsible for the maintenance and monitoring of security at that site and system-wide. As a group the individual Security Officers should comprise a Security Committee for the BAR-HSDS.
9. A qualified independent auditor should be engaged from time to time to review and verify the security provisions of the BAR-HSDS.
10. Firewalls at BAR-HSDS sites should implement a "least privilege" configuration that permits traffic to the minimum set of necessary ports and may restrict access to certain ports to pre-designated Internet addresses.
11. A mechanism should be established for capturing and tracking trouble reports, whether from automated monitors or from system users.
12. Non-critical software upgrades and patches should be coordinated among the sites and "staggered" to verify that no roll-back will required prior to applying changes to all sites.

Section I - Confidentiality

Definition of "confidentiality"

Different authorities define the term "confidentiality" in different ways, varying chiefly as to the narrowness or broadness of the term's scope and its precise relationship to other related words such as "privacy." For the purposes of this document we use the word in the following broad sense:

> **Confidentiality**: Assurance that specified information, meta-information (e.g., regarding the sourcing of information, patterns of communication, etc.) and services are available only to those individuals and organizations permitted by policy to possess them.

Implicit in this definition are four key subsidiary concepts:

- **Authentication** – Assurance that information suppliers and users are who they claim to be;
- **Access Control** – Assurance that information and information-processing functions are only available to users permitted by policy to access them;
- **Attribution** – Assurance that information can be traced to its original source; and,
- **Assertion of Rights** – Assurance that any restrictions on the dissemination or use of information are provided along with the information itself.

This definition of confidentiality is both broad and demanding. For example, by this definition confidentiality is breached whenever information becomes *available* to unauthorized actors, whether or not they actually or consciously access the information.

However, like all the parameters of security, confidentiality is never absolute. Ultimately, every security choice involves a cost/benefit decision, the goal of which is to make the cost of a security violation significantly higher than the benefit to the violator.

Confidentiality Requirements and Constraints

Interviews and surveys of BAR-HSDS participants have revealed that confidentiality requirements arise from a variety of concerns:

- The actual sensitivity of the anticipated geospatial data from a homeland security perspective, as assessed according to FGDC guidelines, is generally acknowledged to be minimal at present, but could become significant factor later, if, for example,

commercial infrastructure (e.g., electrical, fuel or telecommunications) datasets were added to the BAR-HSDS;

- Several jurisdictions have internal policies restricting the redistribution of their geospatial data, which, while they may be more rigorous than suggested by FGDC, are nonetheless binding on their GIS managers;
- Some datasets have inherited restrictions from third-party data sources involved in their preparation; and,
- Several participants expressed concern that the sheer convenience offered by a comprehensive compilation of high-quality geospatial data might provide significant, if not unique, assistance to unauthorized users.

While there is broad support in principal among BAR-HSDS participants for limiting access to their datasets to known and authorized users, a number of constraints have limited the actual deployment of such controls, including:

- Lack of a funded formal management framework to ratify security policies, administer security controls, and to investigate and adjudicate reported violations;
- Pre-existing designs and software that did not provide all the features required for the recommended security safeguards;
- Lack of experience with security technologies and practices among technical implementers; and,
- Lack of complete control of client devices (computers, PDAs, etc.), some of which will from time to time contain downloaded or cached subsets of the BAR-HSDS content.

As a result, the confidentiality goals set forth here are meant to be achievable, but in some cases may only be aspirational.

In addition, there is the threshold question of sensitivity when the geodatabase is "thin sliced" by access technologies. By limiting the geographic scope or resolution of queries, a degree of de-facto compartmentalization can be achieved that could mitigate the damage from a compromise of sensitive data.

In this regard the BAR-HSDS Viewer application can be seen as a valuable tool for security by limiting the amount of data a user can access at any one time to that visible in a Web Map Service image. (This is a rare instance where the interests of convenience and security are aligned rather than conflicting.) However, as other access tools such as other WMS or Web Feature Service clients are enabled, the risks associated with transferring large segments of the BAR-HSDS database to uncontrolled platforms will increase.

Confidentiality Goals for the BAR-HSDS

The following specific goals will lead toward a secure implementation of the BAR-HSDS:

Authentication

- All access to the BAR-HSDS, at any level of privilege, will be controlled on an individual user basis (i.e., no shared logins for groups, offices, agencies, etc.)
- All BAR-HSDS access will be ruled by the principal of "least privilege"; that is, that each user shall be granted only the minimum set of capabilities required to perform that user's required functions.
- Three distinct categories of users, at a minimum, will be identified:
    - System Administrators – Individuals with full administrative privileges on the BAR-HSDS servers, including "ETL" database maintenance functions;
    - Data Providers – Individuals associated with BAR-HSDS participating jurisdictions who are authorized to submit new and updated datasets for inclusion in the BAR-HSDS, and who may be permitted to download geospatial datasets from the BAR-HSDS; and,
    - Viewer Users – Individuals associated with participating jurisdictions and public safety and emergency management agencies, who may access the BAR-HSDS via the provided web-enabled user interface or other web map service applications.
- All users will be reviewed annually for current eligibility in accordance with BAR-HSDS policy.
- All passwords will meet specified standards for complexity and must be updated at least twice a year.
- All users will undertake a written agreement with the BAR-HSDS managing agency, committing the user to observe access and use policies and acknowledging that misuse or negligence may result in suspension or revocation of access as well as any other sanctions provided by law.

Access Control

- Each of the BAR-HSDS servers will be physically secured in a restricted space, and the hosting agencies will maintain positive control and auditing records of individual access to that space.
- Each hosting agency will be required to undertake a written agreement with the BAR-HSDS managing agency, committing the hosting agency to maintain site security and safety at agreed levels, and acknowledging that failure to do so may result in that site being removed from the BAR-

HSDS network.

- The BAR-HSDS servers will, at a minimum, be maintained in compliance with the relevant SANS "Top Ten" recommendations for operating system and application security, and also with the latest hardware and software providers' updates and recommendations.
- The BAR-HSDS servers will be installed behind a separate network firewall device, which at a minimum shall block Internet connections to any ports not required by the BAR-HSDS applications.
- Secondary data storage media, such as portable disk drives and backup tapes, will be physically secured in a locked space whenever not in transit, and will be enclosed in a locked container while in transit.
- All network communication of BAR-HSDS data, whether among servers or with client computers, including geospatial datasets, user account data and any other information related to the operation of the BAR-HSDS, will be encrypted at a level equivalent to 128-bit AES or TLS.
- The BAR-HSDS viewer will automatically purge any cached or temporary files on the client computer on exit, and this practice will be recommended for any other web map service applications accessing the BAR-HSDS servers.

Attribution and Assertion of Rights

- Details of the source, update history, copyright and any access restrictions of each data item derived from a participant-provided dataset or layer will be stored in the BAR-HSDS data structure.
- All applicable source, history, copyright and restriction information will be provided as metadata in any download of geospatial data from the BAR-HSDS.
- All applicable source, history, copyright and restriction information will be displayed in the screen of the BAR-HSDS viewer.

The following technical, procedural, personnel and organizational resources are suggested in order to achieve these goals.

Technology for Confidentiality

Simple Password Authentication

Simple authentication is generally achieved by requiring the user to demonstrate knowledge of a secret password, number or phrase. Since the security of this scheme relies entirely of the secrecy of that password, it is desirable that it be difficult to guess either at random, or based on any other knowledge

about the user (e.g., birthdate, children's names, favorite books/movies/meals, etc.) The best password, from this perspective, is one that is relatively long and essentially random.

The difficulty with such passwords lies in the psychological fact that humans have great difficulty remembering random patterns, especially long ones that change frequently. This leads to two potential problems:

- Users select easily-remembered but also easily-guessed passwords based on personal information, or,
- Users store their hard-to-guess/remember passwords in insecure aids to memory (e.g., in their wallets or on sticky notes in their desks.)

As a result, the design of simple password schemes involves striking a balance between the security and the usability of the passwords. A standard for relatively secure passwords might require that they be:

- Be a minimum of 8 characters and maximum of 14 characters in length.
- Contain at least 4 alphabetic, 2 numeric and one special character. There must be 2 upper case and 2 lower case alphabetic characters.
- Not contain any two identical consecutive characters.
- Be changed every 90 days.
- Not be identical to any of the previous 6 passwords.

However, such policies require considerable effort to maintain and enforce.

Two-Factor Authentication

Two-factor authentication requires presenting both of something the user knows (a password or PIN) and something the user has physically. A significant operational advantage of two-factor authentication is that the secret component can be allowed to be more predictable, and therefore easier to remember, since it is useless for impersonation without the physical factor.

A number of technologies are available for two-factor authentication. The most common is the hardware-token technique, where a password or PIN must be used in combination with a separate digital device to access the service. The fastest growing is the fingerprint-token

technique, in which a miniature scanner is used to examine the user's fingerprint. Other biometric techniques include iris (eye) scanning, voiceprint scanning, and keystroke dynamics analysis.

It would be attractive to require two-factor authentication at least for the system administration accounts and data providers. A trial package from CryptoCard to support one server with five users costs approximately $500. The well-known RSA SecureID system is significantly more expensive.

Network Link Encryption

Encryption adds several key capabilities to open networks. In addition to providing privacy, encryption can be used to verify the identities of both data sources and data recipients.

Although both proprietary and open-source implementations are widely available, the actual encryption techniques used are relatively few and well-known. The Advanced Encryption Standard (AES), which utilizes the Rijndael algorithm, was developed by the National Institute of Standards and Technology as a replacement for the earlier Data Encryption Standard (DES). Other widely-used encryption algorithms include triple DES (3DES), RC4 and Blowfish.

The Transport Layer Security (TLS) is a formalization by the Internet Engineering Task Force of the Secure Sockets Layer (SSL) protocol first devised by Netscape Communications Corporation to secure Web communications. TLS utilizes a variety of encryption algorithms and is provided as a built-in capability of most major operating systems.

The BAR-HSDS will user link encryption to protect the privacy both of the geospatial data and of the user identifiers and passwords used to access data and services.

Virtual Public Networks (VPNs)

VPNs use encrypted connections to extend trusted connectivity over untrusted networks. A VPN expands on the capabilities individual encrypted connections by supporting multiple simultaneous connections over a number of services and ports. In this way a VPN simplifies security administration by allowing a single identity realm and authentication process to control multiple applications.

The chief application of VPN technology in the BAR-HSDS would be to provide continuous trusted connections between the multiple servers. This approach simplifies the management of access controls and reduces the chance of

errors, oversights or inconsistencies. It also permits realtime/anytime connectivity for data backup and replication among the servers as the system's operational tempo accelerates over time.

Digital Rights Management (DRM)

DRM uses encryption technology to control access to digital assets such as music, films and other large files. Basically, DRM systems encrypt data and then control the dissemination of the keys needed to utilize them.

The DRM approach has been developed primarily for the protection of binary large objects ("BLOBs") such as music and video files, where the value resides in the unique arrangement of a large number of bits. DRM becomes less effective when the data are reassembled dynamically into new and unique configurations, as in response to a query from a geodatabase.

DRM also assumes a "cooperative client," that is, a client device that can be relied on to obey the strictures of the DRM framework by, for example, not making copies of the decrypted data. This is typically achieved by building DRM into the hardware or firmware of consumer devices such as DVD players. However, general-purpose computers can readily be programmed to circumvent DRM protections.

Digital Watermarking

Digital watermarks are distinctive patterns of tiny distortions to the content of a binary object, used to indicate the source of a file. This technique is directly analogous to the traditional mapping practice of introducing small errors or additional features to provide evidence in support of a copyright claim.

Digital watermarking does nothing to prevent unauthorized copying or distribution of data; it only provides evidence once a copy has been located. Also, the unique digital pattern of a digital watermark tends to "dissolve" when the dataset is merged with others in a geodatabase.

Digital watermarking could be applied to WMS images produced by the BAR-HSDS, but this would have little value without a mechanism for locating and examining possible copies. No direct equivalent is available for WFS outputs, other than the traditional one of inserting discrete fictitious features into the dataset itself.

Client-Side Data Protection

One potential for exposure of sensitive data in the BAR-HSDS framework arises from the presence of temporary, cache or other residual files (as well as deliberate and download and saves) on client computers. Since client platforms may not always be secured between uses with the BAR-HSDS, any sensitive data remaining on those computers could be compromised. Automatically generated cache and temporary files are particularly troublesome, since the typical user may not be technical sophisticated enough to realize that they exist or that they might contain sensitive data.

Two technologies have potential for improving the security of client-side data. The first is on-disk encryption, by which any data stored to disk by an application is stored in encrypted form. The other is "lost data destruction" as offered by Beachhead Solutions. This technique a special application to automatically destroy client-side data under specified circumstances, such as theft of a portable computer.

On-disk encryption is supported by most computer operating systems, but would require integration into the Viewer and any other client applications. Lost data destruction involves special software and configuration; one vendor offers this service for approximately $129 per user per year.

Another, perhaps less expensive safeguard would be to modify the Viewer to ensure that it automatically deletes any cache and temporary files that a typical user might not realize need to be treated as sensitive. This could be combined with a policy requirement that any other client applications provide the same safeguard, although enforcing that policy would be difficult.

Multi-Level Security

The initial BAR-HSDS implementation has adopted a simple all-or-nothing model for data access. However, data providers have expressed interest in a more sophisticated multi-level security scheme wherein different categories of users can access different portions of the geodata.

Implementing such a capability would require extending the capabilities of ArcSDE with a product such as St. George Consulting's "SecureMap" product. It might also require further extensions of the data model to bind specific access category values to individual features or layers.

Procedures for Confidentiality

Simplicity Principle

Closely related to the principle of least privilege, the simplicity principle holds that the fewer and less complex the security measures imposed, the less chance of error or unforeseen interactions there will be in their maintenance over time. (E.g., using a single VPN among the servers is simpler than maintaining distinct encryption arrangements for different types of replication.) Therefore, a general guideline is that all security provisions and procedures should be made as simple as possible.

Management and Enforceability

Security provisions are meaningless without an ongoing capability for their maintenance, and for evaluation of and response to security threats.

Maintenance tasks include:

- Review and approval or rejection of requests for user accounts;
- Security training of new users and refresher training for experienced users;
- Maintenance of access control databases (users and privileges);
- Regular review of security system performance and system usage patterns; and,
- Referral of possible security policy violations to the enforcement process.

Enforcement tasks include:

- Collecting information about possible security violations;
- Evaluating the reality and severity of suspected violations;
- Taking steps to mitigate the effects of any actual violations; and,
- Taking steps to prevent recurrence of violations.

In cases where violations of security policy involve acts of commission or omission by authorized users, additional training may be a sufficient response. However, the managing entity must have the authority to impose additional sanctions, such a restriction or revocation of access, in the case of wanton or repeated user security violations.

Regular Account Audits

One common vector for attack on computer security is the exploitation of expired, abandoned and/or compromised user accounts. In addition to ongoing maintenance and purging of

user accounts, an complete audit of user accounts should be conducted on a regular basis, at least annually and preferably quarterly, to verify that all account-holders are still authorized to access the system and are maintaining proper security (e.g., password changes).

Regular Activity Audits

Many security breaches can be prevented or at least detected through examination of system logs. In addition to any automated intrusion-detection or -prevention technology, a manual "spot check" of system logs should be performed by qualified personnel at least monthly.

Periodic Configuration Audits and Testing

A qualified security consultant should be engaged shortly after initial deployment, and periodically thereafter, to perform an independent verification of BAR-HSDS security provisions, possibly including "red team" penetration testing.

Backup Security

Backup media should be kept in locked storage accessible only to designated system-administration personnel.

Human Resources for Confidentiality

Designated Security Officers

Each host site and data provider should have a single designated BAR-HSDS Security Officer who is personally responsible for the maintenance and monitoring of security at that site and system-wide. (Each Security Officer should be allowed to designate an alternate in case their are unavailable, but should not be allowed to delegate or subdivide their personal responsibility for security.)

User Training and Documentation

Prior to being issued access credentials (passwords, tokens, etc.) all users should receive a security briefing from one of the site Security Officers to familiarize them with their responsibilities for secure use of the system and its data and to acquaint them with some common security threats (leaving workstations unattended, "phishing" attacks, etc.)

The content of the security briefing should be agreed upon by the Security Committee (as described in "Organization for Confidentiality," below) and all briefings should cover the same material (as appropriate for the level of access being

granted.)

At the conclusion of the security briefing the user should be required to sign a document acknowledging receipt and understanding of the information provided and agreeing to abide by system security policy as a condition of access.

All users should be required to attend "refresher" security briefings and update their signed agreements at least every two years.

(Note that this process has two goals: The first is to provide users with information and to secure their explicit agreement to policy. The second is as a "ritual" to demonstrate and dramatize the seriousness of security for the BAR-HSDS.)

Independent Security Auditor

No matter how careful, every security implementer is prone to unconscious assumptions and "blind spots." Therefore, a qualified independent auditor should be engaged from time to time to review and verify the security provisions of the BAR-HSDS.

Organization for Confidentiality

Managing Entity

The reference point for all of system security is the Managing Entity, the organization (or individual) that sets, implements and enforces BAR-HSDS security policy. Without an active Managing Entity with a mandate and the resources to maintain the security program, any investment in technology, procedure and human resources are ultimately futile.

Responsibilities of the managing entity include:

- Adopting and updating BAR-HSDS security policy;
- Undertaking operating and access agreements with hosts, data providers and other users;
- Funding of ongoing expenses for security system maintenance, including software licenses, encryption certificates and professional services;
- Policy oversight of the Security Committee; and,
- Adjudication of reported security violations and imposition of sanctions in accordance with policy. (Some of this authority may be delegated to the Security Committee with the Managing Entity acting as a board of appeal.)

Security Committee

The Security Officers from all sites should convene regularly

as a Security Committee to share information and to report to and advise the managing entity on security-related activities, requirements and events.

Security Policy

The Managing Entity, advised by the Security Committee, should adopt and publish a security policy, which will be referenced in all hosting, user and mutual-aid agreements.

Section II - Integrity

Definition of "integrity"

As with "confidentiality" the term "integrity" has acquired various shadings of meaning. We use the word in the following sense:

> **Integrity**: Assurance that specified information retains its original content, without unauthorized or undocumented changes.

Note the implication that all additions to, deletions from or changes to an item of data must be documented. A stricter requirement can be called "auditability," under which the state of a data item can be reconstructed ("rolled back") to any previous values; this requires a more detailed record (audit trail) than a mere change-recording requirement.

Note also that this definition of "integrity" is scale-independent in that it can be applied to an entire database, to an individual file, or to a particular data element or attribute.

In some cases a distinction may be drawn between "representational integrity," under which the information retains its format, and "functional integrity" under which the meaning is retained even if the format changes. E.g., as shape files are loaded into a geodatabase, they lose representational integrity (because their format is changed) but the functional integrity of the information is retained (assuming the integrity of the extraction and loading process.)

That last caveat introduces one other specialized sense of the word "integrity" when used to describe a data transformation process. In that extended sense the word means that the output of the transformation accurately reflects the inputs and the intended transformation function. Perfect integrity in that sense would imply that the transformation could be reversed to reproduce the precise original inputs. Such perfection is rarely achievable in actual operations.

Again, it is vital to bear in mind that security is always relative, never absolute.

Integrity Requirements and Constraints

The basic operational criteria of functional integrity for the BAR-HSDS are:

- Data received by clients from the server will accurately reflect the current state of the geodatabase;
- The geodatabase will accurately reflect the most recent and/or best data supplied by the data providers; and,

- Operational data entered during an emergency or exercise will be available to other authorized users accurately as entered.

Meeting these requirements is primarily the task of the BAR-HSDS application software and the geodatabase edit-transform-load (ETL) procedures, which are beyond the scope of this report.

In addition, "man in the middle" attacks whereby data is intercepted and modified in transit must be prevented both during data upload, during geodatabase maintenance and storage, and during query operations.

Special integrity issues arise from the unique design of the BAR-HSDS as a distributed, high-reliability service intended to support public safety, emergency management and homeland security applications. The purpose of deploying more than one server site is to provide continuous service even during disasters. Therefore, maintaining the integrity of the service as a whole requires rapid, accurate and consistent replication of data among the sites.

Integrity Goals for the BAR-HSDS

The following specific goals will lead toward a secure implementation of the BAR-HSDS:

- All instances of the geodatabase will reflect the same view of the latest data sets made available by the data providers;
- All instances of the geodatabase will reflect the latest incident data provided by users, so that users can switch to another server should their initial server site or its telecommunications be disabled;
- Data providers will have the ability as well as the responsibility to verify that the geodatabase reflects the latest and/or best data sets they have provided.
- All updates to the geodatabase and any operational (incident) data entered into the system will be logged and auditable.

Technology for Integrity

The primary technologies for integrity, other than the geodatabase, application and replication software, are the same encryption, authentication and access-control mechanisms addressed under Confidentiality.

Procedures for Integrity

In addition to the ETL and replication procedures provided by the BAR-HSDS applications, data providers should be encouraged to "spot check" the currency and accuracy of the geodatabase layers for which they provided data. These checks should occur immediately after updates and from time

to time between updates.

The replication of base data sets should be synchronised so that the data available to users from all servers is the same at any given time.

Any incident data should be replicated to all servers immediately upon posting.

Human Resources for Integrity

Data providers will need to be educated as to the procedures for providing data updates and for conducting data integrity checks after updates and from time to time.

Client users will need to be educated as to the procedures for "fail over" should the server they are using become unavailable, and also the trouble-reporting procedure should they become aware of a replication failure or other data integrity problem.

Organization for Integrity

The Managing Entity should set standards and secure resources for user education regarding integrity issues.

The Managing Entity should create a receiving, tracking and response mechanism for integrity problem reports from users.

Section III - Availability

Definition of "availability"

Availability is the apparent functioning of the system:

> **Availability**: Assurance that the information and functions of the system will be available to authorized users when and where needed in accordance with the system's design.

Availability entails a variety of design choices aimed at ensuring that the system's design and construction can achieve its goals. It also involves a variety of defensive measures to avoid loss of data or of access to data due to adverse conditions or events.

Availability is the easiest of the three aspects of security to quantify, at least in statistical terms. However, as with confidentiality and integrity, availability can never be a perfect 100%. The challenge, as always, is to choose wisely among the available trade-offs.

Availability Requirements and Constraints

The primary rationale for deploying multiple, geographically distributed servers for the BAR-HSDS is to enhance the availability of its data and services, even under adverse or even disastrous circumstances. Threats to availability of the BAR-HSDS may be posed by natural, accidental or deliberate forces and could include:

- Physical destruction of or damage to equipment at one or more server sites;
- Disruption of network connectivity between users and one or more server sites;
- Disruption of operating system, application software or data at one or more server sites (e.g., by a virus or worm);
- Denial of service by means of cyber attack on one or more server sites or on a shared network service required for operation of the system (e.g., DNS); and,
- Loss of power to one or more server sites or to critical network facilities serving one or more server sites.

Availability Goals for the BAR-HSDS

The following specific goals will lead toward a secure implementation of the BAR-HSDS:

- All data provider and client accounts will be accessible though any of the servers, even if it is isolated from the rest of the BAR-HSDS cluster or other network services;
- All necessary functions of the BAR-HSDS will be supported at

all server sites. (This does not preclude an operational designation of one site as "primary" for ETL or other administration purposes; however, all sites should be capable of immediately assuming the primary role if necessary);

- All server sites will be evaluated for physical security, seismic and fire safety, power backup and network security. Each site host agency will agree in writing to maintain the agreed-upon level of site security and preparedness;
- All software will be maintained at current patch levels (although non-critical updates may be "staggered" among the sites to allow evaluation of any side-effects of the updates before applying them to all sites); and,
- While third-party network services such as DNS and NTP may be utilized for convenience, a temporary failure of such services will not substantially impair the usability of the system.

Technology for Availability

Firewalls

A first line of defense against a variety of possible denial-of-service and other network attacks is the basic address and port filtering capabilities of a network firewall. The BAR-HSDS relies primarily on the site hosts' own firewalls, with a "least privilege" configuration that permits traffic to the minimum number of necessary ports and may restrict some particular ports to access from pre-designated Internet addresses.

Intrusion Detection and Prevention

Software intrusion detection systems (IDS) and intrusion protection systems (IPS) act as "intelligent firewalls" that can signal (in the case of IDS) or block (in IPS) network traffic that matches certain "signatures," i.e., certain recognizable patterns of activity that indicate an attack.

The challenge of IDS/IPS administration is to develop a set of rules that will minimize both false-positive (alarms or blockages of legitimate traffic) and false-negative (permitting inappropriate traffic) types of error. This can entail relatively laborious analysis of server traffic over time.

The Managing Entity may wish to evaluate IDS/IPS products for future upgrades, but at present the BAR-HSDS has limited resources for managing such technologies.

Virus Protection

Email is the chief vector for virus and worm propagation, and the BAR-HSDS servers will not process email (except possibly for generating email notifications of alarms and other unusual events,) so the exposure of the BAR-HSDS servers to

such "malware" is relatively limited. Nonetheless, there is a risk of infection through other means (web and database server attacks, data backups, unauthorized software loads, etc.) Therefore, the installation and maintenance of virus-scanning software on all servers is recommended, along with an aggressive policy of applying software patches when issued.

Technological Diversity

Both software and administration constraints have made it necessary to build all the BAR-HSDS servers using the same hardware and software. While this provides efficiencies, it also creates shared vulnerabilities. Over time the BAR-HSDS Managing Entity may wish to investigate whether it can diversify its technology platform, especially the underlying operating system(s), to reduce the possibility that a single shared vulnerability (also known as a "class break") could be exploited to bring down the entire system.

Procedures for Availability

Server Monitoring and Reporting

The BAR-HSDS is designed as a distributed system to provide high availability during emergencies. However, its day-to-day workload may not be sufficient to ensure that malfunctions or other problems will be detected. An application-level "heartbeat" system could be devised whereby each server site automatically checks on the availability of the others and reports any outages; such monitoring could also be performed by a completely separate system.

In any event, a mechanism will be needed for capturing and tracking trouble reports, whether from automated monitors or from system users. A web-based reporting form could be used for this purpose, as could an email address or a voice mail box. Responsibility for monitoring and acting on such reports will need to be determined by the Managing Entity.

Configuration Control

The software and settings of the individual servers could begin to diverge over time due to site-specific practices. This could eventually lead to significant differences in system behavior and maintenance or restoration procedures. Any departures from the as-delivered configuration of the systems should be documented on site and also reported to the Managing Entity.

Coordination of Maintenance and Updates

Information about planned and unscheduled work on any of the server sites should be shared with the system administrators at all sites. Non-critical software upgrades and patches should be coordinated among the sites and "staggered" so the "pioneer" sites can verify that no roll-back will required prior to applying changes to the rest. This will require ongoing communication among the system administrators.

Periodic Site Re-Surveys

At least every two years for all server sites, and as soon as possible after any substantial changes at any individual site, an independent inspection should be made and a written report files to verify that the security, safety and preparedness of the site has been maintained.

Human Resources for Availability

The system administrators at the host sites are the primary human resource for maintaining system availability.

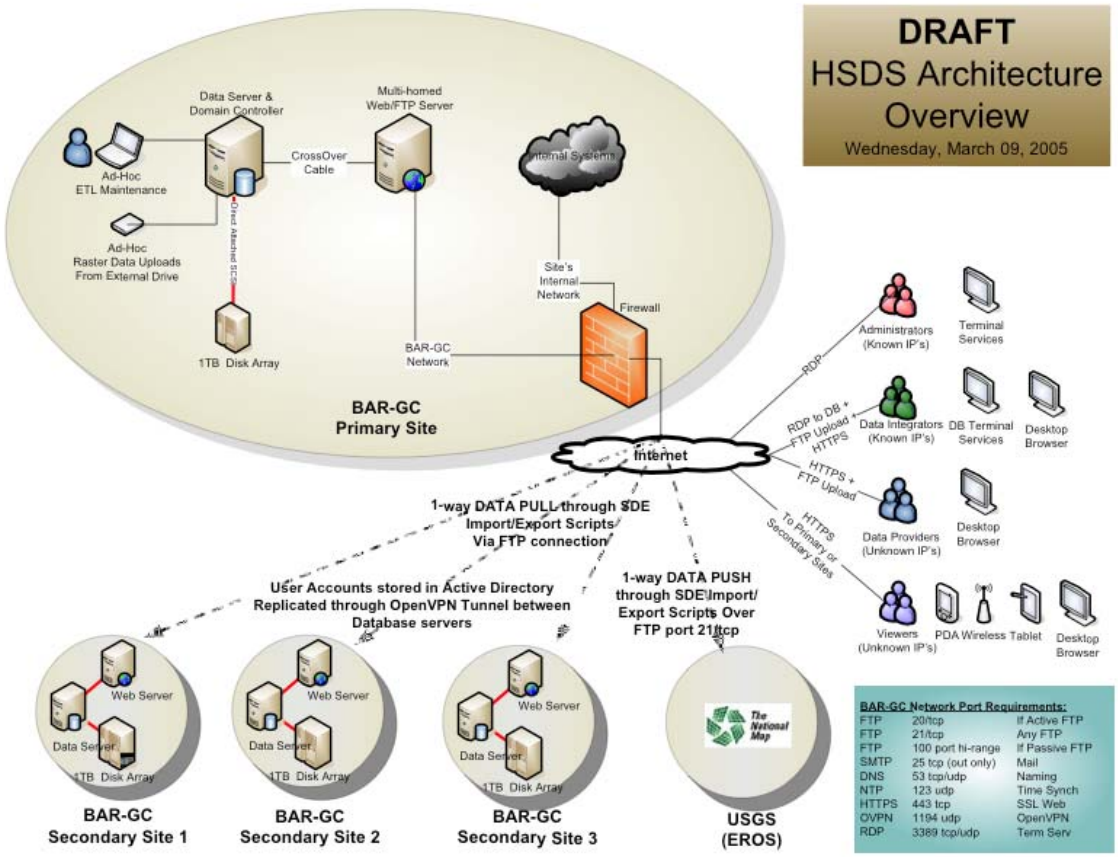
Organizing for Availability

The Managing Entity will have responsibility for:

- Monitoring the implementation of these availability provisions;
- Providing a point of contact for trouble reporting and a tracking system for availability and performance issues;
- Maintaining all necessary software licenses;
- Establishing and maintaining a mechanism for coordination of activities among the various site system administrators;
- Resolving any disputes regarding maintenance policy and server configuration control; and,
- Maintaining documentation of all configuration changes and making that documentation available to system administrators. (This information should be treated as Sensitive But Unclassified and access restricted to designated system administrators and technical contractors.)

Appendix A – "As-Built" Diagram

The following is a schematic overview of the BAR-HSDS architecture as implemented in the first phase.



Appendix B – Data Access Categories

The following is a proposed scheme for multi-level access control developed by BAR-HSDS participants. Note that the initial BAR-HSDS deployment does not implement this scheme.

The following matrix identifies the User Type Restrictions (in the left-hand column) and Use Category Restrictions (in the top row) that the Participating Agency may specify for individual data layers, at the discretion of the Participating Agency.

USER TYPES \ USE CATEGORIES	(1) Internal Use Only	(2) Bitmap display via web	(3) Free distribution to third parties	(4) Free distribution to third parties via Internet
(A) Emergency Service Provider	A1	A2	A3	A4
(B) Government Agencies or their delegated agents	B1	B2	B3	B4
(C) Other Public or EducationalInstitutions	C1	C2	C3	C4
(D) Data Contributors	D1	D2	D3	D4
(E) Public Domain	E1	E2	E3	E4

Explanation of User Categories

1. Internal Use Only – Geospatial Data may be used only by the Licensee or its designated agents to conduct the operations of the Licensee
2. Bitmap display via web – Geospatial Data may be displayed on the Licensee's internet website in a raster display format only, but not made available for download
3. Free distribution to third parties – Licensee may distribute data to third parties at no fee. Licensee is restricted from distribution via the internet.
4. Free distribution to third parties via Internet – Licensee may distribute data to third parties at no fee. No restriction from Internet based distribution.

Explanation of User Types

A. Emergency Service Provider –*To be defined by Bay Area Regional Homeland Security Data Server (BAR-HSDS) Project Team*

B. Governmental Agencies or their delegated agents
C. Other Public or Educational Institutions
D. Data Contributors – Organizations or individuals that sign this MOU and contribute data to the BAR-GC data repository.
E. Public Domain – Any person or agency who requests access to the data

The User Type Restrictions and Use Category Restrictions at each level are inclusive of all less-restrictive levels. For example, a C2 equates to the following:

- Access granted to Emergency Service Providers (A), Government Agencies or Their Delegated Agents (B) and Other Public or Education Institutions (C), and
- Access granted for the above users to use the data for Internal Use (1) and Internet/Web based display (2).

Appendix C – Security Related Laws, Regulations, Policies and Guidelines

Source	Document	Description
Federal Geographic Data Committee (FGDC)	"Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns" (2004)	"...the guidelines provide a method for balancing security risks and the benefits of geospatial data dissemination. If safeguarding is justified, the guidelines help organizations select appropriate risk-based safeguards..."
FGDC	Content Standard for Digital Geospatial Metadata (1998)	"This standard is intended to support the collection and processing of geospatial metadata. It is intended to be useable by all levels of government and the private sector. The standard is not intended to reflect an implementation design..."
National Institute of Standards and Technology (NIST)	FIPS 140-1: "Security Requirements for Cryptographic Modules" (1994)	"...This standard specifies the security requirements that are to be satisfied by a cryptographic module. The standard provides four increasing, qualitative levels of security intended to cover a wide range of potential applications and environments..."
NIST	FIPS 190: "Guideline For The Use Of Advanced Authentication Technology Alternatives" (1994)	"This Guideline describes the primary alternative methods for verifying the identities of computer system users, and provides recommendations to Federal agencies and departments for the acquisition and use of technology which supports these methods..."
NIST	Special Publication 800-12: "An Introduction to Computer Security – The NIST Handbook" (1995)	"This handbook provides assistance in securing computer-based resources (including hardware, software, and information) by explaining important concepts, cost considerations, and interrelationships of security controls..."
NIST	Special Publication 800-19: "Mobile Agent Security" (1999)	"This report provides an overview of the range of threats facing the designers of agent platforms and the developers of agent based applications. The report also identifies generic security objectives, and a range of measures for countering the identified threats and fulfilling these security objectives."
Open GIS Consortium	GeoDRM Working Group Charter (2004)	"...This paper presents a business case and market analysis for geospatial digital rights management..."
RAND Corporation, National Defense Research Institute	"Mapping the Risks: Assessing the Homeland Security Implications of Publically Available Geospatial Information" (2004)	"...This report assesses the homeland security implications of publicly available geospatial data and information... Although the report mainly focuses on federal geospatial information sources, the analytical framework could be relevant to other decisionmakers and analysts at the state and local government levels, and those in the private sector, who deal with similar issues..."

SANS Institute (Andrew Helyer)	"Sensitive But Unclassified" (2002)	"...In this report, one will learn about the differences between classified and unclassified information. One will also learn about the many names by which sensitive information may be labeled. The history of the United States laws that affect the dissemination of sensitive information is addressed, and guidance is provided for identifying and protecting sensitive information..."

Appendix J

The National Map/Palanterra™ Synchronization System Architecture Design

NGA Homeland Security Enterprise GIS Support Project

The National Map/Palanterra™ Synchronization System Architecture Design

Prefinal National Geospatial-Intelligence Agency
March 2005

Prepared for:

Brad Lucas, Contracting Officer's Representative
National Geospatial-Intelligence Agency
4600 Sangamore Road
Bethesda, Maryland 20816-5003
Tel.: 301-287-6572

Prepared by:

ESRI
380 New York Street
Redlands, California 92373-8100

Document History

Revision History

Date and Sections	Description	Person Responsible
7/28/04	Initial Draft	Kate Taylor
7/29/04	Updates to Initial Draft	Kate Taylor
10/28/04	First Prefinal	Michael Young
11/2/04	Second Prefinal	David Kehrlein
3/25/05	Reference Architecture for Regional and State Hubs added (Section 4)	Michael Young

Table of Contents

Section **Page**

1.0 About This Document

1.1 Purpose

This prefinal system architecture document serves to describe *The National Map*/Palanterra™ Synchronization system architecture that ESRI will develop for the National Geospatial-Intelligence Agency (NGA) Homeland Security Enterprise GIS Support Project. It defines an architecture that is planned to evolve to satisfy the operational and technical goals and constraints of the project. This document is organized in accordance with ESRI's Quality Management System (QMS) and meets the requirements of that system.

The system architecture will evolve during the project to reflect lessons learned from the prototypes, pilots, and spirals planned for the project's three major tasks, which are data management, synchronization with *The National Map,* and Palanterra™ enhancements. This approach supports the program goals to incrementally deliver operational capabilities into a number of existing and new operating environments.

1.2 Intended Audience of This Document

This prefinal system architecture document is intended primarily for use by the NGA and USGS Homeland Security Enterprise GIS Support Project management team, their partner agencies, and the ESRI Homeland Security Enterprise GIS Support Project team.

1.3 Document Overview

The document is organized into five sections including two appendixes as summarized below.

- **Section 1—About This Document:** Provides an overview of *The National Map*/Palanterra™ Synchronization System Architecture Design document
- **Section 2—Project Homeland Overview:** Presents an overview of the stakeholders and the architectural framework utilized
- **Section 3— The National Map/Palanterra™ Synchronization Task:** Provides a detailed breakdown of the task's operations, systems, data and security
- **Appendix A—Glossary:** Defines important terms used in the prefinal system architecture document

- **Appendix B—Acronyms and Definitions:** Defines various acronyms used in the prefinal system architecture document

1.4 Assumptions, Dependencies, and Constraints

- The synchronization of NGA Palanterra and the *The National Map* will be implemented within the budget and schedule constraints of the program, and within the bounds of technical feasibility of a system being built on COTS technology. The implementation priorities will be negotiated between NGA and ESRI.
- One Pilot Study (BAR-GC) will be conducted for this task during CLIN 0001.
- A second Pilot Study has been identified (Colorado Springs) and will be scoped during CLIN 0001, for execution during CLIN 002.
- If additional Pilot Studies are defined and prioritized as a result of the workshops and partner agency requests, they will need to be defined and prioritized to fit within the budget and schedule constraints of the program.
- Any equipment purchased for use by partners under this contract will remain in place at the partner location for their continued use after completion of the task activities.
- The USGS and NGA will play an active role in each Pilot Study.
- NGA will identify large scale analysis requirements, and the USGS will work with ESRI to identify and implement mechanisms for collaborative efforts to provide the data to support them.

1.5 Referenced Documents

1.5.1 Project Documents

This section identifies documents referenced in the prefinal *The National Map*/Palanterra™ Synchronization System Architecture Design document as well as documents that will be referenced during the development of the architecture. These include project documents that are deliverables. External documents referenced by this document are listed in Section 1.5.2.

Document Name	Date	Document Owner
Technical Proposal for Homeland Security Enterprise GIS Support	May 10, 2004	Peter Bottenberg
Program Schedule for Homeland Security Enterprise GIS Support	May 28, 2004	Peter Bottenberg
Project Homeland Draft System Architecture	July 2004	John Ellis
The National Map/Palanterra™ Synchronization System Architecture Design—Draft	July 2004	Kate Taylor
Technical Proposal for Homeland Security Enterprise GIS Support, Revision 1	August 11, 2004	Peter Bottenberg
Project Homeland Pre-final System Architecture	September 2004	Michael Young
Project Homeland Data Management Plan, Final	September 2004	Charlie Wells

1.5.2 External Documents

- Institute of Electrical and Electronics Engineers, 2000, *IEEE Std. 1471-2000*. Piscataway, New Jersey: IEEE Computer Press.

- Wood, W., M. Barbacci, P. Clements, S. Palmquist, H. Ang, L. Bernhardt, F. Dandashi, D. Emery, S. Sheard, L. Uzzle, J. Weiler, and A. Krummenoehl, 2003, "DoD Architecture Framework and Software Architecture Workshop Report," Software Engineering Institute, Carnegie Mellon University.

- Department of Defense Architecture Framework Working Group, 2003 (August), "DoD Architecture Framework, Version 1.0, Volume 1: Definitions and Guidelines."

- National Imagery and Mapping Agency (NIMA) and United States Geological Survey (USGS). "*Homeland Security Infrastructure Protection TIGER Team Report"* Version 1.1, September, 2002.

- GIDI Site Document – Bethesda, CDRL #13, NGA, April 2004.

- National Oceanic and Atmospheric Administration (NOAA) submitted by Applied Geographics, Inc., "*Boston Preparedness Pilot Project: GeoSpatial Data Development in Support of Critical Infrastructure Protection",* September, 2003.

- ESRI white paper entitled *GIS Portal Technology,* June 2004.

- ESRI white paper entitled *Interoperability: ESRI's Support of OGC Specifications,* July 2004.

- ESRI white paper entitled *ESRI's System Design Strategies,* July 2004.

2.0 Project Homeland Overview

Project Homeland's primary goal is to improve and enhance the ability of the National Geospatial-Intelligence Agency to support homeland defense (HLD) and homeland security (HLS) customers with timely, relevant, tactical, operational, and strategic data products and services. The efforts conducted as part of this project will provide a proof of concept for the larger implementation of COTS-based solutions to support HLD and HLS missions.

Four primary objectives will be addressed as part of the NGA HLS enterprise GIS effort.

1. Utilize a rapid integration and implementation process to improve and enhance the existing NGA HLS Data Model.

2. Automate the sharing of spatial databases between the NGA Palanterra™ Systems and the USGS National Map system. The synchronization efforts will focus on (1) synchronization between state and local geospatial data sources and *The National Map* of the United States Geological Survey (USGS), enabled by the HLS Data Model and (2) synchronization of *The National Map* with NGA Palanterra™ Systems.

3. Improve and enhance the operational capabilities of Palanterra™, NGA's Web Based Geospatial-Intelligence Decision Support System (Analytical system, Visualization system, Dissemination system). These efforts will provide a proof of concept for the larger integration and implementation of the COTS-based solutions to support the homeland security and homeland defense missions.

4. Provide training and technical resources in support of objectives 1, 2, and 3.

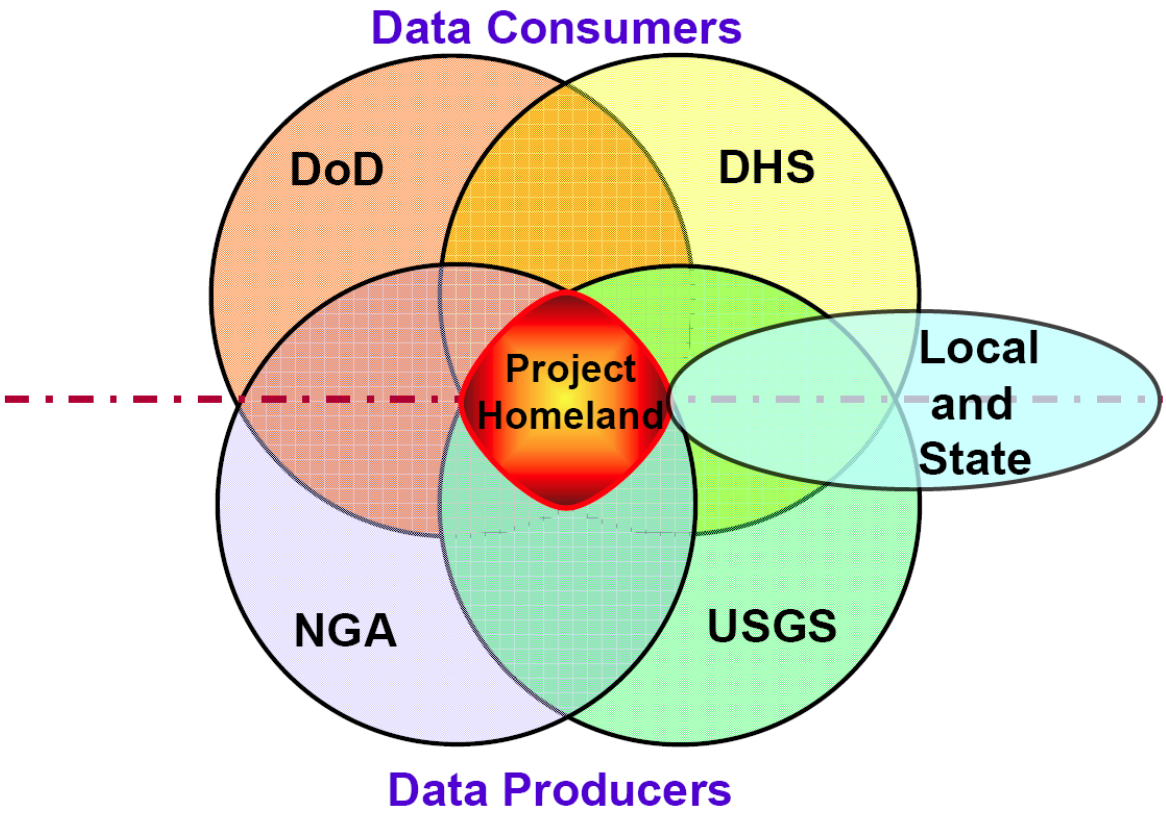
To accomplish these objectives ESRI will employ cutting edge commercial technology, undertaking five broad tasks supporting the four primary objectives to support the NGA HLS enterprise GIS effort.

This document focuses on the second objective listed above, *The National Map*/Palanterra™ Synchronization.

2.1 Stakeholders

Project Homeland's primary stakeholders during Contract Line Item Number (CLIN) 1—NGA, Department of Homeland Security (DHS), Department of Defense (DoD), and United States Geological Survey (USGS)—and their roles as data consumers or data producers are illustrated in Figure 2-1.

Figure 2-1
Project Homeland Primary Stakeholders (CLIN 1)



Each organization utilizes various applications to help distribute its data. For this project, the organizations will utilize systems as follows:

- **NGA—**Palanterra™—Views, collects, and distributes data
- **USGS—***The National Map*—Centralized baseline of national maps
- **DHS—**COMFORCE—Primarily consumes data from NGA and USGS

- **DoD—**NorthCOM—Primarily consumes data from NGA and USGS

- **Local and State—**Bay Area Regional GIS Council (BAR-GC)—Provides detailed data and consumes USGS data from The National Map.

There are currently three pilot programs for this project within CLIN 1. Two of the pilots focus on *The National Map*/Palanterra™ Synchronization, and the COMFORCE pilot focuses on interoperability improvement with DHS Geospatial Enterprise Architecture.

- *The National Map*/Palanterra™ Synchronization Pilots (BAR-GC, Colorado Springs)

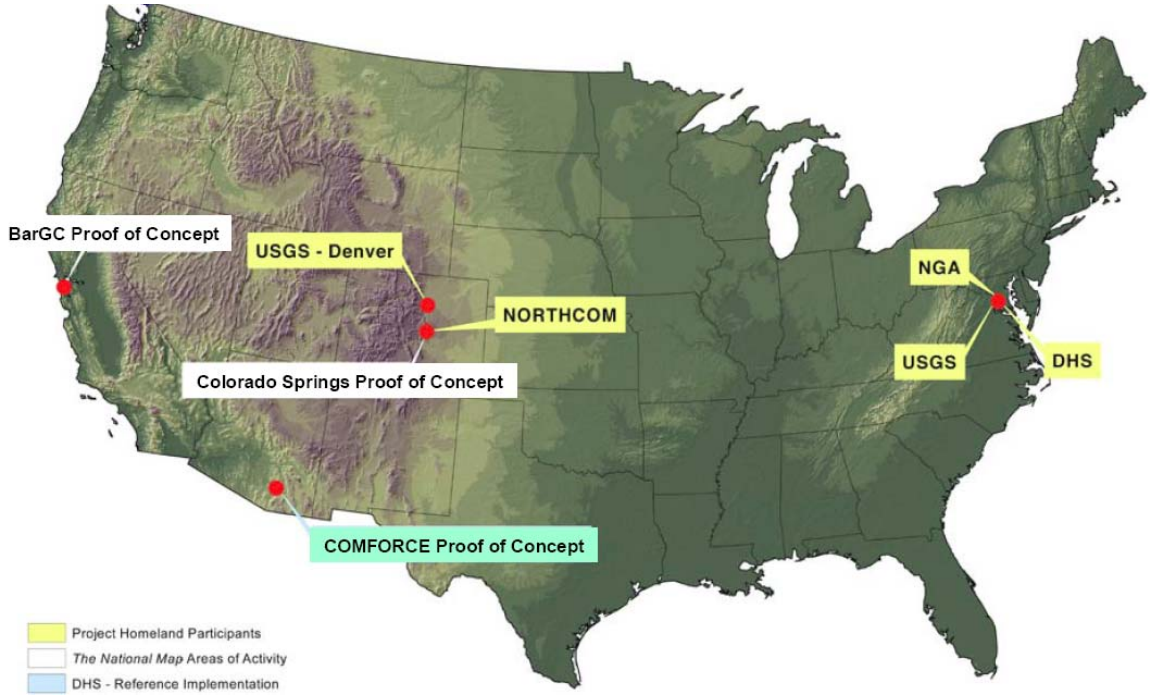
  - These two pilots are based on exploring and elaborating the technical and policy issues associated within the data synthesis between state, local, and national systems and in the collaboration/communication between participating agency databases (*The National Map* from USGS and NGA's Palanterra™).

- Interoperability improvement with DHS Geospatial Enterprise Architecture (COMFORCE)

  - This pilot involves the customization and implementation of NGA's Palanterra™ for DHS. COMFORCE is an application based on Palanterra™ that will be utilized by the Arizona Border Patrol for improving interoperability and compliance with the developing DHS Geospatial Enterprise Architecture.

The locations of these proof of concept pilots and the primary stakeholders are shown in Figure 2-2.

Figure 2-2
Project Homeland Participants and Pilots



There are currently two additional pilot programs scheduled for CLIN 2.

- New York
- Boston

2.2 System Concept

The foundation of Project Homeland is the concept of providing end-to-end, Web-enabled, and services-oriented geospatial intelligence between multiple agencies as seen in Figure 2-3. Further details of the services and interface layers utilized are shown in Figure 2-4.

Figure 2-3
Project Homeland Concept View

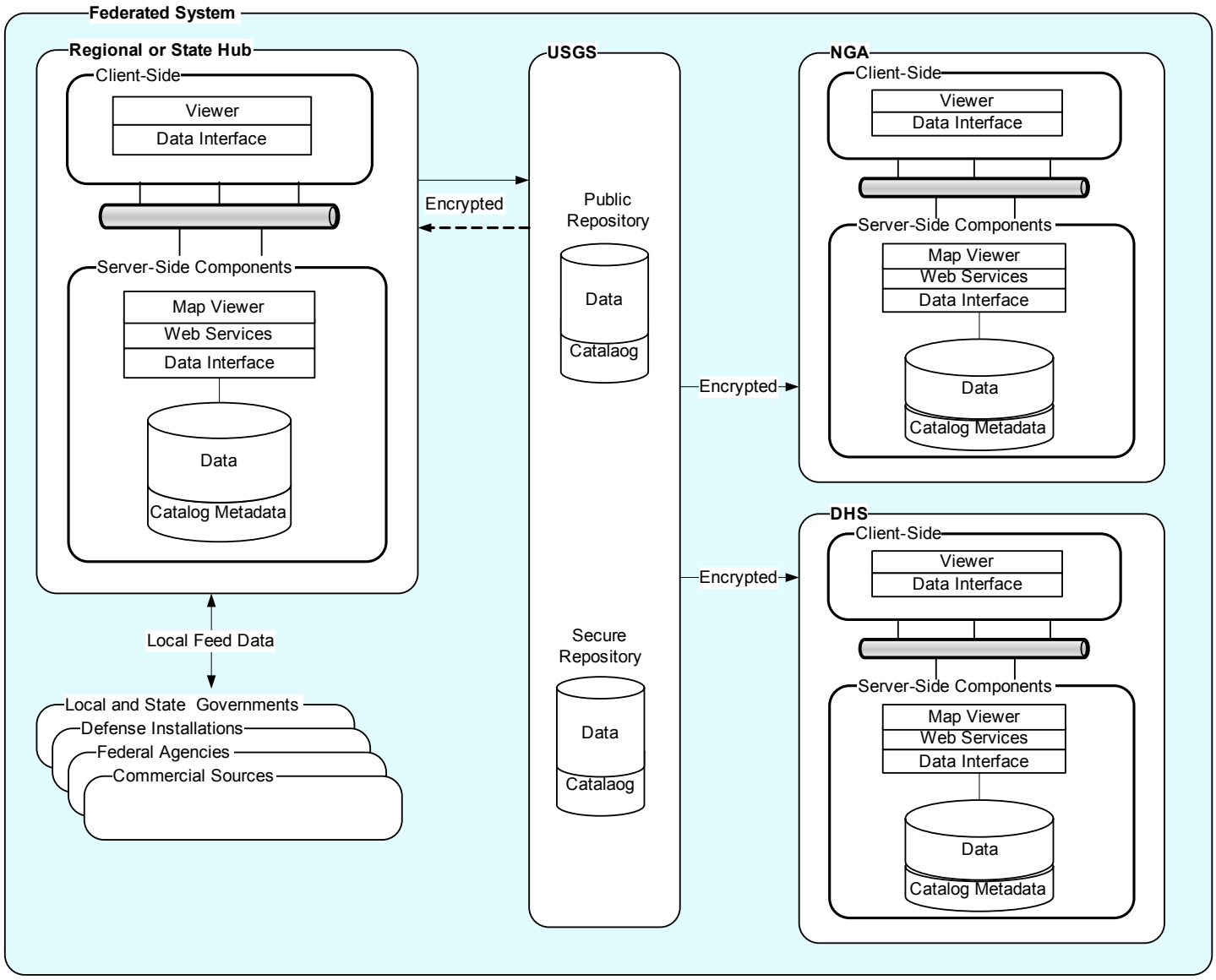
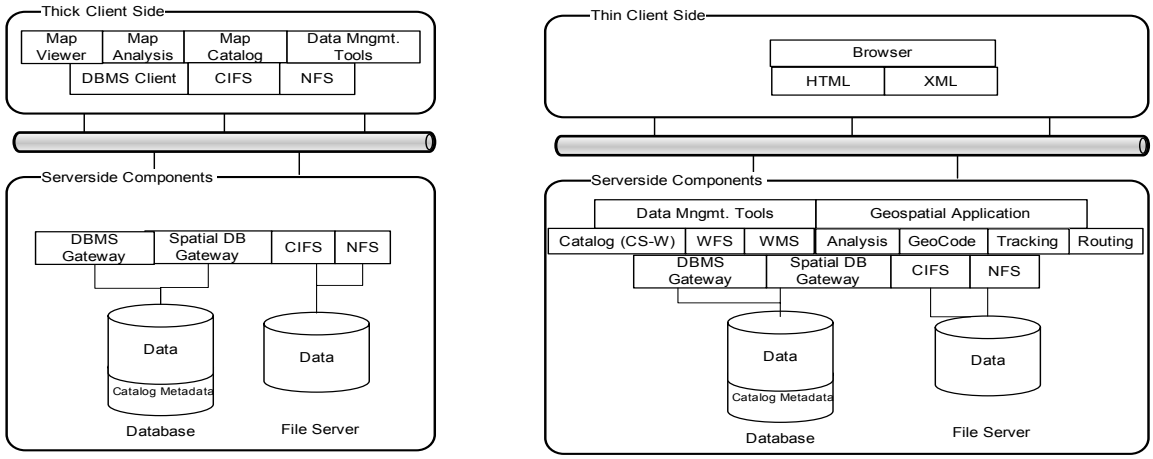
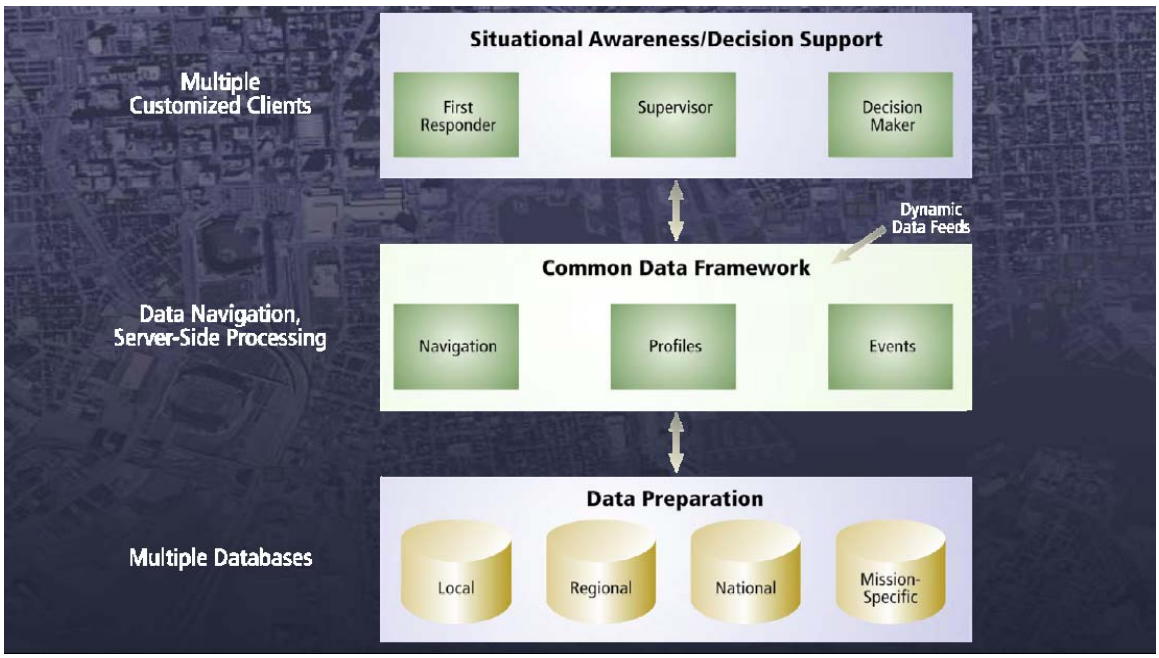


Figure 2-4
Overview of Server and Clientside Components



At a conceptual level there are three tiers in the system, which are illustrated in Figure 2-5 below.

Figure 2-5
Concept View of Information Flow



The top tier (Situational Awareness/Decision Support) is a set of customized interactive geospatial presentations that will be designed to present the information a particular type of user needs for either situational awareness or decision support. These users will be operators, analysts, managers, or decision makers.

The middle tier (Common Data Framework) organizes multiple data sources into a common data framework that feeds the upper (presentation) tier. The goal of this framework is to enable multiple coherent views of the current situation. Since different users require different information, there will not typically be just one common operating picture (COP), but multiple pictures, each filtered and presented to meet the information needs of its intended audience. The framework will also provide capabilities to navigate through the system's multiple data sets to quickly discover relevant data. Real-time data feeds, such as events or tracking data, will be processed at the common data framework tier and the results made available to the presentation tier. This common data framework is a core component of providing a federated services-oriented architecture (SOA). The *Project Homeland System Architecture* document contains further details on the SOA, and what types GIS web services are available.

The lower tier (Data Preparation) consists of the many data sources for the system. These include *The National Map* from USGS, NGA's own data holdings, data from other agencies that has been supplied to NGA to support this mission, and mission specific data that is held by partner agencies.

3.0 *The National Map*/Palanterra™ Synchronization Task

The National Map/Palanterra™ Synchronization task's primary goal is to improve and enhance the availability of geospatial data to support NGA in their support of Homeland Defense (HLD) and Homeland Security (HLS) customers with timely, relevant, tactical, operational, and strategic data products and services. This objective will be accomplished through several proof-of-concept activities that test and demonstrate key components of a future, more comprehensive implementation of COTS-based solutions and *The National Map* to support HLD and HLS missions.

The Synchronize NGA Palanterra and The National Map of the USGS task has three main objectives:

- Support NGA and the USGS in the data synthesis between state, local, and national systems
- Support NGA and the USGS in the data collaboration and communication between the two agency databases
- Collaborate with NGA and the USGS in reviewing and updating the maintenance strategy for The National Map database based on the needs of the NGA

The National Map/Palanterra™ Synchronization architecture will evolve in parallel with the development work on the project. This prefinal document is written to capture only the highest level of detail, so that agreement may be reached between the stakeholders before more detailed design work is undertaken.

3.1 Operations

Operationally, *The National Map*/Palanterra™ Synchronization activities are focused on the processes for creating and maintaining *The National Map* and fulfilling special requests for data to support the operational use of Palanterra™.

There are four main role types within Project Homeland:

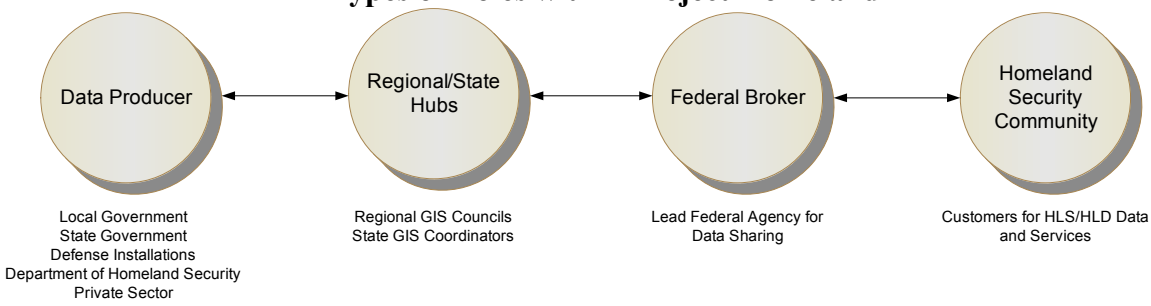
- Data Producer
  - Create data within local jurisdictions and then made available through Web Services, maps and datasets
- Regional/State Hub
  - Consolidates data from the data producers

- Federal Broker
  - Limits the need for every federal agency to go to every state to the information it needs to fulfill HLS/HLD responsibilities

- Homeland Security Community
  - Utilize information similar to local first responders

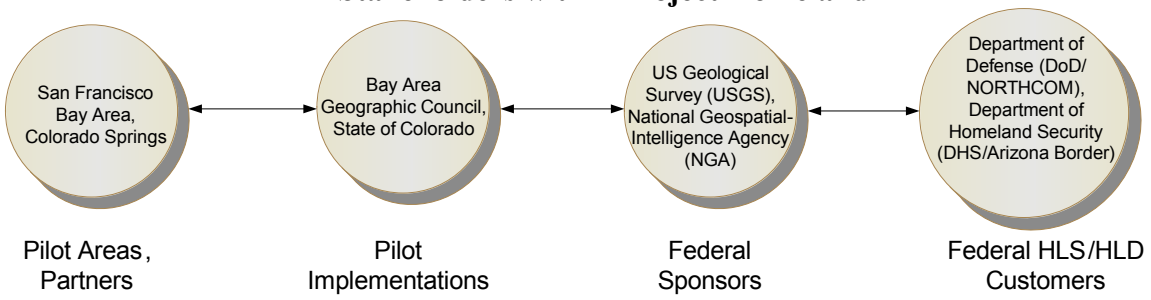
The relationship of dataflow between the types of roles is shown in Figure 3-1 below.

Figure 3-1
Types of Roles within Project Homeland



Project Homeland's primary stakeholders that fit each of these roles and their relationships to each other are illustrated in Figure 3-2 below.

Figure 3-2
Stakeholders within Project Homeland



The USGS acts as a liaison to supply state and local data to members of the HLS/HLD community in support of their missions. In order to support this role, USGS has worked with NGA to instantiate an instance of *The National Ma*p data holdings within the NGA classified systems. Additionally, the USGS receives requests from NGA for data to support a specific event or mission. USGS routes these special requests for data to their local partners and acquires the available resources for provision to NGA.

To describe the required processes for building and maintaining *The National Map* and sharing these data resources with the HLS/HLD community, a series of cross-functional (swim-lane) flow diagrams are included in this section. These charts document the current (As-Is) processes for synthesis and collaboration and identify a long-term vision for how the current processes might be improved to better support the HLS and HLD missions. A set of intermediate processes will be implemented through *The National Map*/Palanterra™ Synchronization spiral activities. The intent is to evaluate the options for incrementally reengineering the existing synthesis and collaboration processes.

The following cross-functional charts are provided below:

1. As-Is synthesis activities for creating and maintaining *The National Map*

   This chart describes the current processes for USGS to create and maintain *The National Map*. USGS has also been providing a copy of *The National Map* data to NGA for use within their network architecture.

2. As-Is synthesis and collaboration activities for fulfilling special data requests from NGA

   This chart describes the current processes for NGA to acquire additional data from local partners, with USGS acting as a liaison. This information was developed through discussions with USGS Menlo Park.

3. As-Is collaboration activities for installing and maintaining *The National Map* at NGA

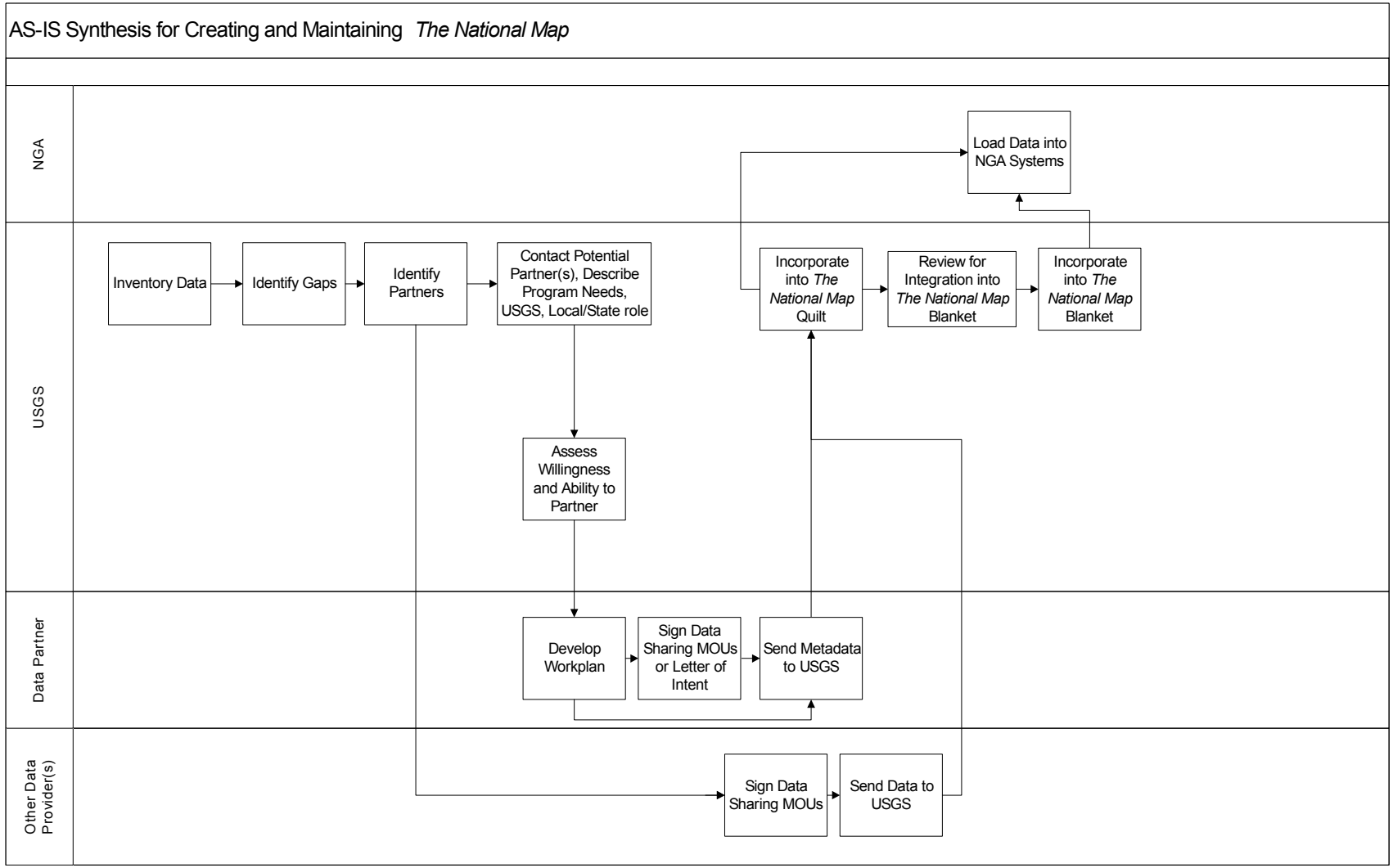
   This chart describes the current processes for creating and maintaining *The National Map* instance at NGA. Note that this is an elaboration of the last process described in chart (1) above.

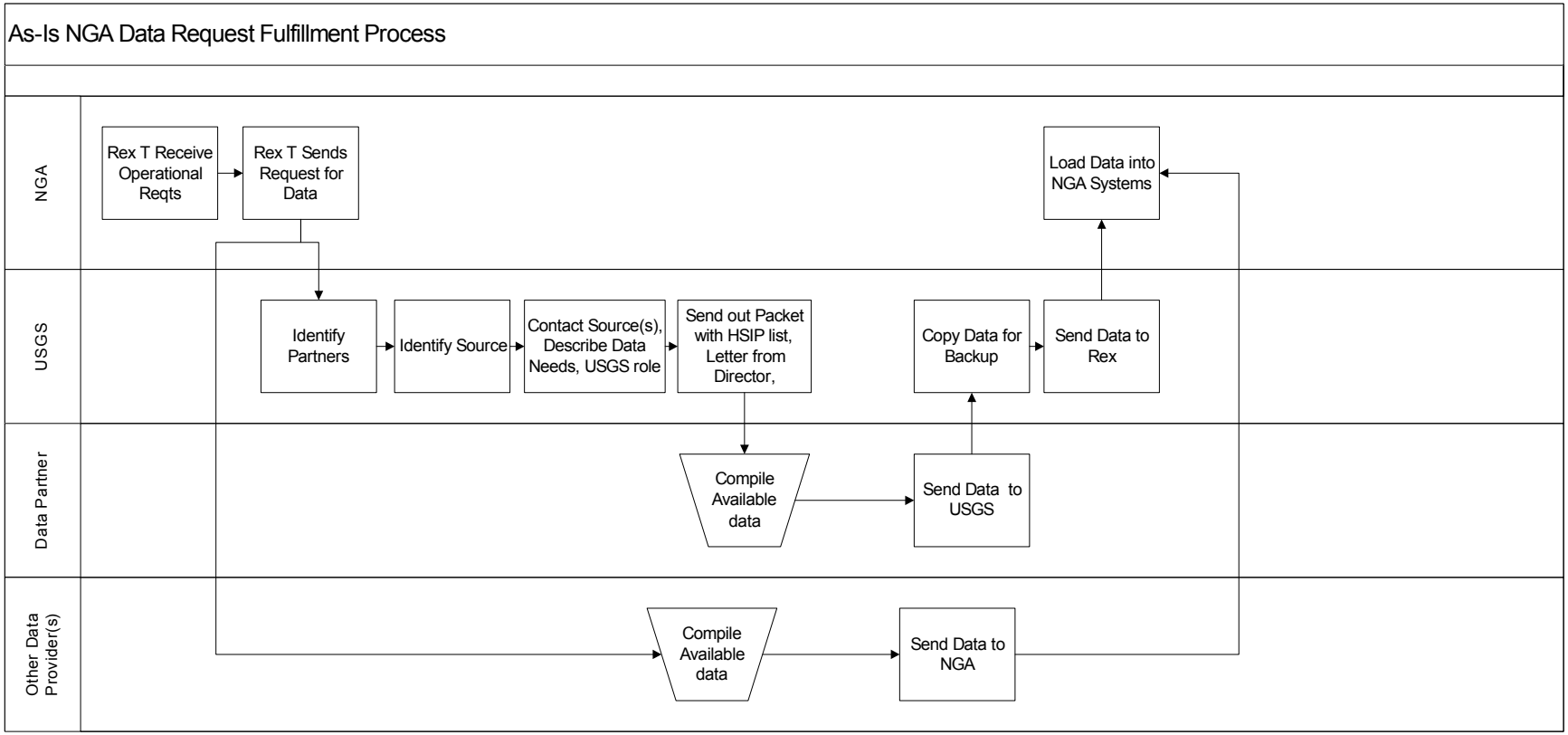
4. Intermediate synthesis and collaboration activities to be evaluated through the Bay Area Regional GIS Council (BAR-GC) spiral

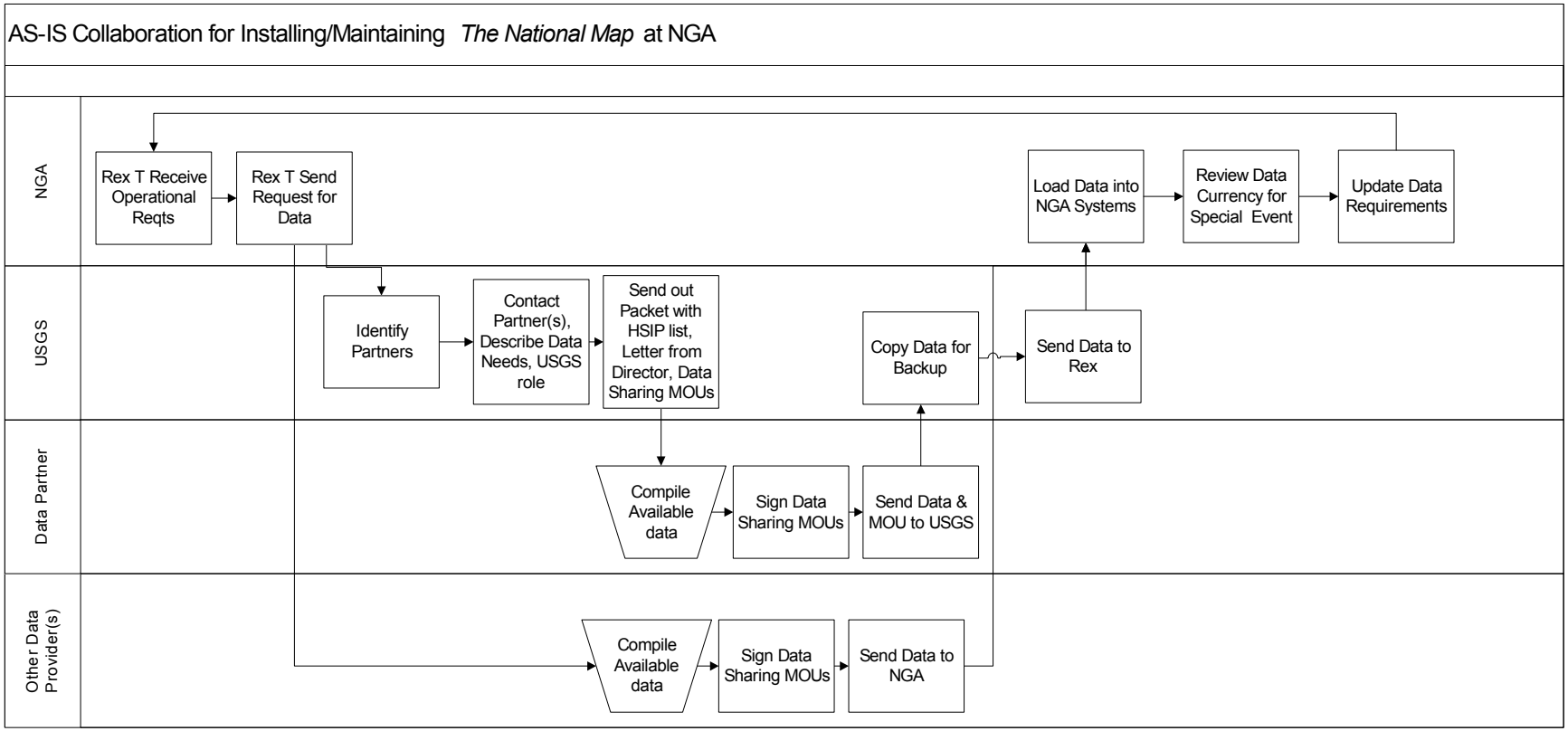
   This chart describes the proposed processes for improving and automating the current data fulfillment activities. The tools and architecture to support these processes will be implemented during the BAR-GC spiral activities.

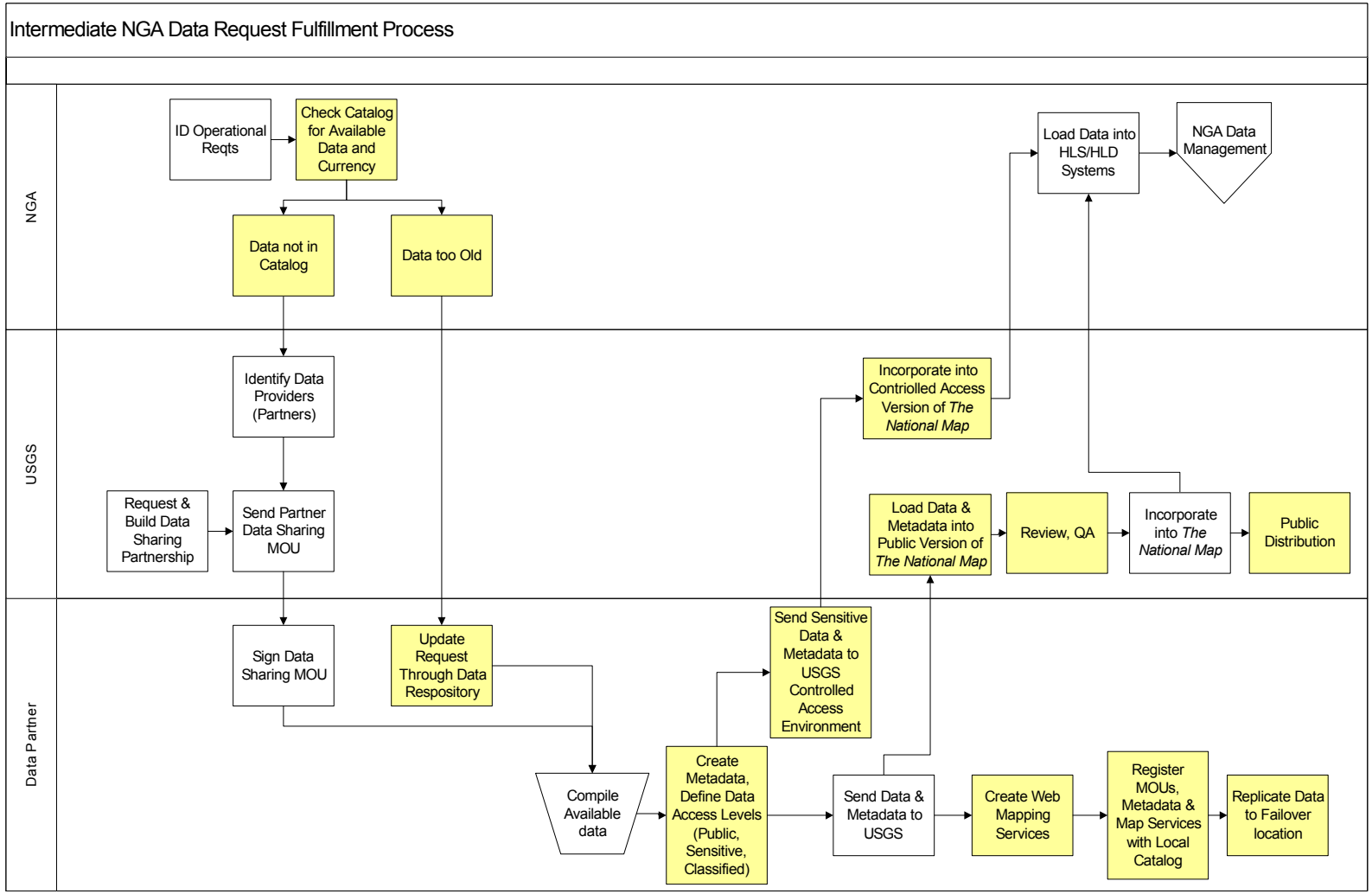
5. Long term synthesis and collaboration activities in support of HLS/HLD activities

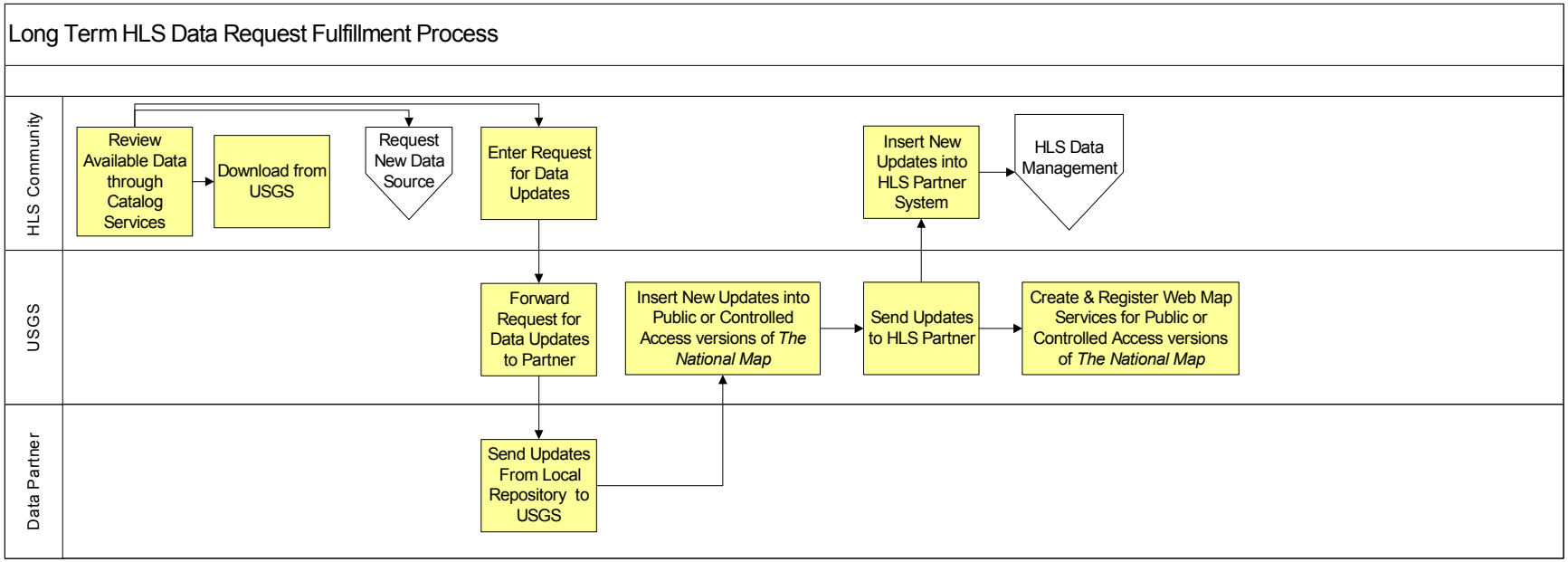
   This chart proposes a long term solution for supporting and automating HLS/HLD data needs in collaboration with USGS.











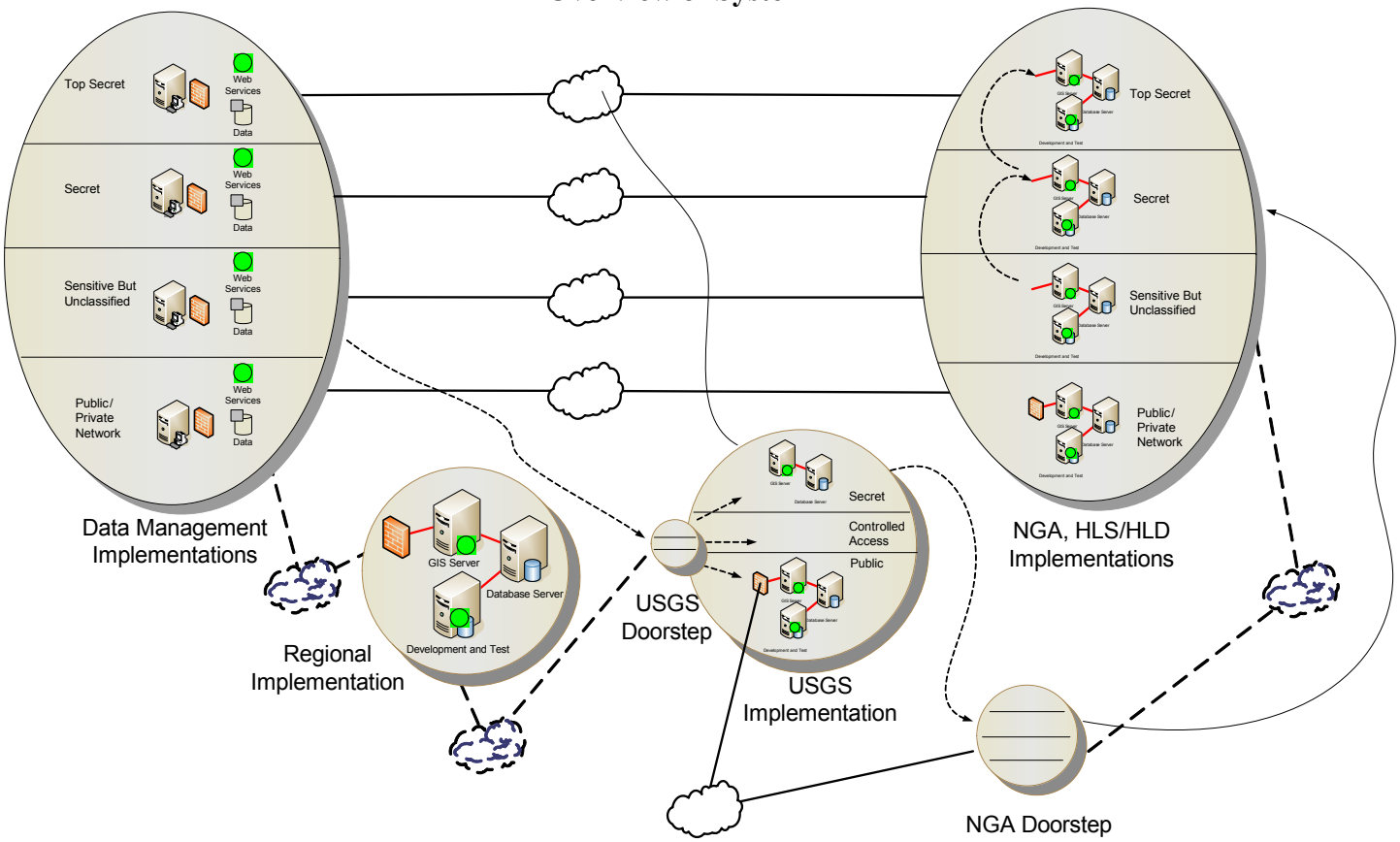
3.2 System

The system resources associated with The National Map/Palanterra synchronization activities include the local partner elements, USGS elements and the HLS/HLD elements required to support the data synthesis and collaboration activities.

There are four major computational elements in the system, as illustrated in Figure 3-3:

- Data Management (Local) Implementations
- Regional Implementations
- USGS Implementation
- NGA, HLS/HLD Implementations

Figure 3-3
Overview of System



3.2.1 Data Management (Local) Implementations

This part of the system consists of data producers who are participants in *The National Map* pilots for The NGA Homeland Security Enterprise GIS Support Project. Data collected by these local participants are rolled up to their corresponding Regional Implementation.

3.2.2 Regional Implementations

There are currently two regional pilots facilitating research on The National Map/Palanterra synchronization, BAR-GC and Colorado Springs/State.

3.2.2.1 BAR-GC

BAR-GC will synthesize (load) data from the local producers to implement two regional spatial data warehouses that will host the data from the participating counties and provide failover and redundancy. A GIS portal will be implemented to facilitate secure data sharing and data discovery for the local participants and first responder community.

The BAR-GC Pilot establishes the Bay Area Regional Homeland Security Data Server (BAR- HSDS). This consists of redundant servers in several locations in the San Francisco Bay Area. Each server will be capable of providing critical geospatial information region-wide for the homeland security community, at all levels of government.

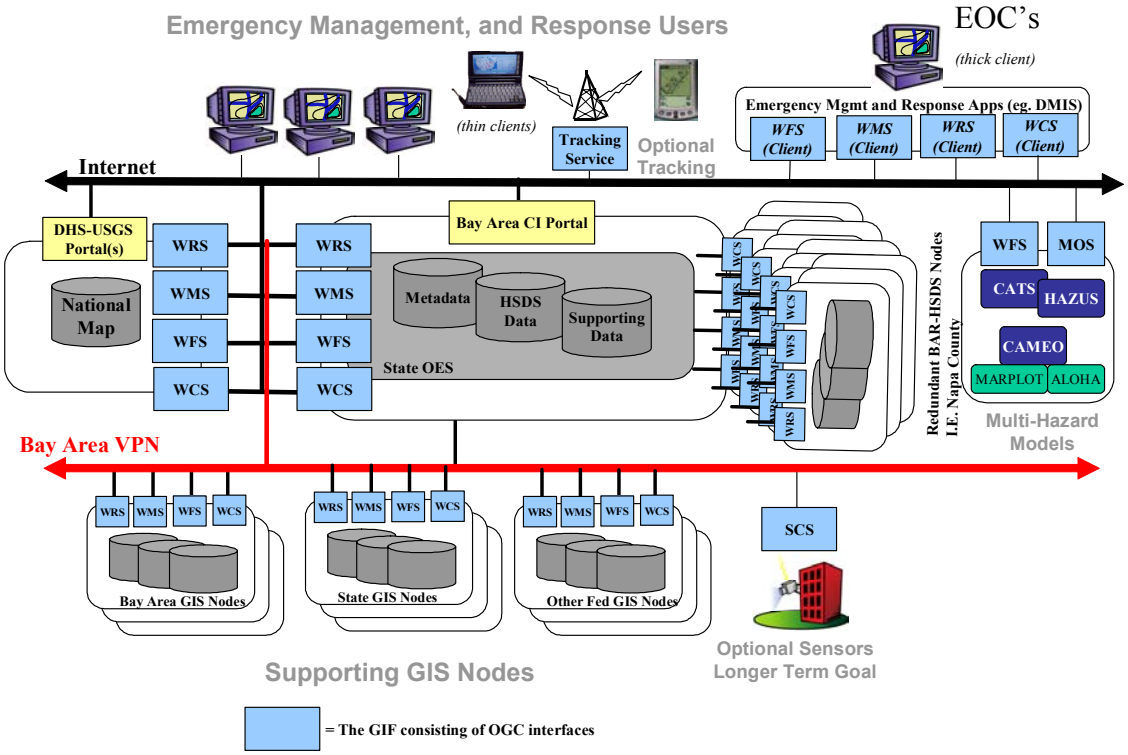
Information loaded to the system will also be sent to the USGS for incorporation into *The National Map* using the best method available. For this project, the USGS is not expected to perform any additional processing. The automation of the ETL (extract, transform, load) process will be explored to enable incorporation of local "quilt" data into national seamless "blanket" coverage within *The National Map*. The USGS will then forward the data to NGA and other HLS/HLD locations, as the data sharing agreements allow.

Components of the BAR-HSDS

A network of servers (2)	Build/Load as much data as possible
Data sharing agreements	Configure Palanterra to access
Secure login access for first responders	Model Results (CATS)
Demonstrate system interoperability	Region exercise
A data model – HLS Template	Documentation of lessons learned
Implementation plan	Long-term maintenance plan

Figure 3-4

BAR-HSDS Architecture Concept



The initial architecture depicted in Figure 3-4 conveys the distributed, interoperable nature of the BAR-HSDS, featuring GIF. Acronyms used in this diagram and key features of this architecture are described below.

GIF – The *Geospatial Interoperability Framework (GIF)* provides secure access to shared regional assets, based upon OGC interfaces.

Bay Area Regional Homeland Security Data Server Portal – This is a secure, critical node with multi-node backup and 24/7 capability that would contain all Critical Infrastructure & Key Asset metadata for the Bay Area. Plus, it would contain Emergency Planning and Operational Support Layers. Although illustrated here with a central BAR-HSDS Node for the data, this server will be a distributed, redundant capability, and will be further defined based on regional operational requirements.

DHS-USGS Portals – This is a national-level capability in support of national emergencies and includes the National Map.

Bay Area GIS Nodes – These are the various Bay Area GIS capabilities for collecting, storing, and managing detailed geospatial data across the region, using existing multi-vendor assets. These "Producer" nodes would always contain the latest, authoritative data for their applicable domains. Although some of this data will be replicated at the BAR-HSDS Node, it is suggested that remote nodes also provide 24/7 assured access to authorized users, with backup capabilities.

State GIS Nodes – These are the various State GIS capabilities for collecting, storing, and managing relevant geospatial data for the region, using existing multi-vendor assets.

Federal GIS Nodes – These are the various other National GIS capabilities for collecting, storing, and managing relevant geospatial data for the region, using existing multi-vendor assets.

Project EOC – An EOC capability that operates at one or more sites within the Bay Area, and that hosts one or more emergency response, recovery, reporting, assessment and incident management applications including the DMIS. Contra Costa County (CCC) EOC currently meets this criteria and is also a registered user of Cameo, Marplot, HAZUS, DMIS, and the Consequences Assessment Tool Suite (CATS).

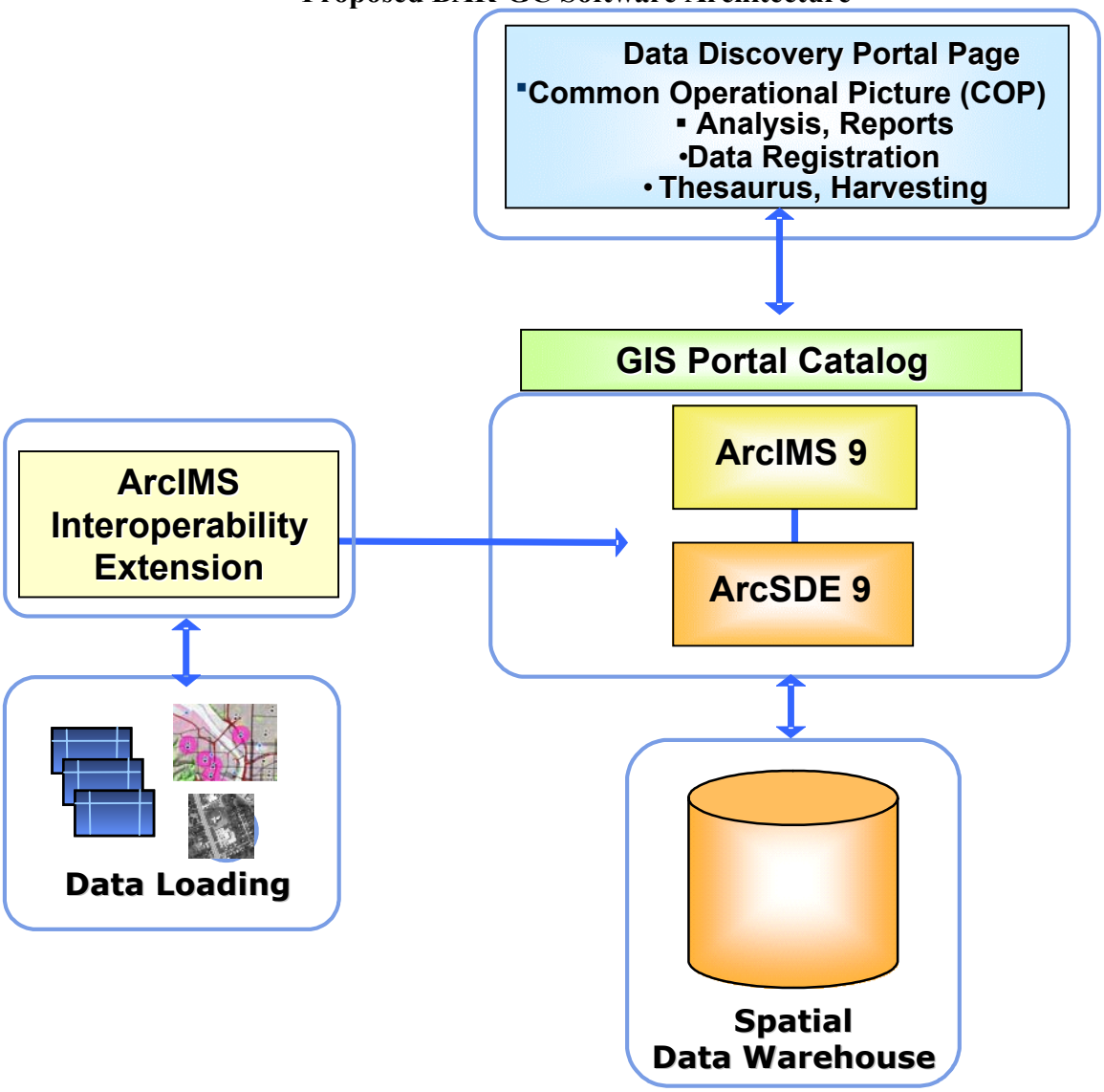
Mobile Terminals (optional) – These will host emergency response applications that run on mobile terminals (laptops, PDAs and cell phones) that provide Incident Commanders and their teams with timely, relevant data at the scene of an incident. Keeps field deployed units in the loop with latest common operating picture. This will be evaluated as a possible component of the Operational Exercise. It will be coordinated with Region 9 Federal Emergency Management Association (FEMA), California State Office of Emergency Services (OES), Metropolitan Transportation Commission (MTC), the Association of Bay Area Governments (ABAG), the Golden Gate Safety Network (GGSN) and the CCC Mobile Field Force and other federal, state, and local agencies.

OGC Web Services – Geospatial services that are accessible as Web resources. Consists of:

Web Map Service (WMS) – Produces stylized maps with desired layers
Web Feature Service (WFS) – Fetches select feature sets
Web Coverage Service (WCS) – Fetches select (raster) coverages
Web Registry Service (WRS) – Accesses OGC registries with metadata
Web Tracking Service (WTS) – Determines tracking information for mobile terminals
Sensor Collection Service (SCS) – Fetches observations from select sensors

The proposed software to support the BAR-GC implementation is described in Figure 3-5.

Figure 3-5
Proposed BAR-GC Software Architecture



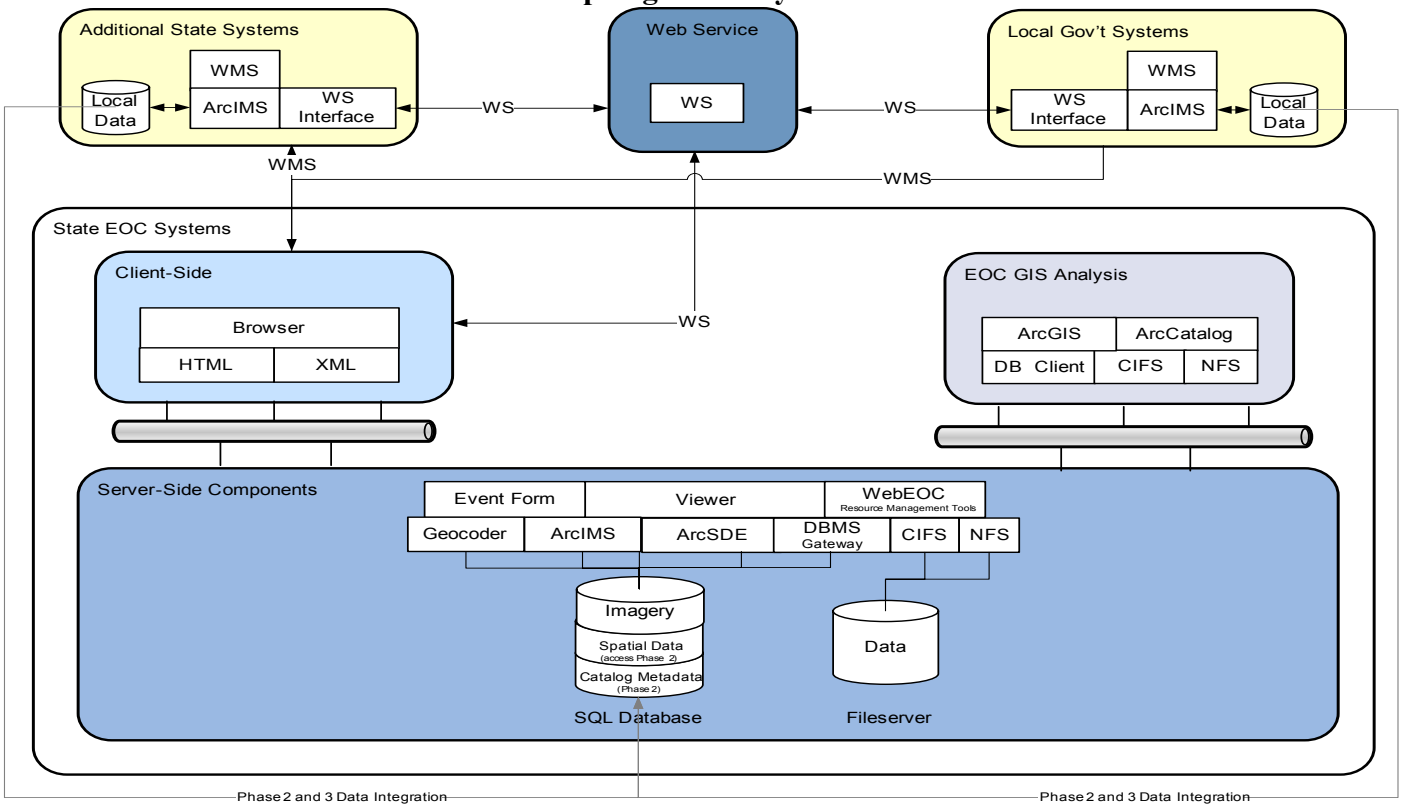
3.2.2.2 Colorado Springs/State

The Colorado HLSGIS is envisioned as an enterprise Web utility providing end-to-end, Web enabled, services-oriented geospatial intelligence integrated between state, local and federal jurisdiction, keeping abreast of current activities from its own Emergency Operations Centers (EOC) and offices. Key components of the Colorado Springs/State pilot include:

- Common Operational Capability for multi-levels of HLS/HLD
- Standards Based / Interoperable
- Integration with other HLS/HLD Systems
  - DMIS
  - National Map
  - Palanterra
  - COMFORCE
- Highlight the issues and costs of implementation at a national level
- Gather and distribute lessons learned by current players

Further details of the services and interface layers utilized are shown in Figure 3-6.

Figure 3-6
Colorado Springs/State Systems Overview



Data entering the USGS system from the Colorado HLSGIS will have two main branches, *The National Map* public domain section and a controlled access section. Data must be processed before addition to *The National Map.* The controlled access data will be forwarded to NGA and other HLS/HLD partners as designated in the MOU. Public domain data will be sent to NGA in both pre- and post-processed formats. The method of delivery will vary but may include data bricks, deposits on FTP sites, or processing the data through a secure location within the USGS and sending it over a secure network.

3.2.3 USGS Implementation

Data received from the local partners which is designated as publicly accessible will be loaded onto USGS data servers and registered with *The National Map* Catalog. Hardware and software to support this function will be drawn from existing infrastructure at USGS.

Data received from the local partners, which is designated as sensitive but unclassified (SBU), will be loaded onto controlled access data servers at USGS. Hardware to support this function will be drawn from existing infrastructure at USGS. The specifications for the environment will be described or referenced in a later draft of this document. Software prototypes to support the controlled (secure) access requirements will be implemented as part of the spiral activities.

3.2.4 NGA HLS/HLD Implementations

NGA data holdings will consist of the public and controlled access data from *The National Map,* NGA's data (e.g., 133 cities MEDS), and data from other government agencies (e.g., the DoD base facilities data). This data may be held at multiple security levels to support clients on the NIPRNET, SIPRNET, and JWICS networks.

NGA currently hosts a copy of *The National Map* data from USGS, which augments NGA's own data holdings (see GIDI Site Document, Bethesda CDRL #013, April 2004 for more information). This data has been transmitted from USGS to NGA using media such as "data bricks." During The National Map/Palanterra™ Synchronization activities, ESRI will work with USGS and NGA to streamline the process for updating the NGA data holdings provided by USGS.

3.3 Data

One of the key components of the program is the flow of data and metadata between stakeholders. This presents a number of technical and organizational challenges, and the project approach is to exercise these parts of the system through a series of pilot projects. Some of the key technical challenges are:

- Aggregation of data from local providers into regional/state systems has implications for data sharing, security, and database design best practices.
- Groups at each level need flexibility to self-define the database design that works for them, but at another level, organizations have data requirements that may be different from what local providers currently make available.
- The needs of HLS/HLD users are not well defined for the local data producers. There is a need for better guidance on the information needs of the HLS/HLD community.

There are design, organizational, and funding aspects to these technical considerations. The goal of this project is to make progress on the technical aspects and document implications in the other areas.

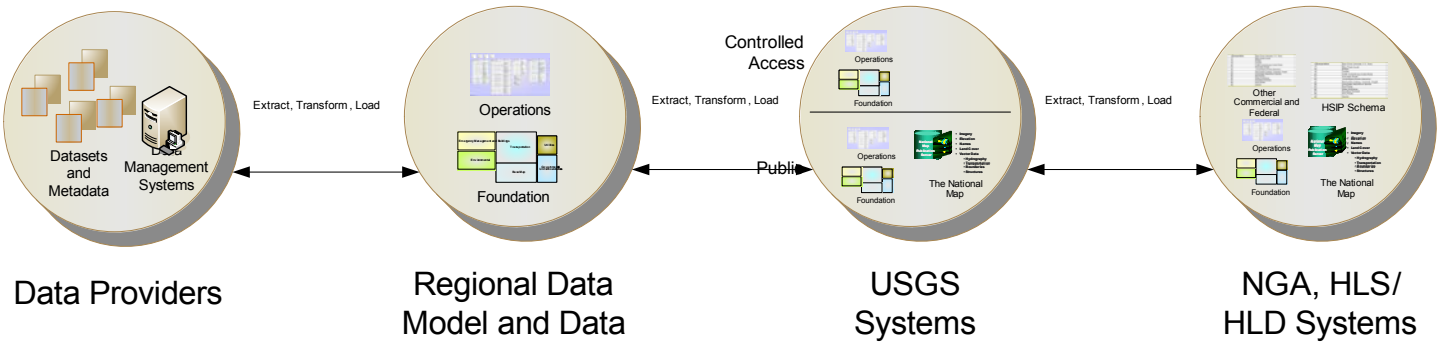
3.3.1 Data Types

In terms of the high level content of the data model, there are two important categories of information.

1. **Foundation data:** This is the basemap, transportation, administrative areas, environmental, utilities, and emergency management/buildings content. This model follows local best practices, and the starting point for each regional design is a common database template.
2. **Operations data:** This content includes incident, operations, infrastructure, and natural event information. The structure is derived from the FGDC Homeland Security Working Group (HSWG) standard.

The location of these data types is illustrated in Figure 3-7.

Figure 3-7
Data and Metadata Paths



ESRI is currently working with the stakeholders to identify the types of data that will be contained at each location and which security classification system will hold that data. An example of this "mapping" is shown in Figure 3-8 below. Additional data will be added to this "mapping" as it is collected from the stakeholders.

Figure 3-8
Data and Web Services by Network

	BARGC	CO Springs	CO DISDI	CO State	USGS RMMC	USGS EROS	NGA	DHS-ABCI
JWICS							Palanterra™, Other Commercial and Federal, HSIP Schema, Operations, Foundation, The National Map	
SIPRNET			DISDI		The National Map		Palanterra™, Other Commercial and Federal, HSIP Schema, Operations, Foundation	
NIPRNET							Palanterra™, Other Commercial and Federal, HSIP Schema, Operations, Foundation	
Public Internet			DISDI (Subset)		The National Map			
Private Network	BAR-HSDS, Operations, Foundation	Operations, Foundation		TSAP, Operations, Foundation	The National Map, Operations, Foundation			COMFORCE

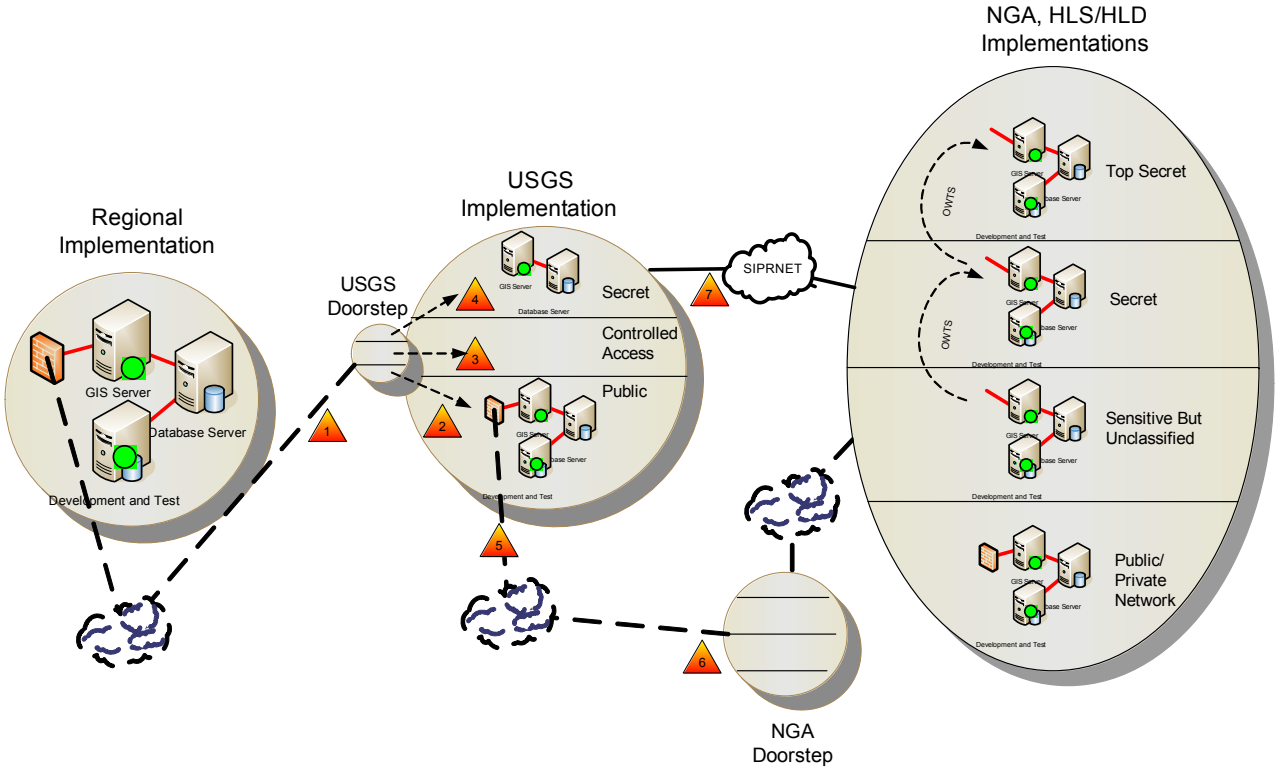
3.3.2 Data Flow

Data will be an essential component of the NGA Homeland Security Enterprise GIS Support Project system architecture. Modeling of data flows is one of the major challenges of the project. Data flows will involve moving (and sometimes transforming) data among multiple entities including NGA, the USGS, other NGA partners, and state and local organizations. It will also involve managing data with varying levels of security classifications.

3.3.2.1 Current Data Flow

A high level overview of the general data flow within the system is presented in Figure 3-9.

Figure 3-9
Current Data Flow Through the USGS



Facts about each of the data flows shown in the diagram (keyed to Figure 3-9 by number) are summarized below:

1. *Data into USGS*
   a. Transfer methods include CD, DVD, external drive (FireWire or USB2). In general, transfers are not online with ftp or transaction type database to database operations.

b. Elevation, orthoimagery, BTS roads come from EDC. Other layers come from other places. See http://seamless.usgs.gov for details on formats and user options.
c. Once data arrives at the USGS it is typically loaded into SDE, and the original media is deep archived. For orthoimagery, it is loaded into a near-line (STK silo) prior to SDE ingest.
d. Data is typically re-projected to geographics and NAD83, NAVD. An exception is high-resolution orthoimagery which is generally .3m cellsize in UTM, with USNG grid used for origin. In some cases, multiple instances of images are created to allow for a version in native form, plus a version in geographics that provides for re-projection on the fly (through ArcIMS).
e. Controlled access data is protected (per partner agreement) with a password web application and secure map services.
f. Generally data are downloaded for an application and then operated on locally for USGS Science.

2. USGS Maintenance Operations for Public data

a. Rights management and inventory are currently maintained through database or spreadsheets with partner agreements for urban areas, EDC registry or catalog that controls data layers served through ArcIMS. This is only for layers served by EDC, not *The National Map* catalog that is hosted and operated at MCMC.

3. USGS Maintenance Operations for Controlled data

a. Further details to be provided in next revision.

4. USGS Maintenance Operations for Secret data

a. Only data from The National Map will be migrated to the Secret classification at USGS.

5. Data out of USGS

a. A push method is currently utilized for transferring data out of the USGS.
b. Users are able to direct the automated transfer of data. Bulk transfers are ad hoc, and are usually externally driven with ortho “chips” of 1500x1500 meters, or (preferably) Oracle tablespace “drawoffs” that save loading time for the user end.
c. There are currently no automated means of determining that the NGA has the latest data on their servers. Regular teleconferences to discuss schedules on orthoimagery are the primary means of determining data status.
d. Currently the transfer of data is not secure

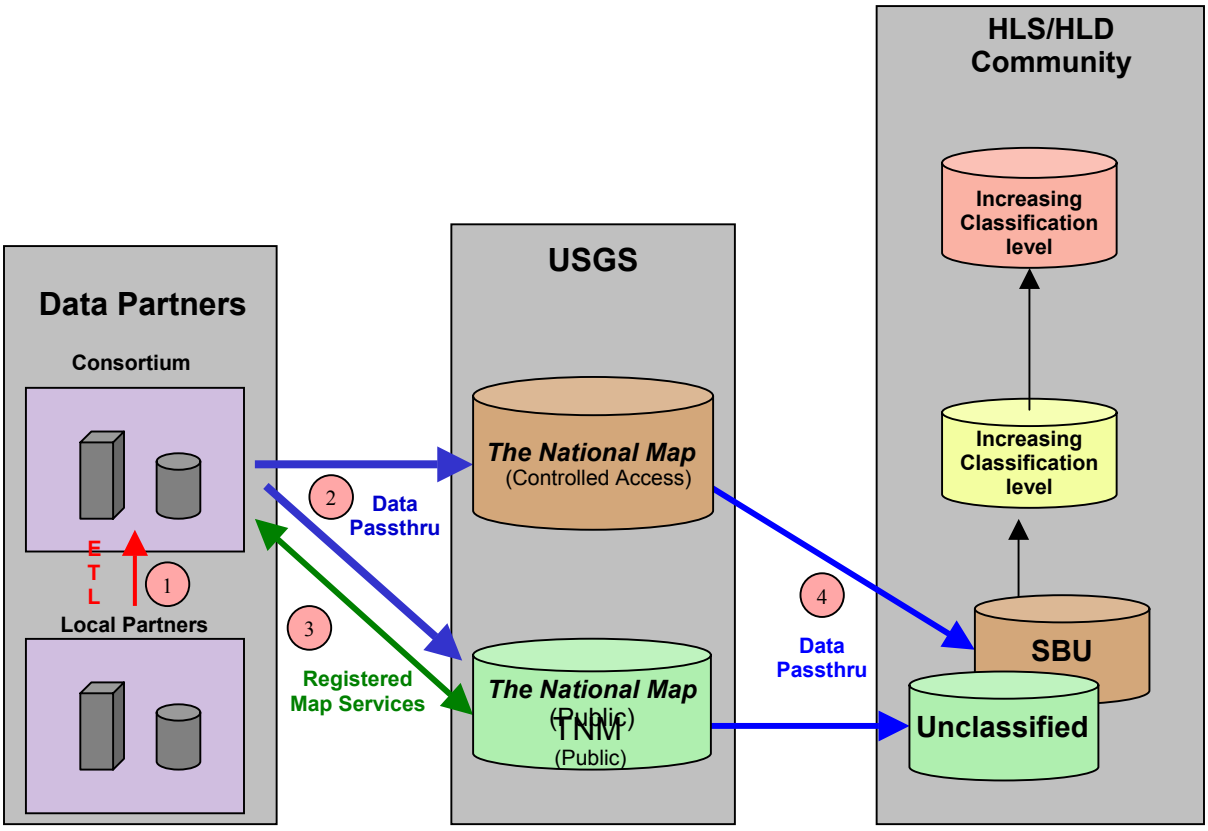
6. *Data into NGA*
   a. Further details to be provided in next revision.

7. *Data into SIPRNET from USGS*
   a. RMMC has JWICS capability, and possibly SIPRnet. EDC will be setting up SIPRnet in FY05 (for another project at EDC).

3.3.2.2 Intermediate Synthesis and Collaboration Data Flows: BAR-GC

Figure 3-10 depicts the data flows that are being investigated during the BAR-GC spiral activities. This is envisioned as an intermediate step towards the long term goals for a more automated data flow from the data partners to USGS and the HLS/HLD community.

Figure 3-10
Intermediate Synthesis and Collaboration Data Flows: BAR-GC



ETL = Extract, Transform, Load

1. The local partners synthesize their data into integrated regionally hosted data sets using an extract, transform, and load (ETL) process.

2. The consortium passes data to the USGS without making changes to the data schema. Data that is identified by the local partners as "sensitive but unclassified" will be loaded into the controlled access version of *The National Map* data stores. Data that is identified by the local partners as publicly accessible are loaded into the data stores for the public version of *The National Map*.

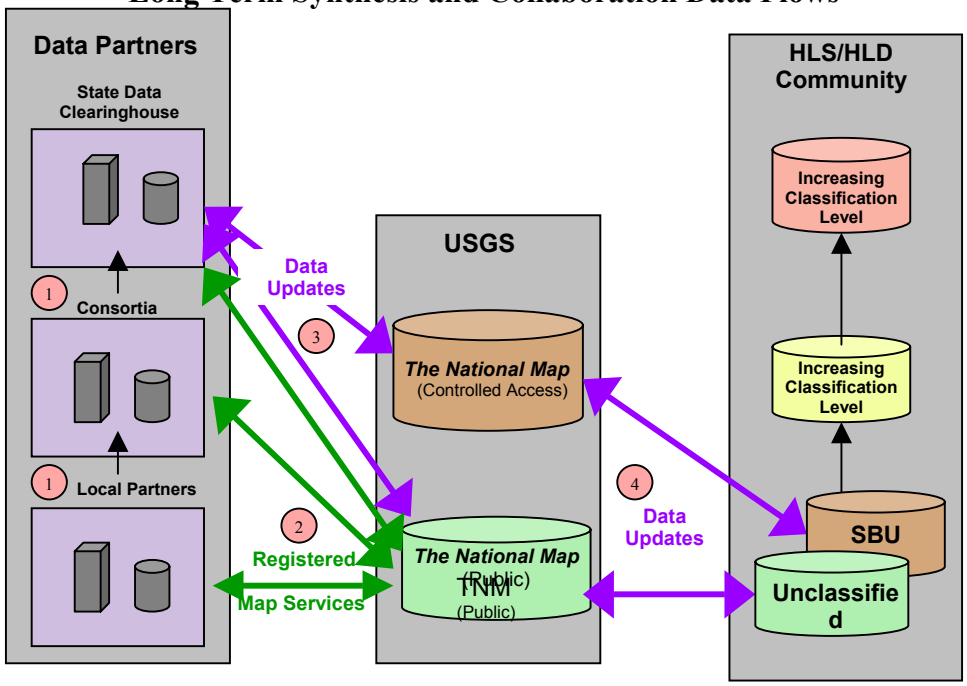
3. The consortium will create Web mapping services and register them with *The National Map* catalog.

4. The USGS passes data from the public and controlled access data stores to the Homeland Security (HLS) and Homeland Defense (HLD) community without making changes to the data schema.

3.3.2.2 Long-Term Synthesis and Collaboration Data Flows

Figure 3-11 depicts a proposed long term solution for automating the synthesis and collaboration activities.

Figure 3-11
Long Term Synthesis and Collaboration Data Flows



1. Spatial data flows from local partners to regional GIS consortia to state-level partners.

2. The local partners, consortia, and state agencies will create Web mapping services and register them with *The National Map* catalog.

3. The partner data will be loaded onto the controlled access and publicly accessible data stores, based upon the partner-identified access privileges. After the initial data load, updates made by the USGS or the data partners will be passed to each other on an ongoing basis.

4. The data from USGS will be loaded onto the Homeland Security (HLS) and Homeland Defense (HLD) data stores, based upon the partner-identified access privileges. After the initial data load, updates made by the USGS or the HLS/HLD community will be passed to each other on an ongoing basis.

3.4 Security

Different security controls should work in a layered approach to protect assets.

- Administrative Controls

  Policies, standards, procedures, guidelines, screening personnel, security awareness training

- Technical Controls

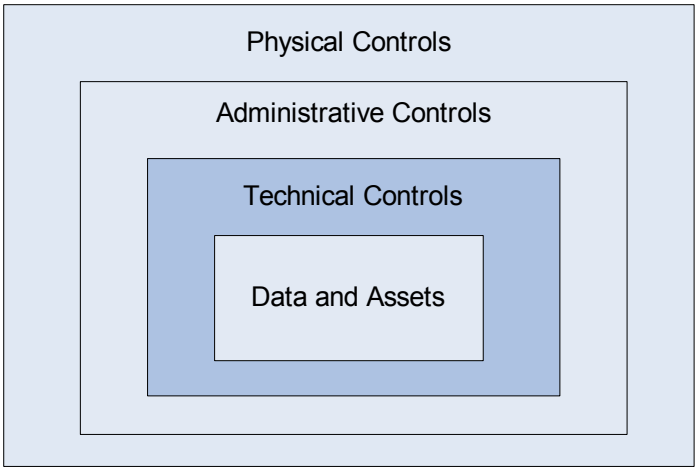
  Logical access controls, encryption, security devices, identification and authentication

- Physical Controls

  Facility protection, security guards, locks, monitoring, environmental controls, intrusion detection

This project assumes that physical and administrative controls are fully managed by the customer; therefore, security issues discussed in this document focus on the technical controls as depicted in Figure 3-12. An overview of these Technical Controls may be found in the *Project Homeland System Architecture Prefinal* document.

Figure 3-12
Layered Security Approach



3.4.1 Data Security Classifications

The Project Homeland stakeholders currently do not have a standard convention for classifying the security levels of the data they share as illustrated in Figure 3-13. Every effort will need to be made to identify or characterize access groups in a way that matches up with other levels of governments so downstream administrators can effectively manage access.

Data Providers and Regional Hubs typically classify their data as:

- Public domain data
- Sensitive
- Restrictive

Each data provider will need to determine the level of access for their data and may utilize the Rand Report and FGDC decision tree to make this determination.

The USGS classifies their data as:

- Public
- Controlled
- Top Secret (JWICKS)
- Secret (currently under review)

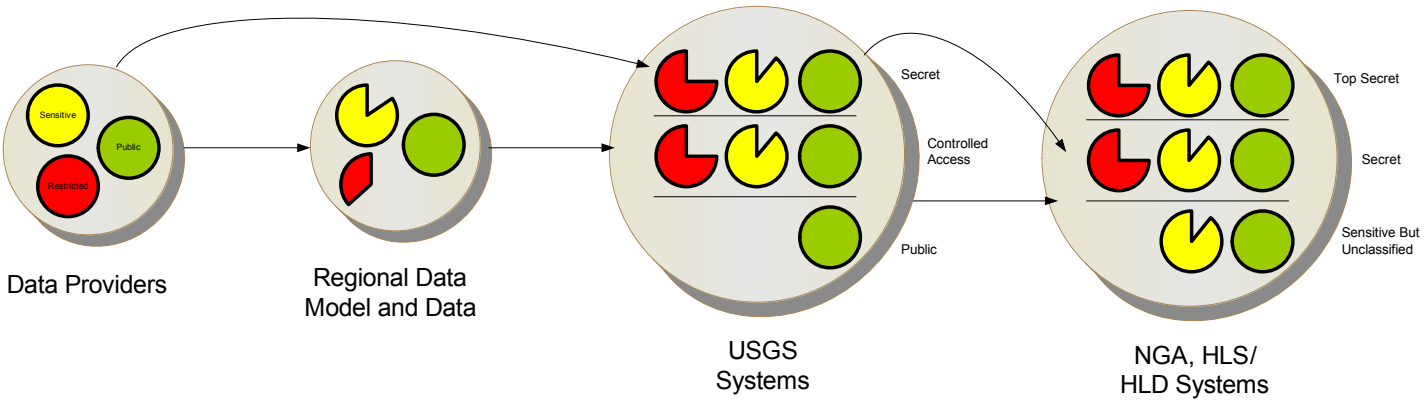
Some local data providers will pass restrictive or sensitive data directly to the USGS, bypassing the Regional Hub.

The NGA, HLS/HLD systems classify their data as:

- Sensitive But Unclassified (NIPRnet)
- Secret (SIPRnet)
- Top Secret (JWICKS)

The feasibility of transferring data at a classified level on the SIPRnet between the USGS and NGA is currently being investigated.

Figure 3-13
Security Classifications Between Stakeholders



3.4.2 Data Labeling

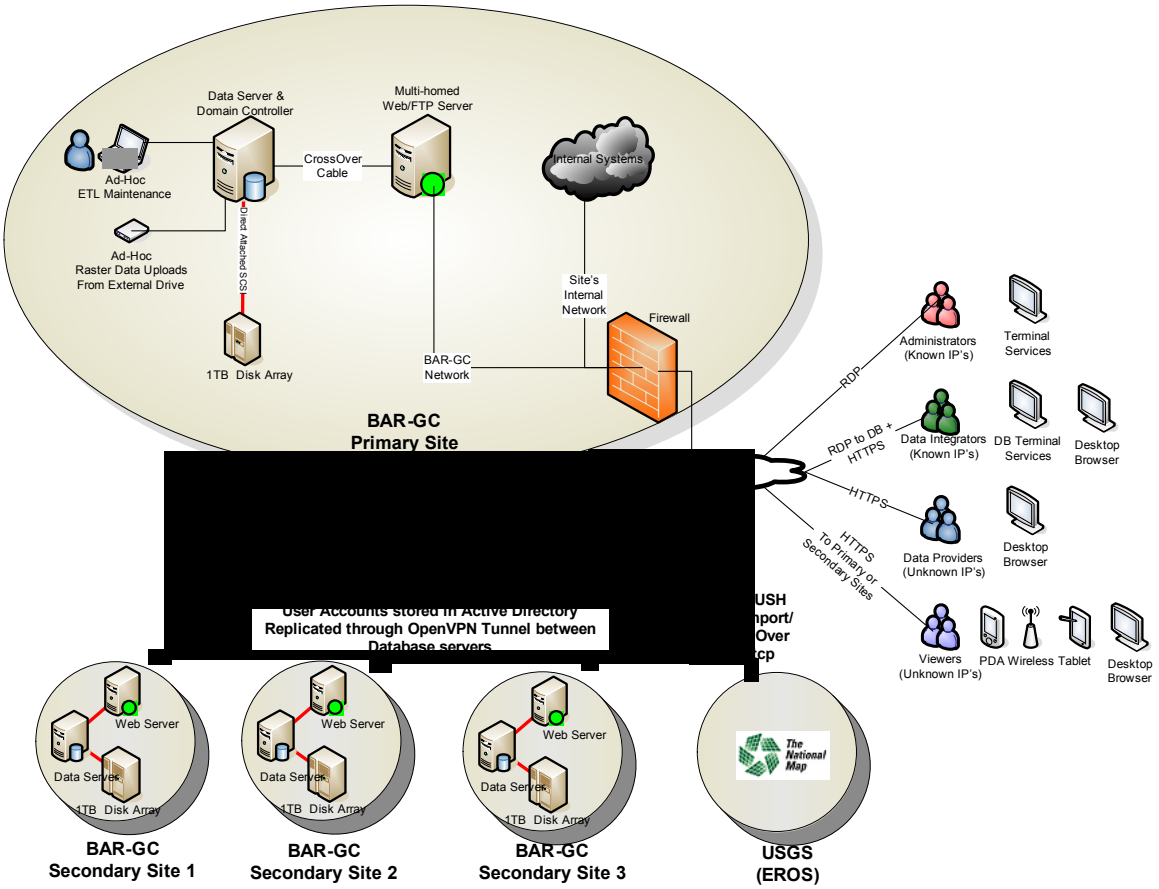
ESRI is current researching techniques to provide data security labeling at the Feature Level. Some of this research focuses on:

- Tracking through a new database table that includes
  - Security Classification
    - Unclassified, SBU, Confidential, Secret, Top Secret
  - Security Categories
    - Groups that a subject must have a need-to-know of
- Extract, Transform, Load applications distribute data based on the above security information

4.0 Reference Architecture for Regional and State Hubs

The BAR-GC pilot is the primary source for the current Regional and State hub architecture described in this section. This reference architecture will evolve during the project to reflect lessons learned during the Colorado State pilot. An overview of the "AS IS" BAR-GC architecture is shown in Figure 4-1.

Figure 4-1
BAR-GC HSDS Architecture Overview



The BAR-HSDS is a group of 4 redundant sets of servers strategically located in Berkeley, San Francisco and 2 sites yet to be named. Each server set (or host site) is capable of providing critical geospatial information region-wide across 9 counties in the Bay Area for the Homeland Security community at all levels of government. Information loaded to the system will be "pushed" to the USGS via FTP or HTTPS. This allows the USGS to act as a broker and

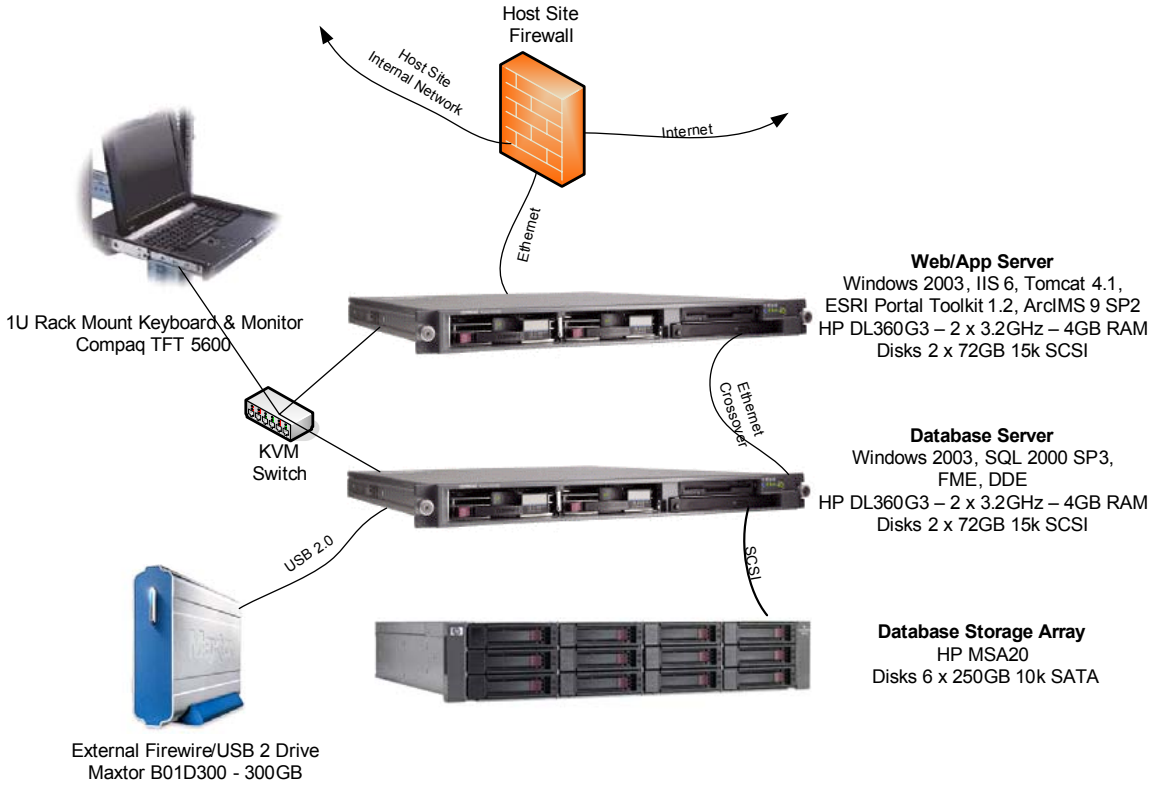
forward the data to the NGA and other federal HLS/HLD agencies within the constraints of the data sharing agreements.

Built upon commercial-off-the-shelf (COTS) framework, the BAR-HSDS provides the Bay Area with the ability to quickly mobilize for emergency response, provide for redundant data repositories, capabilities to track events, and assess response activities. The primary HSDS system components include:

- Heterogeneous Data Loading and Downloading Tools (DDE & DIE)
- Spatial Data Repository (SDE & SQL 2000 database)
- File Transfer Services (FTP, WebDAV)
- Web Map Services (ArcIMS, IIS, Tomcat)
- Common Operating Picture (COP) Application (BAR-GC Viewer based on JITK)

Figure 4-2 provides an overview of the equipment for an individual host site. All equipment together takes up less then 7u of space in the 36u rack which was provided. This allows for host sites to add their power supplies in the bottom of the rack and future system expansion.

Figure 4-2
Equipment for a Single Host Site



Note that since some host sites are strictly regulated data centers, system power requirements are listed in Figure 4-3. This information may be utilized to help determine if additional power may be needed for potential host sites or if additional air-conditioning is necessary.

Figure 4-3
Equipment Power and Heat Requirements

Equipment	Current (Amps)	Power (Watts)	Heat (BTU)	# Plugs & Type
HP DL360G3 - Web	5	450	1535	2 - NEMA5-15P
HP DL360G3 - DB	5	450	1535	2 - NEMA5-15P
MSA20 - Disk Array	6	345	1175	1 - NEMA5-15P
Monitor	2	150	550	1 - NEMA5-15P
KVM	1	9	30	1 - NEMA5-15P
External FW Drive	2	150	550	1 - NEMA5-15P
Totals	21	1554	5375	-

There are five defined solution architectures in the BAR-GC Reference Architecture. These architectures are:

- Network
- Storage
- Application Infrastructure
- Management
- Security

4.1 Network

It is important to make a distinction between the logical "role," which is a requirement of the network, and the physical "device" that can fulfill the requirement. In many cases, the enterprise-level device used for a role may have the capability to perform other roles as well. Implementation and recommended configuration for the following network devices is provided below:

- Router
- Switch
- Load Balancer
- SSL Accelerator
- Firewall
- Intrusion Detection Systems (IDS)
- Virtual Private Networks (VPN)
- IP Address Scheme

- Router devices

  Locking down incoming external communication ports is recommended, but it is recognized that many installations do not have the authority to adjust open ports as necessary on the router and that many hubs prefer to consolidate their port management to their firewall and switch systems. No router port configuration changes were made for the BAR-GC implementation.

- Switch Devices

  Virtual LANs are useful for limiting communication between networks. For the BAR-GC systems, different VLAN switch configurations for the web and database server are eliminated by multi-homing the web server – One crossover connection for the database server, and one connection to the Internet (public facing network). Some hosting sites will still want to create a separate VLAN for the servers, which is not a problem as long as the VLAN allows the same types of communication with the web server as the firewall.

- Load balancing devices

  The individual BAR-GC implementations are currently not large enough for load balancing. If desired Microsoft Network load balancing may be utilized to increase availability and throughput of individual sites. Hardware load balancing devices may be utilized, but should be avoided unless the host site administrators have extensive experience with the devices.

- SSL Accelerators

  SSL accelerator hardware is not utilized within the BAR-GC pilot. This hardware should only be used for implementations expecting very large numbers of concurrent users. Applications that have not been explicitly tested for usage with SSL accelerators have been known to fail.

- Firewall Devices

  The BAR-GC systems are connected directly off a firewall port as a separate network segment. The following ports have been opened at the firewall for communication with the BAR-GC network segment:

  ```
  ftp-data 20/tcp
  ftp      21/tcp
  SMTP     25/tcp (Outbound Only)
  domain (DNS)53/tcp/udp
  ntp      123/udp
  https    443/tcp
  OpenVPN  1194/udp
  RDP      3389/tcp/udp
  ```

  As the usage of web services increases the importance of XML/SOAP Firewalls increases. Host site firewalls should be reviewed for these relatively new

capabilities. Note that it is recommend that sites use Network Address Translation (NAT) in order to keep systems configured similarly, but is not a requirement.

- Intrusion Detection Systems (IDS)

  An IDS gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and misuse (attacks from within the organization). IDS uses vulnerability assessment (sometimes referred to as scanning), which is a technology developed to assess the security of a computer system or network. Due to the ongoing, significant administration requirements of IDS and relative high costs, the BAR-GC solution does not include IDS hardware.

- Virtual Private Network (VPN) devices

  Since these systems utilize a common user rights management system (Microsoft's Active Directory), a hardware VPN site to site connection would provide the most robust, secure, live data capability. Recently, SSL VPN software has started to provide an alternative to expensive hardware solutions. OpenVPN is one such product that is being utilized to transfer user and system management data securely between servers. In addition to encapsulating Active Directory traffic within VPN, host sites may further improve their security by forcing Remote Administrators (using the RDP protocol) to use VPN.

- IP Addressing Scheme

  Both the web and database servers are to utilize private IP Addresses. This is supported by using a public IP address for the web server on the Firewall and NATing to the private IP address of the physical web server. IPv4 addressing is to be utilized due to familiarity and wide-spread device support. Class C assignment for both network segments is adequate.

4.2 Storage

Data storage is an important consideration in supporting the Project Homeland environment, due to the amounts of data involved. Storage requirements for individual cities alone have increased from several 10's of gigabytes to over a terabyte of data within the last couple of years. It is not uncommon to find the cost of the storage solution to exceed the cost of the enterprise server, both in terms of initial hardware and overall administrative cost. Data storage and management tasks reviewed in this section include:

- Storage Equipment
- RAID and Location
- Replication
- Backups

4.2.1 Storage Equipment

The BAR-GC database servers utilize the HP StorageWorks Modular Smart Array 20 Enclosure (MSA20), see Figure 4-4 below. It is a SATA 1.5 Gb/s disk drive storage enclosure with Ultra320 SCSI host connectivity. Direct Attached Storage (DAS) was selected for the BAR-GC pilot database servers for the following reasons:

- Minimal administrative overhead, low entry cost and simplicity
- Scalability
  Additional disk arrays may be daisy chained together to create a larger repository. SATA drives were chosen to reduce costs, but SCSI drives can be added at a later date if performance requirements drive it.
- Availability
  Individual systems are not currently designed for high availability. The combination of the 4 BAR-GC sites provides a high-availability solution.

Figure 4-4
HP MSA20 SATA Direct Attached Storage

Larger data repositories at the state level will potentially require Storage Area Network (SAN) solutions. Current direct attached SATA storage solutions easily scale to 3 TB when utilizing 250GB hard drives. If an HP MSA1500 controller is added, 7 additional MSA20 systems may be connected together for a total storage capacity of 24TB.

4.2.2 RAID and Location

There are several basic considerations that must be addressed in selecting the appropriate storage solution. Figure 4-5 provides a comparison of price and performance for various Microsoft SQL Server disk configurations.

Figure 4-5
Characteristics of SQL Server Disk Configurations

# Physical Drives	Data Location	Log Location	Relative read performance	Relative write performance	Eases I/O hot spots and does not require handcrafting physical layout	Relative Cost	Up-to-the-minute data protection	Any single drive failure does not cause immediate outage	Can sustain multiple drive failures without an outage
1	Disk 1	Disk 1	Very Low	Very Low	No	Very Low	No	No	No
2	Disk 1	Disk 2	Low	Low	No	Low	No	No	No
3+	RAID 0	Disk 3	High	High	Yes	Low-Mod	No	No	No
4+	RAID 0	RAID 1	High	High	Yes	Moderate	Yes	No	No
4+	RAID 1	RAID 1	Varies	Varies	No	High	Yes	Yes	Partial
5+	RAID 5	RAID 1	High	Low-Mod	Yes	Moderate	Yes	Yes	No
6+	RAID 10	RAID 1	Very High	High	Yes	High	Yes	Yes	Partial

For SQL Server, Microsoft recommends locating the SQL log files on a RAID 1 mirrored pair of disks with index and data files on RAID 5 (striping with parity disk) volumes. To further improve performance, the index and data files may be placed on a RAID 10 (mirror and striping) disk set. A separate RAID 10 array is usually not an appropriate choice for the transaction log. Because the write activity tends to be sequential and is synchronous, the benefits of multiple disks are not realized.

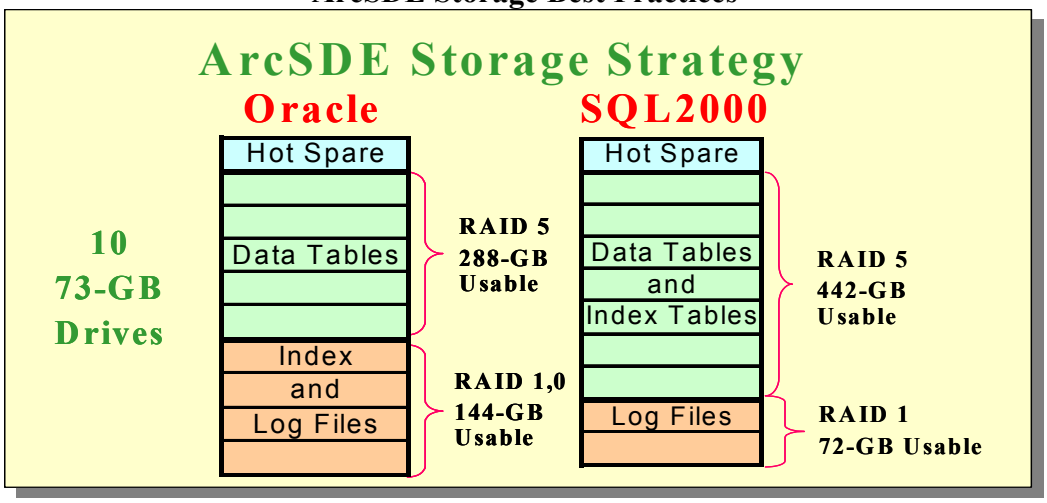
For the BAR-GC implementation RAID 1 was used for local server drives and RAID 5 utilized for the Direct Attached Storage (DAS) array as seen in Figure 4-6 below.

Figure 4-6
BAR-GC Disk System Configuration

Server	Logical Drive	Data Stored	Logical Size	Location	Array	# Disks	Disk Size (GB)	RAID	Controller
Web	C	OS	20GB	Local	A	2	72	1	Built-in
Web	D	Uploads	50GB	Local	A	2	72	1	Built-in
Database	C	OS	20GB	Local	A	2	72	1	Built-in
Database	D	DB Log	50GB	Local	A	2	72	1	Built-in
Database	E	DB Data	1000GB	DAS	B	6	250	5	SA6402

Note that Oracle recommends that high-usage index and log tables be stored on RAID 10 and data tables on RAID 5. For best practice, the RAID 5 volumes should be supported on 5 (4+1) or 9 (8+1) disk volumes (five-disk RAID 5 supports highest peak capacity). Storage recommendations are summarized in Figure 4-7.

Figure 4-7
ArcSDE Storage Best Practices



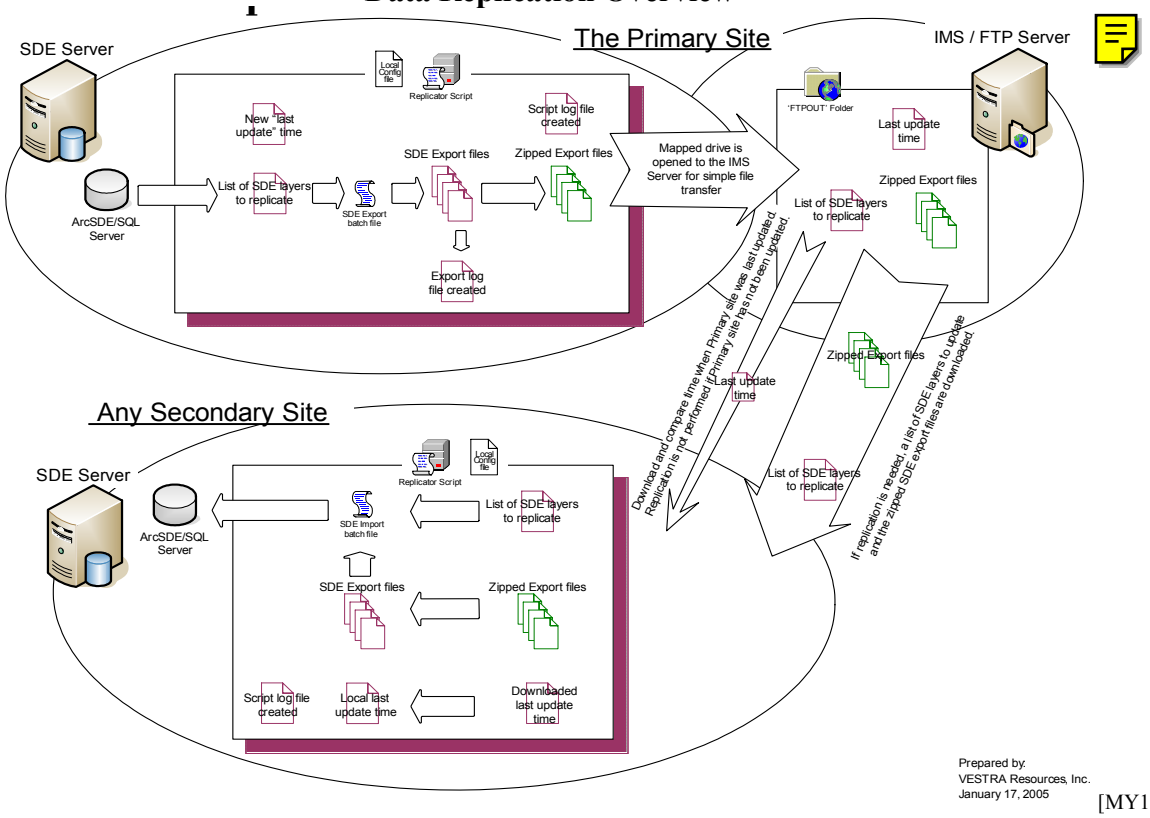
Several factors contribute to disk access performance. Data striping along with the array data cache improves access performance. Too few disks in the array (with lots of concurrent client queries) can result in disk contention (I/O performance bottlenecks). Small database environments with a large number of peak user loads present the highest probability of disk contention. As a general guideline, a minimum of five-disk RAID 5 array should support the ArcSDE data files.

4.2.3 Replication

The BAR-HSDS hosts sites have redundant data incase an emergency situation disrupts any of the hubs. To keep all hubs synchronized with current data, a data replication scheme has been developed. The scheme allows all data in a 'primary' site to be replicated to any and all of the secondary sites. Any site can take on the role of primary or secondary site, but there can only be one 'primary' site at a time.

The replication process is fundamentally divided into two independent but complimentary processes, an export process and an import process. The export process runs on the primary server. The process ultimately creates a compressed and encrypted file of all the vector data in the local ArcSDE geodatabase and two ancillary files the support data import at the secondary sites. The import process runs on a secondary site and retrieves, from the primary site, the latest replication data set. The dataset is then used to update all of the layers on the secondary sites ArcSDE geodatabase. Control files on the primary and secondary servers, generated by the import and export processes are used to control if replication occurs and which feature classes are updated. An overview of the replication process is shown in Figure 4-8.

Figure 4-8
Data Replication Overview



Technology used for the replication of ArcSDE feature classes includes SDE Command line tools, Python programming language, WinZip file compression and encryption and FTP. ArcSDE command line tools were chosen because they are database independent and robust. The Python programming language was chosen because of its simplicity, broad functionality, geoprocessing ability (through in ArcGIS scripts) and platform independence. WinZip was chosen for compression, AES-128 bit encryption, and because of its broad support and command line encryption tools. FTP was chosen as a transfer protocol because of its speed and programmability within Python.

Note that a replication process is currently not in place for "raw" datasets that have been uploaded by Data Providers. One possible service solution is to utilize the Windows File Replication Service (FRS).

4.2.4 Backups

Standard data center backup procedures should be followed at each site. No separate backup solution is provided as part of the BAR-GC system. For disaster recovery of the Viewer application, the Web server will be "Ghosted" to DVD/CD once the site is approved, along with any custom SDE/Python scripts for the database server.

4.3 Application Infrastructure

There are many possible options for choosing which applications to utilize to support a web site such as the BAR-HSDS site. The primary considerations for the BAR-GC pilot were COTS, functionality, cost, and ease of management/maintenance. This section provides an overview of the applications and services utilized in the BAR-GC environment, see Figure 4-9.

Figure 4-9
System Application and Service Summary

Web Server		Database Server	
Applications	**Win Services**	**Applications**	**Win Services**
ESRI ArcIMS 9.0 SP2 ESRI JITK Tomcat 4.1	IIS 6 FTP RDP 5.2 WebDAV NTP	MS SQL 2000 SP3 OpenVPN Data Delivery Extension (DDE) Data Interoperability Extension (DIE) ArcSDE 9 SP2	Domain Controller DNS RDP 5.2

4.3.1 Service Details

Details for the following services are provided in this section:

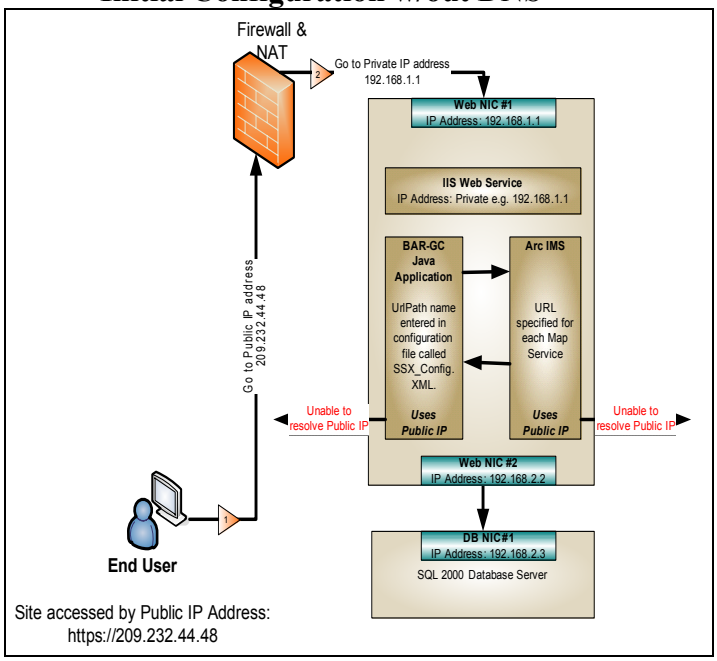
- DNS
- FTP
- WebDAV
- EFS
- CA
- RDP
- SMTP
- NTP

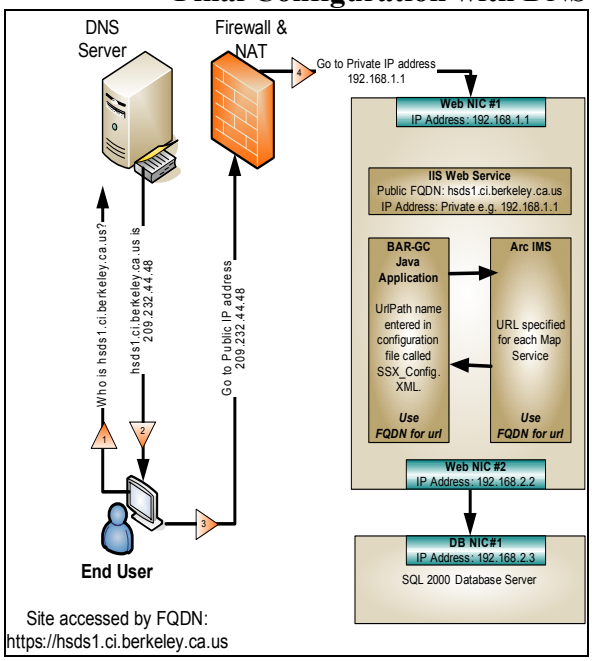
DNS (Port 53/tcp/udp)

The preliminary configuration for the BAR-GC project attempted to avoid usage of DNS and in the process avoid the potential outage risks associated with DNS. However, the BAR-GC Viewer is currently not designed to support using Network Address Translation (NAT) on the firewall while using IP addresses for the URL. The initial and final DNS configurations are show in Figure 4-10.

Figure 4-10

Initial Configuration w/out DNS **Final Configuration with DNS**





Each HSDS site utilizes a different Fully Qualified Domain Name (FQDN), and uses the format HSDS#.x.x. DNS is not used to round-robin load balance access to the sites utilizing a single FQDN, since the BAR-GC Viewer requires sticky sessions.

FTP (Port 21/tcp):
Active-mode FTP is the current protocol utilized for transferring data replication information from the primary site to secondary sites. Recognizing that this protocol is inherently insecure the following mechanisms have been implemented to ensure safe data transfer:

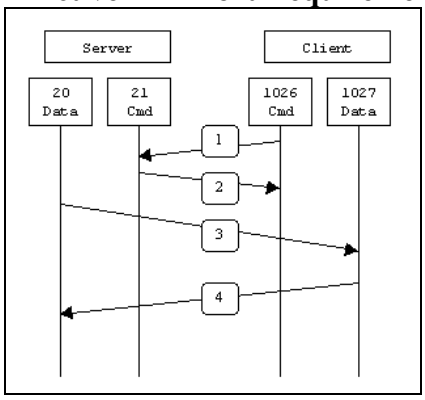
- Access to the FTP site is limited by IP address of the other BAR-GC host sites.
- The user account used for FTP site access has been denied log on local rights and is limited to specifically reading data from the Primary site's FTP folder.
- Files are automatically encrypted with WinZip 9.0's AES-128 bit encryption before being sent to secondary sites.
- Active-mode limits FTP port requirements to two ports.

Long-term, this solution should be supplemented or replaced with a more secure means of transferring data. Utilizing the WebDAV protocol with EFS is believed to be one of the primary future data transfer candidates or utilizing a 3rd party SFTP product. This is discussed more extensively in the WebDAV service section below.

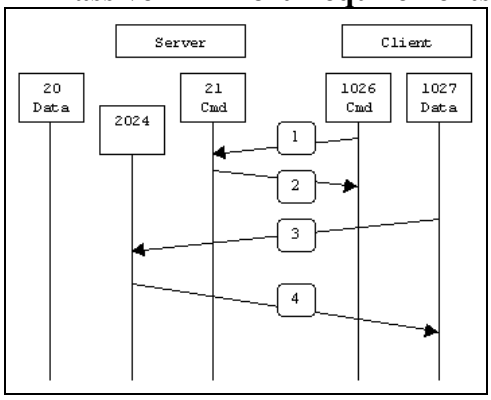
The BAR-GC configuration currently utilizes active FTP to minimize the number of open ports required. Windows 2003 servers support either active or passive FTP, so the service type available is driven by the ports open on the firewall and the method specified by the client. Figure 4-11 compares the port requirements of active vs. passive FTP.

Figure 4-11

Active FTP Port Requirements



Passive FTP Port Requirements



Active-mode FTP connections are sometimes referred to as "client-managed" because the client sends a port command to the server, over the control connection. The command requests the server to establish a data connection from TCP port 20 on the server to the client, using the TCP port that is specified by the port command.

Passive-mode FTP connections are sometimes referred to as "server-managed", because after the client issues a pasv command, the server responds with one of its transient ports used as the server-side port of the data connection. After a data connection command is issued by the client, the server connects to the client using the port immediately above the client-side port of the control connection.

The most common problem encountered with FTP over the Internet involves data transfers through a proxy server, a firewall, or a Network Address Translation (NAT) device. In most cases these network security devices allow the control connection to be established over TCP port 21 (that is, the user can successfully log on to the FTP server), but when the user attempts a data transfer such as DIR, LS, GET, or PUT, the FTP client appears to stop responding because the network security device is blocking the data connection port that is specified by the client. If the network security device supports logging, you can verify port blocking by reviewing the deny/reject logs on the network security device.

If passive FTP becomes a requirement for future implementations, the following guidelines should be followed. By default, the FTP server allocates ports for passive-mode connections from the WinSock dynamic range, 1024 to 5000. If the server itself is behind a firewall, the 1024 to 5000 port range must be open for incoming connections at the firewall. This requirement can expose the server to potential attack because many applications share the WinSock dynamic port range. To reduce the attack surface, you can configure the FTP service to allocate ports for passive-mode connections from a port range above 5000, which allows you to assign a port range to be used exclusively by FTP passive-mode connections and to create firewall and router policies that open that range for incoming TCP connections.

To set the port range for passive-mode connections, edit the PassivePortRange property in the metabase. When you set the port range, consider the number of anticipated concurrent file transfers because each client might need a distinct port. For performance reasons, the port range should be at least two times the number of anticipated concurrent file transfers.

WEBDAV (Ports 80 and/or 443)

Research during the BAR-GC implementation determined that WebDAV would provide a simple, secure way for Data Providers to upload data to the BAR-GC sites. WebDAV is an industry-standard set of extensions to HTTP 1.1. WebDAV is an alternative to FTP for secure publishing of resources because it enables the same strong authentication, encryption, proxy support, and caching capabilities as any other HTTP-based Web site. In addition, with WebDAV, you can send multiple file transfers through a single TCP connection, whereas FTP requires a new connection for each file transferred.

WebDAV provides support for:

- *Editing*: creating, updating, deleting
- *Properties*: title, author, publication date, etc.

- *Collections*: analogous to a file system's directory or desktop folder
- *Locking*: prevents the confusion and data corruption caused by two or more people editing the same content at the same time

WebDAV is platform independent, both in terms of client and server. This means that Macintosh, Unix, and Windows users can collaborate on Web content without all the usual conversion problems. Furthermore, it doesn't matter whether your documents are hosted on an Apache or Microsoft IIS server. The two leading Web servers—Apache and Microsoft Internet Information Server—are compliant, as are document and content management systems from Documentum Inc., Vignette Corp. and BroadVision Inc.

Popular software tools that support DAV include Microsoft's Office XP, Office 2003 and SharePoint Portal Server, as well as Macromedia Inc.'s Dreamweaver, and Adobe Systems Inc.'s GoLive and Acrobat. WebDAV functionality is embedded in operating systems including Windows (from Windows 95 onward), Apple Computer Inc.'s Mac OS X and Novell Inc.'s Netware. Windows XP's integrated support enables any application running on it to be WebDAV-enabled.

WebDAV makes use of the standard authorization and authentication methods built right into every Web Server. In the same manner as one restricts access to a portion (whether a file, folder, or entire site) of one's Web site to a particular set of users or machines, so too can one finely tune WebDAV access to resources.

Currently, the BAR-HSDS configuration requires all WebDAV users authenticate and use HTTPS 128-bit encryption for all communications. Data Providers have no FTP access, so all data transfers between the site and end user are encrypted without requiring any 3rd party applications.

Note that security of data transfer and storage can be further increase by combining WebDAV with Encryption File System (EFS). This is explained further in the EFS section below.

However, a more effective way to transfer files securely could entail utilizing the Encryption File System built into Windows. This would mean data is stored encrypted on the web server without needing to password protect / encrypt with WinZip.

EFS

Encryption File System (EFS) services are currently not utilized for BAR-HSDS, however, new functionality offered between Windows 2003 servers and Windows XP clients show the future direction of encrypted files transfers. Bottom-line, combining EFS and WebDAV allows for:

1. Storing encrypted files on the server without needing to password protect/encrypt with WinZip
2. Transferring server-side encrypted files over HTTP to a client
3. The client machine automatically un-encrypts the file for usage on the local user's machine

This distributes the burden of encryption to the client machines allowing for a more scalable security solution and improving performance. Both uploads to and downloads from Web folders are raw data transfers, so even if an attacker could access the data during the transmission of an encrypted file, the captured data would be encrypted and unusable.

EFS with Web folders eliminates the need for specialized software to securely share encrypted files between users, businesses, or organizations. Files can be stored on common intranet file servers or Internet communities for easy access while maintaining strong security through EFS.

CA

The BAR-HSDS implementation currently does not utilize a separate Public Key Infrastructure (PKI) and Certificate Authority (CA). Currently, a self-signed 128-bit SSL certificate is utilized on each server to provide encrypted HTTPS communication. The certificate was created with Microsoft's SelfSSL utility that comes with the Windows 2003 Resource Kit. Non-Repudiation could be incorporated into this solution by requiring user certificates for access to the site.

Self-Signed certificates provide additional changes such as:

- Requires adding the self-signed certificate to the Trusted Authorities list for server Java components using the Keytool. Note that Tomcat, ArcIMS, and the Java SDK potentially all have separate Java directories.
- Java client applications such as ArcExplorer fail to operate against HTTPS sites that have self-signed certificates unless the certificate is added to the Trusted Authorities list for the applications Java with Keytool.
- Think clients such as ArcMap cannot access self-signed encrypted ArcIMS web services unless the certificate is added to the system's Trusted Authority list. The easiest way to add the certificate is through a browser, access the secured site, and when the security alert dialog appears, view the certificate, choose install, and select all the defaults.

RDP (Port 3389/tcp/udp)
The Remote Desktop Protocol (RDP) is utilized for remote server administration and by Data Integrators for running ETL processes and applications on the database server. Opening RDP at the firewall should be avoided where possible as the protocol allows for full administration of servers. However, to simplify the BAR-HSDS deployment, VPN is currently not required for RDP users. RDP access requires:

- The remote client having it's IP address in the approved list for RDP access (currently managed at the Firewall level, but can be supplemented at the server level)
- Authenticating to the domain with a valid Active Directory username and password
- The Active Directory user must be a member of the Administrator group or Data Integrator Group
- If the user is a Data Integrator, they only have RDP access to the database server as a standard user (non-administrative rights) and no access to the web server

Long-term, RDP access requirements by Data Integrators should reduced and possibly eliminated as more ETL functions are automated, and also ported to web based interfaces as necessary. RDP access by remote system administrators should be encapsulated within VPN and possibly utilize a token for 2-factor authentication to significantly improve the security of the systems.

SMTP (Port 25/tcp Outbound)
SMTP will be utilized to send notice to Data Integrators of when there is new data on the server at the primary site that needs to be loaded. Additional email notifications will be implemented as they become necessary. Limiting SMTP as an outbound only protocol reduces the likelihood of system compromise and eliminates usage as an external "junk mail" relay server.

NTP (Port 123/udp):
To help ensure the BAR-HSDS servers time all stay in synchronization, they utilize the Network Time Protocol (NTP) service. NTP is a protocol designed to synchronize the clocks of computers over a network. The servers have been configured to synchronize time with the servers from the National Institute of Standards and Technology (NIST) in Boulder, Colorado, listed in Figure 4-12.

Figure 4-12
NTP Servers Utilized by BAR-HSDS Server

FQDN	IP Address
time-a.timefreq.bldrdoc.gov	132.163.4.101
time-b.timefreq.bldrdoc.gov	132.163.4.102
time-c.timefreq.bldrdoc.gov	132.163.4.103

4.3.2 Application Details

Details for the following applications are provided in this section:

- Tomcat – Servlet connector
- ESRI ArcIMS – WebServices
- ESRI JITK – BAR-GC Viewer

Tomcat

Many Project Homeland pilots utilize IIS for web access and Tomcat as the "Connector" for ArcIMS. It is worthwhile to note that a potentially missed security item is to prevent access to Tomcat over the cleartext port 8080. In other words, even though a firewall might prevent port 8080 traffic with Internet users, a client located within the firewall could have cleartext communication with Tomcat over port 8080. Two techniques to prevent this are:

a. Use an IP Sec filter on the server with Tomcat that restricts port 8080 calls from and to the server itself. This is in addition to an external firewall not allowing port 8080 traffic to pass.

b. The other option is to configure Tomcat to use SSL which reduces application performance. Instructions to do this are at: http://jakarta.apache.org/tomcat/tomcat-4.0-doc/ssl-howto.html

For the BAR-HSDS pilot, the IPSec filter option was selected to ensure there was not a reduction in performance, and minimize risk of application compatibility with a secure Tomcat configuration.

The above issue in addition to Tomcat's use of the servlet invoker, suggests that utilizing commercial products such as ServletExec, instead of Tomcat, could potentially provide a more secure system. Note that the Tomcat creators refer to the invoker portion of code as "evil" and ask developers to avoid using it.

ESRI ArcIMS

ArcIMS map services are not only utilized by the BAR-GC Viewer, but may also be accessed via HTTPS with think client applications such as:

- ArcCatalog
- ArcMap
- ArcReader
- ArcGlobe
- ArcScene

Note that ArcPad currently does not support HTTPS web services, but this functionally is slated for availability by June 2005.

ESRI JITK

ESRI's Java Interoperability Tool Kit (JITK) allows for extending ArcIMS functionality further then ever before. The BAR-GC Viewer's advanced functionality is due to the usage of JITK. JITK blurs the lines of functionality offered with ArcIMS versus the newer ArcGIS Server. Clients may both view and download map data via the BAR-GC viewer with a web browser. Note that the BAR-GC viewer is currently limited to usage by Internet Explorer 5 or later.

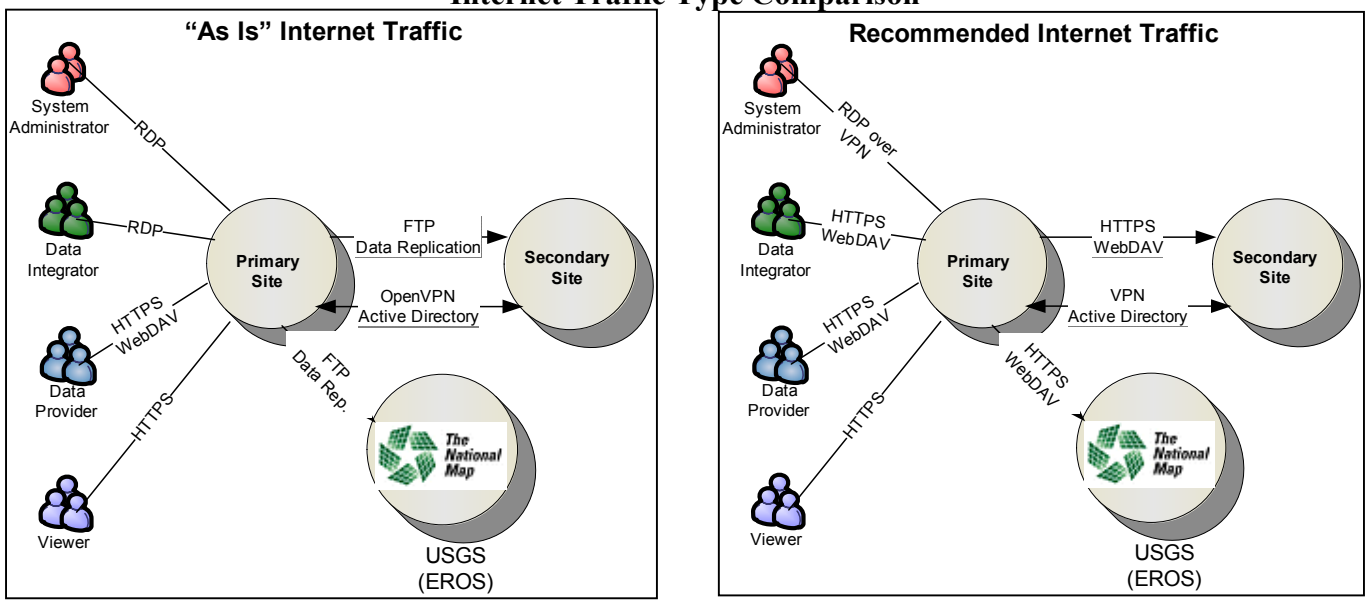
4.3.3 Optimizing Internet Traffic

A key component to the success of a nationwide GIS data repository is minimizing the amount of traffic that needs to cross network lines. Currently, the BAR-HDSD pilot utilizes multiple protocols to traverse the Internet.

- RDP is used for both server administration and by Data Implementers for running ETL processes
- FTP is used to transfer WinZip compressed/encrypted dataset replication information
- OpenVPN is used to encapsulate Active Directory information for user account and systems management replication
- HTTPS is used for BAR-GC viewer users
- HTTPS via WebDAV is used for Data Providers and potentially the USGS for file uploads/downloads

Figure 4-13 compares the current Internet traffic configuration for the BAR-GC HSDS vs. a potentially more secure, higher performing recommendation.

Figure 4-13
Internet Traffic Type Comparison



The future implementation recommendations for Internet traffic are as follows:

- RDP is used only for server administration and is encapsulated within VPN
- Server-side ETL functions are automated so that Data Integrators only need HTTPS access to the server
- Dataset replication information is transferred via HTTPS and WebDAV – Eliminating FTP transfers
- VPN is used to encapsulate Active Directory information for user account and systems management replication
- HTTPS is used for BAR-GC viewer users
- HTTPS via WebDAV is used for Data Providers and the USGS for file uploads/downloads

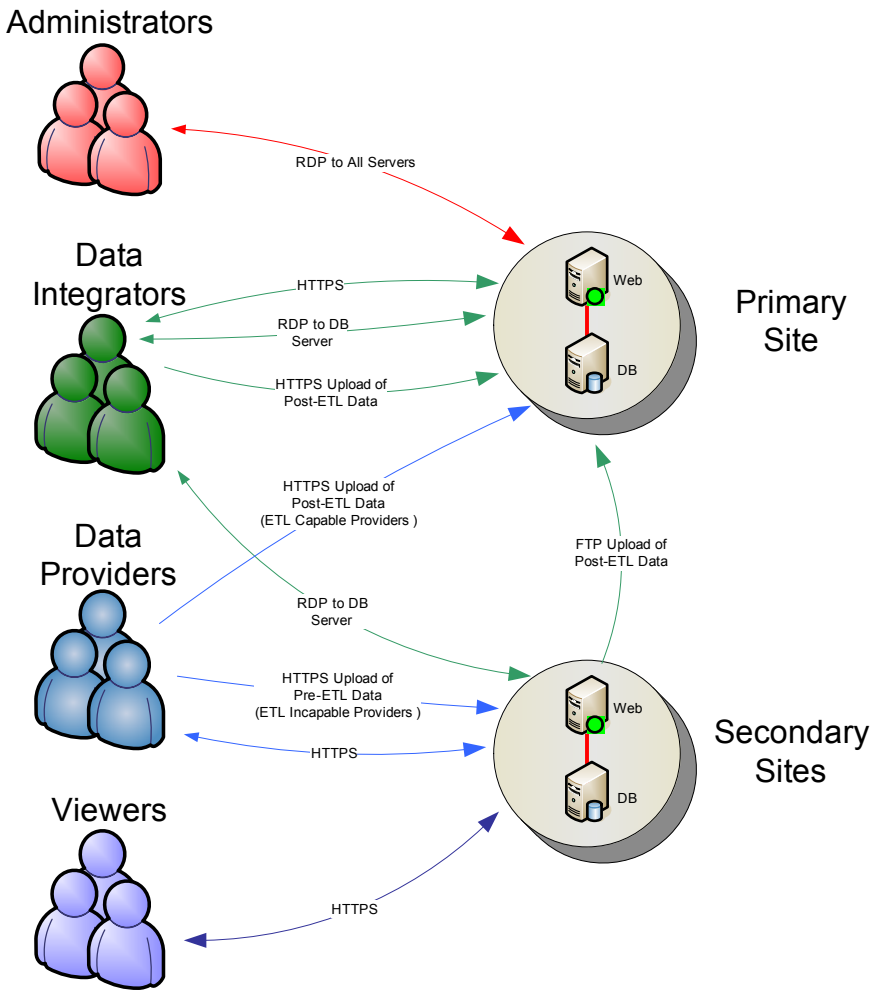
4.4 Management

Management priorities in the enterprise have led to the acceleration and deepening of relationships between business needs and network operations. A primary goal of network managers is the mapping of network infrastructure components to business processes, together with the need to refocus the development of management metrics and event processing rules to address SLAs based on business priorities. The BAR-GC application and systems utilized the same management system, Windows Active Directory. This section describes the various roles of system/application users and the mechanisms used to transfer management information between host sites.

4.4.1 User Roles

Defining role-based management privileges in an organization facilitates clear separation of management responsibilities. Figure 4-14 below provides an overview of the various user roles, the protocols/access they may utilize, and the systems they access.

Figure 4-14
Communication Mechanisms Available for Each Role



Viewers

Website access only

Allowed to log on and access the BAR-GC viewer. Currently, a self-signed 128-bit SSL certificate is utilized on each server to provide encrypted HTTPS communication. The certificate was created with Microsoft's SelfSSL utility that comes with the Windows 2003 Resource Kit.

Data Providers
Website + WebDAV Uploads only
Same permissions as Viewer plus the ability to upload data to a secondary site for Pre-ETL Data and to the Primary site for Post-ETL data. 128-bit AES Winzip encryption will protect files in transport. There are two types of Data Providers:

- ETL Capable providers who have the tools & knowledge to run the extract and transformations on their local machines, they then upload the Post-ETL data
- ETL Incapable providers upload their raw (Pre-ETL) data to a secondary site

Data Integrators
Website + WebDAV + RDP access to database server – Known IP Address
Same permissions and Data Provider plus allowed to run ETL processes on secondary site database servers and final Load process on Primary site database server. This user will initially have admin rights on the database server, but these rights should be reduced after the initial implementation.

Two Administrator Types

Single Site Administrator
Website access + RDP admin access to servers at one site – Known IP Address
Each site has an assigned Domain administrator as determined by the site. The site may change who at their site is assigned to this role.

Full Administrator
Website access + RDP admin access to all BAR-GC servers – Known IP Address
Has Forest level administrator privileges. This will initially only be assigned to Vestra contacts and will be turned over to an appropriate user as BAR-GC specifies.

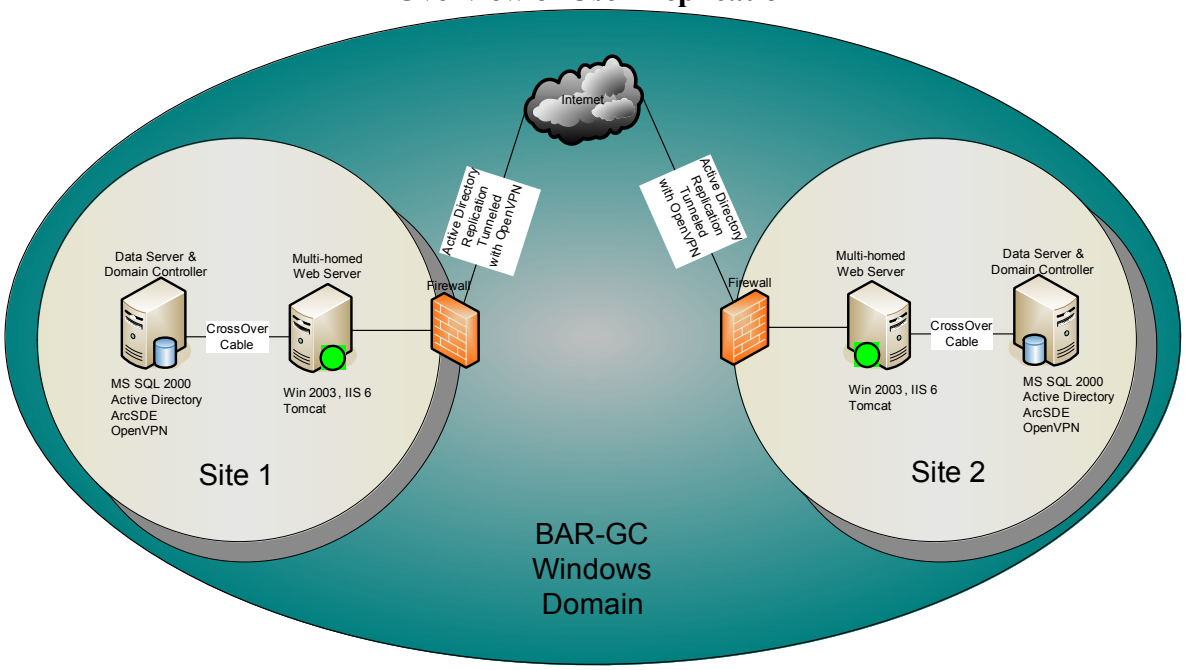
4.4.2 User Replication

To help simplify providing an LDAP compatible solution, Microsoft's Active Directory, is utilized for Account Synchronization, see Figure 4-15. This single identity management system provides access permissions for:

- BAR-GC Viewer
- Server System Administrators
- FTP
- WebDAV folders

Note that SQL server accounts can also be controlled under the Active Directory structure, but this has not been implemented for BAR-HSDS.

Figure 4-15
Overview of User Replication



Active Directory configuration includes:

- Consolidating system administration and application account information into one authentication system
- Active Directory chosen as an industry standard secure repository for user authentication data
- Active Directory account information will be synchronized between all 4 BAR-GC sites through the Windows 2003 "Synchronous" (RPC) mode over OpenVPN tunnels. Synchronization traffic is limited through use of Windows Site schedule configuration.
- The Active Directory structure will be maintained on the database servers which will be configured as Domain Controllers.
- Application interfaces protected with Windows Accounts include the FTP sites, Viewer application, Web Services, and Remote Desktop Protocol.

Future Options

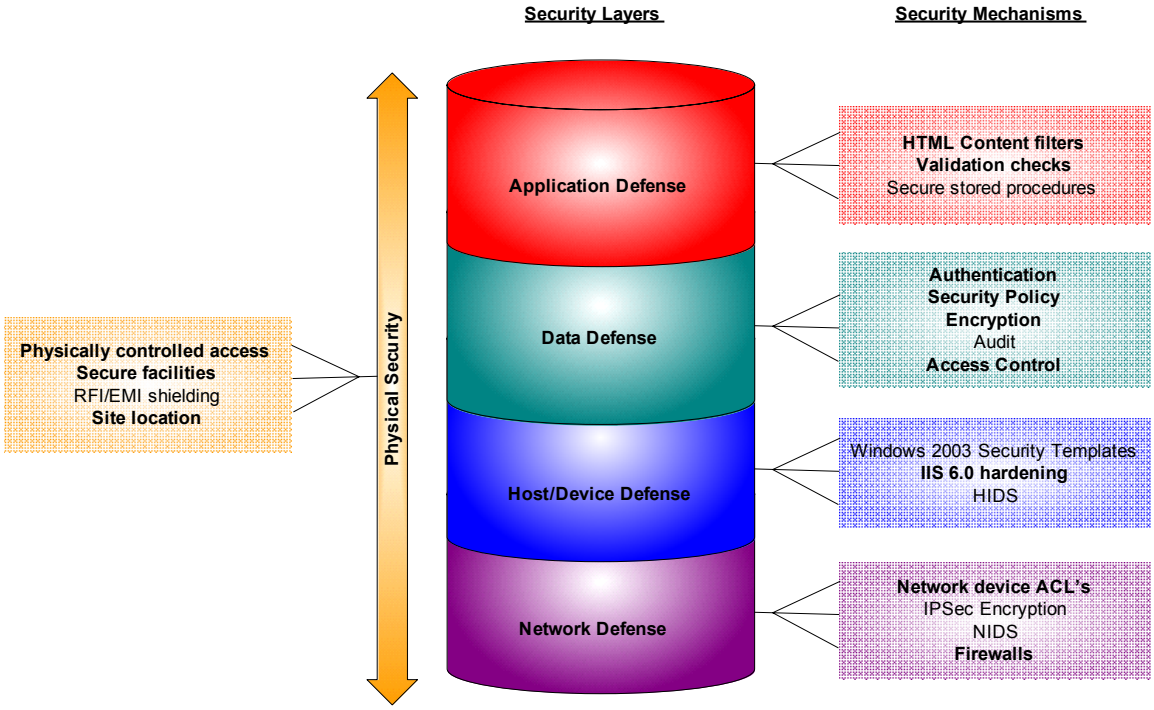
- Further utilize more Windows Group Policies to distribute and enforce standards across all systems.
- Compatible with Secure Map which allows for securing each map layer separately.
- This solution lends itself to future growth potentially utilizing a certificate PKI for client authentication and the use of layer level authentication with such products as SecureMap.

4.5 Security

The BAR-HSDS is intended to handle unclassified material only, although some of its contents may, by their nature and/or their aggregation, fall under "sensitive but unclassified" (SBU). Instead of implementing an MLS system for supporting SBU and unclassified data separately, all data is protected at the level of SBU. Federal Information Protection Standards (FIPS) such as 197 provide our guidance for the handling of SBU data. FIPS 197 may be used by Federal departments and agencies when an agency determines that sensitive (unclassified) information requires cryptographic protection. A key requirement is that all data transferred across public networks be encrypted with at least 128-bit level.

A defense-in-depth strategy has been implemented for the BAR-HSDS pilot to help ensure a secure environment. Some of the mechanisms utilized to secure the environment are listed in Figure 4-16 below.

Figure 4-16
Risk Mitigation Technologies



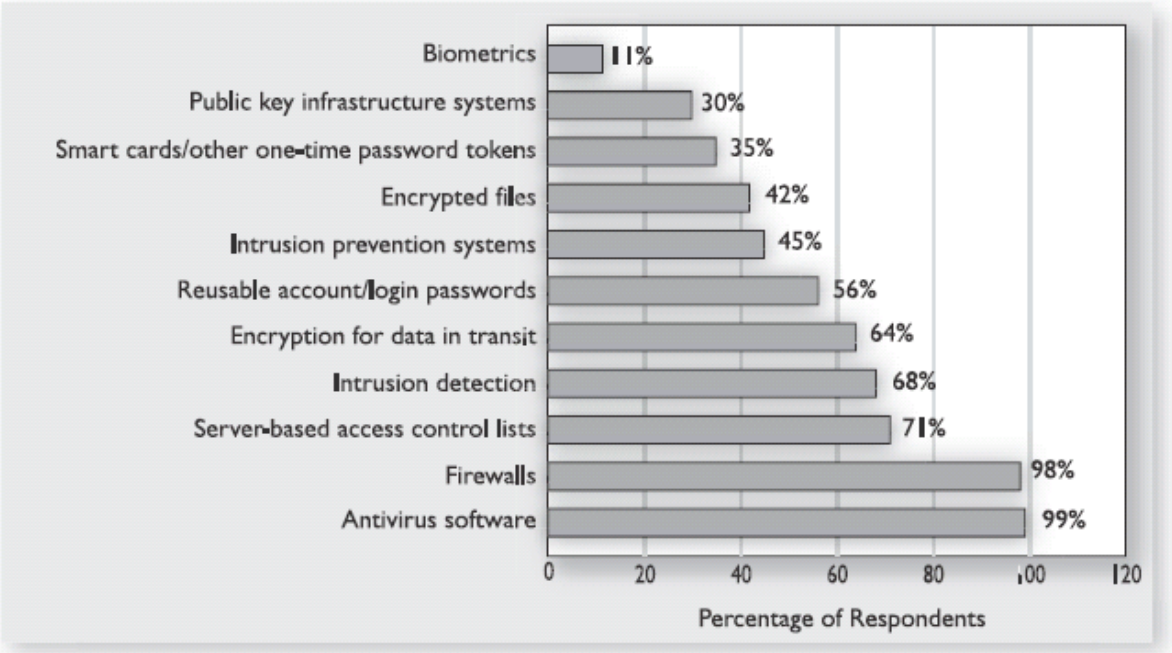
Note: Security mechanisms implemented for the BAR-GC project have been bolded in the figure above.

The following security mechanisms are reviewed in this section:

- Antivirus Software
- Access Control Lists
- Service Accounts
- Multi-homing
- Security Lockdowns
- Authentication
- Network Encryption

Figure 4-17 below shows how widely deployed various security mechanisms were in 2004.

Figure 4-17
Technologies Used to Secure Environments in 2004



4.5.1 Antivirus Software

It is strongly recommended to implement a standard host site solution across the server environment. To ensure the servers are provided to the customers "virus free" a temporary version of Norton Antivirus is initially installed on the web servers. Anti-Virus software should be installed on both web and database servers since data will be loaded periodically on to both.

4.5.2 Multi-homing

Multi-homing allows for separation of traffic, such as the separation of management traffic (generated by backup/restore and remote administration functions, for example) from the network adapter used for communication with the Internet. It also allows each interface to be connected to a separate network segment and facilitates TCP/IP port filtering on each of the hosts' interfaces.

The Web Server is Multi-homed to provide additional security and minimize dependence on customer switch hardware. Two ports are used on the web server, one connection runs to the firewall, and one crossover cable connects to the database server (1 Gbit/s connection).

4.5.3 Security Lockdowns

The Web Server has been locked down based on SANS 10 Top Issues with Windows systems, shown in Figure 4-18 below. Microsoft Baseline Security Analyzer V1.2.1 was run against all servers. MBSA is the free, best practices vulnerability assessment tool for the Microsoft platform. It is a tool designed for the IT Professional that helps with the assessment phase of an overall security management strategy.

Figure 4-18
SANS Top 10 Issues and Protection Steps Taken for HSDS

Issue #	Vulnerability	Protection Performed	Server
1	Web Services	Windows Auto Update turned on IIS Logging enabled Anti-Virus installed	Web
2	Workstation Service	N/A	N/A
3	Remote Access Service	Delete hidden shares/Reg hack Do not allow anonymous Limited access to registry	All
4	Microsoft SQL Server	Service Pack 3a + KB815495 Ran SQLScan Enabled Authentication Logging	Database
5	Windows Authentication	Strong password required+ History & Aging NTLMv2 only Prevent LM hash storage	All
6	Web browsers	Updates IE Security settings	All
7	File-Sharing Apps (P2P)	N/A	N/A
8	LSAS Exposures	N/A	N/A
9	Mail client	N/A	N/A
10	Instant Messaging	N/A	N/A

4.5.4 Access Control Lists

Access control works in conjunction with authentication. Once a client is authenticated, that client must have access to resources via NTFS permissions and web permissions to view content.

NTFS permissions allow administrators to control access to files and directories based on users and/or groups. NTFS permissions can provide strong granular control, but can become seemingly complicated when you have many users or groups to assign different sets of permissions.

Web permissions are set for and always include everyone accessing the BAR-HSDS site. Web permissions work in conjunction with NTFS permissions, although there is not the same granular level of control as with NTFS permissions. Web permissions apply to virtual directories and when a conflict in access between NTFS and Web permissions happens, the most restrictive setting takes precedence.

Access control lists should be further locked down as the BAR-HSDS solution moves out of a pilot to production solution.

4.5.5 Service accounts

The initial BAR-GC implementation does not utilize specialized accounts for services. This should be reviewed for future iterations of the environment.

4.5.6 Authentication

To simplify authentication requirements and ensure a robust solution, accounts are authenticated against the Microsoft Active Directory. Integrated Windows Authentication is required to provide a highly secure method for connecting a client to an IIS server. It uses a hash algorithm for sending credentials before sending them across the network. When the client is running a supported browser (such as IE) and the domain controllers are Windows 2000 or later, the default authentication mechanism is Kerberos v5.

Currently, security is not enforced at the GIS layer level and direct access to web services outside of the BAR-GC viewer is disabled.

Future Implementation Options

Layer level and web service level protection of data could be provided through such products as SecureMap from the St. George Consulting Group. The solution utilizes LDAP for authentication which is compatible with the current MS Active Dir solution. In addition to LDAP authentication at the layer level, it has full request logging capabilities, and authentication requirement for all ESRI clients and custom applications that use the servlet connector. The cost is approximately $5-6k/server installed. More info may be found at: http://www.stgeorgeconsulting.com/downloads/SecureMapBrochure_2004-09-13.pdf

4.5.7 Network Encryption

All HSDS web HTTPS traffic is protected with 128-bit SSL. Initially, the BAR-HSDS sites were set up with free 30-day certificates from InstantSLL which did not result in un-trusted site prompts because their root server is already trusted by client browsers. All sites utilize a different FQDN, therefore separate certificates are required for each host site. Verisign was not used for the test certificates because they last only 14 days.

BAR-GC will need to determine if they see the value of pursuing purchasing SSL certificates from a standard provider such as Verisign, continue utilizing self signed certificate, or implement/utilize a government managed Certificate Authority public key infrastructure.

Appendix A

Glossary

Appendix A—Glossary

NGA Homeland Security Enterprise GIS Support Project Glossary	
Architectural View	A view is a representation of a set of system elements and the relations associated with them. A view documents a particular aspect of the system's architecture while intentionally suppressing others. Different views will highlight different system elements and/or relationships.
Architecture	The organizational structure of a system or component.
Architecture Design	The process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.
Coding	The transforming of logic and data from design specifications (design descriptions) into a programming language.
Coding Standards	Written procedures describing coding (programming) style conventions specifying rules governing the use of individual constructs provided by the programming language and naming, formatting, and documentation requirements, which prevent programming errors, control complexity and promote understandability of the source code.
Conceptual Data Model	Represents the overall logical structure of a database, which is independent of any software or data storage structure. A conceptual model often contains data objects not yet implemented in the physical databases. It gives a formal representation of the data needed to run an enterprise or a business activity.
Configuration Management	A discipline applying technical and administrative direction and surveillance to identify and document the functional and physical characteristics of a configuration item, control changes to those characteristics, record and report change processing and implementation status, and verifying compliance with specified requirements.
Data Collaboration	Allows multi-directional sharing of data, files and services, allowing ideas and thoughts to be communicated interactively.
Data Dictionary	A collection of the names of all data items used in a software system, together with relevant properties of those items (e.g., length of data item, representation, etc.).
Data Flows	Data flows involve moving (and sometimes transforming) data among multiple entities.

NGA Homeland Security Enterprise GIS Support Project Glossary	
Data Model	In a general sense, an abstraction of the real world that incorporates only those properties thought to be relevant to the application at hand. It would normally define specific groups of entities, their attribute values, and the relationships between these. In GIS, it is often used to refer to the mechanistic representation and organization of spatial data, for example, the vector data model and the raster data model. It is independent of a computer system and its associated data structures.
Data Set	A collection of related records.
Data Synthesis	The acquisition and integration of data, files and Web services.
Data Validation	The checking of data for correctness or compliance with applicable standards, rules, and conventions.
Database	A collection of interrelated data, often with controlled redundancy, organized according to a schema to serve one or more applications. The data is stored so that different programs can use it without concern for the data structure or organization. A common approach is used to add new data and to modify and retrieve existing data.
Design	The process of defining the architecture, components, interfaces, and other characteristics of a system or component. See: architectural design, preliminary design, detailed design.
Developer	A person, or group, that designs and/or builds and/or documents and/or configures the hardware and/or software of computerized systems.
Enhancement	A change to a product, which is intended to make it better in some way (e.g., new functions, faster, or occasionally more compatible with other systems).
Flowchart or Flow diagram	A control flow diagram in which suitably annotated geometrical figures are used to represent operations, data, or equipment, and arrows are used to indicate the sequential flow from one to another.
Hardware	Physical equipment, as opposed to programs, procedures, rules, and associated documentation.
Installation	The phase in the system life cycle that includes assembly and testing of the hardware and software of a computerized system. Installation includes installing a new computer system, new software or hardware, or otherwise modifying the current system.

NGA Homeland Security Enterprise GIS Support Project Glossary	
Logical Data Model	This model attempts to describe the data in as much detail as possible, without regard to how the data will be physically implemented in the database. Features of the logical data model include the following: (1) All entities and relationships among them, (2) All attributes for each entity are specified, (3) The primary key for each entity is specified, (4) Foreign keys (keys identifying the relationship between different entities) are specified, and (5) Normalization occurs at this level.
Metadata	Information about data describing a collection of data. Metadata for geographical data may include the source of the data, its creation date and format, its projection, scale, resolution, and accuracy, and its reliability with regard to some standard.
Meta-Model	A data model for metadata
Pilot	Serves as an original model upon which future development will be based.
Prototyping	Using software tools to accelerate the software development process by facilitating the identification of required functionality during analysis and design phases. A limitation of this technique is the identification of system or software problems and hazards.
Quality Assurance	The planned systematic activities necessary to ensure that a component, module, or system conforms to established technical requirements.
Quality Control	The operational techniques and procedures used to achieve quality requirements.
Quality Management System	Storage system for the quality processes, procedures, templates, and other documents used by ESRI's Database Services division. These processes, and their use, are in compliance with ISO 9001.
Release	The formal notification and distribution of an approved version.
Requirements Analysis	(1) The process of studying user needs to arrive at a definition of a system, hardware, or software requirements. (2) The process of studying and refining system, hardware, or software requirements.
Risk Assessment	A comprehensive evaluation of the risk and its associated impact.
Software	Programs, procedures, rules, and any associated documentation pertaining to the operation of a system.
Source Code	The human readable version of the list of instructions (program) that cause a computer to perform a task.

NGA Homeland Security Enterprise GIS Support Project Glossary	
Specification	A document that specifies, in a complete, precise, verifiable manner, the requirements, design, behavior, or other characteristics of a system or component, and often, the procedures for determining whether these provisions have been satisfied.
Spiral Development	A family of software development processes characterized by repeatedly iterating a set of elemental development processes and managing risk so it is actively being reduced.
System Architecture	The architecture for a system is the structure or structures of the system, which comprise components, their externally visible behavior, and the relationships among them.
Test Case	Documentation specifying inputs, predicted results, and a set of execution conditions for a test item.
Test Plan	Documentation specifying the scope, approach, resources, and schedule of intended testing activities. It identifies test items, the features to be tested, the testing tasks, responsibilities, resources, and any risks requiring contingency planning.
Test Report	A document describing the conduct and results of the testing carried out for a system or system component.
Testing	The process of analyzing a software item to detect the differences between existing and required conditions (i.e., bugs), and to evaluate the features of the software items.
User Guide	Documentation that describes how to use a functional unit and that may include a description of the rights and responsibilities of the user, the owner, and the supplier of the unit.
User Interface	The aspects of a computer system or program with which a user can interact and the commands and mechanisms the user uses to control its operation and input data.

Appendix B

Acronyms and Definitions

Appendix B—Acronyms and Definitions

NGA Homeland Security Enterprise GIS Support Project Acronyms and Definitions	
ADL	Alexandria Digital Library
ABCI	Arizona Border Control Initiative
AMOC	Customs and Enforcement Air and Marine Operations Center
AS	Application Server
ASP	Active Server Pages
AV	All-Views
BAR-GC	Bay Area Regional GIS Council
CAD	Computer-Aided Design
CBP	U.S. Customs and Border Protection
CIFS	Common Internet File System
CIGT	Center for Innovative Geospatial Technology
CIP	Critical Information Protection
CLIN	Contract Line Item Number
COMFORCE	Common Enforcement Environment
COP	Common Operating Picture
COTS	Commercial Off-the-Shelf
CPFF	Cost-Plus-Fixed-Fee
CS-W	Catalog Services—Web
DEM	Digital Elevation Model
DGN	Intergraph Design File
DHS	Department of Homeland Security
DoD	Department of Defense
DODAF	Department of Defense Architecture Framework
DS	Data Server
DTC	Defense Technology Center
DWG	Autocad Drawing File
ESRI	Environmental Systems Research Institute, Inc.
ETL	Extraction, Translation, and Loading
FGDC	Federal Geographic Data Committee
FTP	Fire Transfer Protocol
GCCS	Global Command and Control System
GFE	Government-Furnished Equipment

NGA Homeland Security Enterprise GIS Support Project Acronyms and Definitions	
GFI	Government-Furnished Information
GIS	Geographic Information System
HIFLD	Homeland Infrastructure Foundation Level Database
HLD	Homeland Defense
HLS	Homeland Security
HSIP	Homeland Security Infrastructure Program
HSPD-7	Homeland Security Presidential Directive 7
ICMP	Internet Control Message Protocol
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
JWICS	Joint Worldwide Intelligence Communications System
LAN	Local Area Network
LIDAR	Light Detection and Ranging
MEDS	Minimum Essential Data Set
MSOP	Mission Specific Operating Picture
NFS	Network File System
NGA	National Geospatial-Intelligence Agency
NIMA	National Imagery and Mapping Agency
NIPRNET	Unclassified but Sensitive IP Router Network
NOAA	National Oceanic and Atmospheric Administration
NORTHCOM	United States Northern Command
NRE	Nonrecurring Engineering
NSDI	National Spatial Data Infrastructure (U.S.)
NSSE	National Special Security Event
OV	Operational View
OWL	Ontology Web Language
POC	Point of Contact
QA	Quality Assurance
QMS	Quality Management System
RDBMS	Relational Database Management System
RDF	Research Description Framework
RO	Regional Office
SAML	Security Assertion Markup Language

NGA Homeland Security Enterprise GIS Support Project Acronyms and Definitions	
SBU	Sensitive But Unclassified
SDSFIE	Spatial Data Standard for Facilities, Infrastructure, and Environment
SEI	Software Engineering Institute
SIPRNET	Secret Internet Protocol Router Network
SOA	Services-Oriented Architecture
SOAP	Simple Object Access Protocol
SOW	Statement of Work
SSE	Special Security Event
SSL	Secure Sockets Layer
SV	System View
TASC	The Analytical Sciences Corporation
TNM	*The National Map*
TNMC	*The National Map* Commercial
TNMP	*The National Map* Public
TNMS	*The National Map* Sensitive
TV	Technical Standard View
UML	Unified Markup Language
USGS	United States Geological Survey
USOP	Unit Specific Operating Picture
W3C	World Wide Web Consortium
WAN	Wide Area Network
WBS	Work Breakdown Structure
WFS	Web Feature Service
WMS	Web Map Service
WS	Web Server
XCBF	XML Common Biometric Format
XrML	eXtensible rights Markup Language

Appendix K

Bay Area Regional GIS Council Viewer Release Notes

Bay Area Regional GIS Council
Homeland Security Data Server
Pilot Project

BAR-GC Viewer
Release Notes

Draft

ESRI Professional Services
March 2005

Prepared for:

National Geospatial-Intelligence Agency
4600 Sangamore Road
Bethesda, Maryland 20816-5003
Tel.: 301-287-6572

Prepared by:

ESRI
380 New York Street
Redlands, California 92373-8100

Document History

Revision History

Date and Sections	Description	Person Responsible
12/8/2004	Baseline document	Beata Van Esch
12/17/2004	Updates per testing	Beata Van Esch
2/9/2005	Updates for installation and configuration	Beata Van Esch/Ricardo Trujillo

Table of Contents

1.0 About This Document

1.1 Purpose

ESRI has worked with the Bay Area Regional GIS Council (BAR-GC) to identify and prioritize enhancements to be made to the BAR-GC Viewer. This document describes the release of the BAR-GC Viewer.

1.2 Intended Audience of This Document

The primary audience for this document is ESRI Project Homeland management staff, BAR-GC staff, and NGA.

1.3 Document Overview

This document is organized into four sections.

Section 1.0—About This Document: Provides introduction to Release Notes.

Section 2.0—Release Overview: Defines functionality in release, issues resolved, release process, and outstanding issues with release

Section 3.0—Release Package: Describes release media and installation instructions.

Section 4.0—Contact Information: Provides contact information for installation issues.

1.4 Assumptions, Constraints, and Dependencies

Installation is dependent upon ESRI staff or subcontractor for on-site configuration and deployment.

1.5 Referenced Documents

The following documents served as input into the release of the BAR-GC Viewer.

Document Name	Date	Document Owner
Bay Area Regional GIS Council Homeland Security Data Server Pilot Project, BAR-GC Viewer Test Plan	November 2004	Beata Van Esch
Bay Area Regional GIS Council Homeland Security Data Server Pilot Project, BAR-GC Viewer Test Results	November 2004	Beata Van Esch

2.0 Release Overview

2.1 Functionality Included in This Release

The following functionality, based on enhancement requests for BAR-GC Viewer, has been included for this release.

ID	Description
1	Integrate ArcIMS Data Delivery Extension (DDE) (download data to various formats).
2	Add ability to click on the map and return coordinates in other projections. ■ UTM 10 ■ CA State Plane 2 ■ CA State Plane 3 ■ Teale Albers ■ Geographic WGS 84 ■ Geographic NAD 27
3	Palanterra™ Spiral 1 baseline functionality minus the following functionality: ■ Geocoding ■ Reverse Geocoding ■ Routing ■ Find Address ■ Top Organizational Buttons ■ National Geospatial-Intelligence Agency (NGA) Logos

2.2 Major Issues Resolved in This Release

Issues resolved include meeting the three enhancement requests (see Section 2.1) and resolving 24 defects that were identified during testing.

2.3 Test Strategy Applied to This Release

ESRI followed the Quality Management System (QMS) processes for software testing. A test matrix was developed from the requirements and was used as the basis for testing. ClearQuest (CQ) was used to track requirements and defects identified in the system.

2.4 Known Issues

The following enhancement requests have not been met due to rejection, postponement, or known limitations.

- All requirements as outlined in Section 2.1 have been met.
- BAR-GC Viewer is based on an earlier version of the Palanterra™ GUI.

The following defects have not been resolved due to postponement or known limitations.

Known Limitations

ID	Description
CQ00254793	Data: Find by county but no county data
CQ00257647	Legend minimize
CQ00257941	Coordinate system not applied on many file types

Postponed

ID	Description
CQ00255417	QB: Cannot find results using spatial query option "contains the center of"

2.5 Additional Release Information

The BAR-GC Viewer is being released to BAR-HSDS host server sites as a courtesy by NGA. ESRI will make no further code or maintenance updates to the BAR-GC Viewer.

3.0 Installation Instructions

This section provides information for the configuration of the JITK (Java Integration Tool Kit) on a Web server. The JITK serves as the Java building blocks for the BAR-GC Viewer.

3.1 Prerequisites

Minimum Requirements

1. Java Servlet 2.2 specification (Java Servlet 2.3 specification ***recommended***).
2. JDK (Java Development Kit) 1.3.1 or higher.
3. ArcIMS 3.1 or higher.
4. A compatible Web server.

Supported Web Browser

IE version 5.5 or higher*

* JavaScript must be enabled for full functionality support.

3.2 Getting Started

1. Before installation and configuration

   1.1 Make sure that your ArcIMS Map Service(s) is (are) fully functioning.

   1.2 Create a ***profiles.xml*** configuration file using JITK Administrative tools.

   1.3 Create a virtual directory where image files can be served to client (Usually if ArcIMS is installed on the same machine then the <ArcIMS>\Output directory will suffice).

   1.4 Create a log file directory.

Configuration Files

2. All the configuration files are located in the \WEB-INF directory.

2.1 *ssx_config.xml.*

2.1.1 This file is used to configure most of JITK's behavior.

2.1.2 This file is divided into four main sections. The factory (`FACTORIES`) and builder (`BUILDERS`) sections specify the implementation classes to be used for certain components and what factory classes should be used to create them. The ArcIMS server (`AIMS_SERVERS`) and service (`AIMS_SERVICES`) sections are used to configure the ArcIMS Map Services used by the application. The data source (`DATA_SOURCES`) and ArcSDE data source (`SDE_DATA_SOURCES`) sections are used to configure the DBMS and ArcSDE database connection settings. Finally, the initialization parameters (`INIT_PARAMS`) section provides a slew of parameters that configure various aspects of the core functionality.

2.2 *profiles.xml*

2.2.1 This configuration file acts as a filtering proxy to an underlying map service that only exposes a subset of the layers and geographic area.

2.2.2 The application relies on six database tables to maintain configuration settings specific to profiles, themes, theme groups, table registrations, and column mappings. The contents of these tables are read into memory by the application at start-up; therefore, any changes to these tables require that the application be restarted before they can take effect.

2.2.3 This file is created using the JITK Administraive tools.

2.3 *geocode_config.xml*

2.3.1 The "geocode_config.xml" file is used to configure settings specific to the geocoding functionality. Currently, there is only an initialization parameters section, which is used to configure the rendering properties of geocoded points drawn on the map.

2.4 *route_config.xml*

2.4.1 The *"route_config.xml"* file is used to configure settings specific to the routing functionality. Currently, there is only an initialization parameters section, which is used to configure the rendering properties of routes drawn on the map.

2.5 arcweb_config.xml

2.5.1 The arcweb_config.xml file is used to configure settings specific to the ArcWeb Services functionality. Currently, there is only an initialization parameters section, which is used to configure security settings (user name/password) that are required for integration with ArcWeb Services.

2.6 select_config.xml

2.6.1 The select_config.xml file is used to configure settings specific to the theme selection functionality. Currently, there is only an initialization parameters section where settings such as selection rendering can be configured.

2.7 tr_config.xml

2.7.1 The tr_config.xml file is used to configure settings specific to the thematic rendering functionality. Currently, there is only an initialization parameters section, which is used to configure the swatch cleaning task.

2.8 web.xml

2.8.1 The web.xml file is the Web application deployment descriptor defined as part of the Java Servlet Specification. A few sections of this file directly affect the application configuration. Context parameters are used for settings that must be known to the application before the actual configuration files are read. To initialize and finalize the application, either listeners or on-load servlets can be used. While listeners are the preferred choice, they may only be used if the application is being deployed in a servlet container that conforms to the Java Servlet 2.3 Specification. Other servlets used in the application must also be configured here. For more information on configuring a Web application with the web.xml deployment descriptor, please reference the Java Servlet Specification made available by Sun.

2.9 error.properties

2.9.1 The error.properties file essentially maps error codes to messages. This allows the error messages that are presented to the user to be easily changed without recompiling any code. Please note that the root logger can be configured to read the mappings in from a source other than the properties file, for instance, a database table.

Custom Configuration Files (BAR-GC Viewer Configuration Files)

2. These configuration files are also located in the \WEB-INF directory but are customized for each individual instance of the JITK.

3.1 event_config.xml

3.1.1 The event_config.xml file is used to configure settings specific to the BAR-GC Viewer event functionality. Currently, there is only an initialization parameters section where settings event table name, column names, and so on, can be configured.

3.2 gems_config.xml

3.2.1 The *gems_config.xml* file is used to configure settings specific to the BAR-GC Viewer event notification functionality. Currently, there is only an initialization parameters section where settings event pull interval, data source key, and so on, can be configured.

3.3 columnAlias.properties

3.3.1 The columnAlias.properties is used to map the BAR-GC Viewer event layer's column names to specified column aliases.

3.4 init.properties

3.4.1 The init.properties is used to map configuration settings specific to the BAR-GC Viewer's DDE functionality.

Quick Configuration Guide

4.1 This section will walk through a basic configuration of the JITK.

4.1.1 Open up each specified file with a text editor and save the file after each section.

4.2 Modifying the ssx_config.xml file.

4.2.1 Log file configuration

4.2.1.1 Near the top, search for the `<FACTORY type="core" key="logger">` XML tag. Within this tag, there is an `<PROPERTY_NAME>logFile</PROPERTY_NAME>` attribute.

4.2.1.2 Modify the `<PROPERTY_VALUE>[log file path]\[log file name]</PROPERTY_VALUE>` of the above XML tag to specify where the log file will reside and the name of the log file.

4.2.1.3 Example

```
<FACTORY type="core" key="logger">
<FACTORY_CLASS>com.esri.sln.ss.core.util.log.DailyRollingFileL
og4jLoggerFactoryImpl</FACTORY_CLASS>
      <SET_PROPERTY>
        <PROPERTY_NAME>errorMapFile</PROPERTY_NAME>
        <PROPERTY_VALUE>/WEB-
INF/error.properties</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>logFile</PROPERTY_NAME>
        <PROPERTY_VALUE>D:\logs\BAR-GC.log</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
   …
```

4.2.2 ArcIMS Servers

4.2.2.1 Search for the `<AIMS_SERVERS>` XML tag that is located near the middle of the document.

4.2.2.2 Here you will find a `<AIMS_SERVER key="imsserver1" type="com.esri.sln.ss.core.aims.Base40AimsServerImpl">` followed by some `<SET_PROPERTY>` tags.

4.2.2.3 The "key" attribute associated with the `<AIMS_SERVER>` tag is used in the `<AIMS_SERVICE>` tag configuration (Section 4.2.3) to reference the ArcIMS server configuration specified here.

4.2.2.4 Modify the "host", "port", and "connectionType" <PROPERTY_VALUE> XML tag for each of these XML pairing to specify your correct ArcIMS configuration.

4.2.2.5 Example **(Single ArcIMS Server Configuration)**

```
<AIMS_SERVERS>

      <AIMS_SERVER
          key="imsserver1"
          type="com.esri.sln.ss.core.aims.Base40AimsServerImpl">
        <SET_PROPERTY>
          <PROPERTY_NAME>host</PROPERTY_NAME>
          <PROPERTY_VALUE>mod.esri.com</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>port</PROPERTY_NAME>
          <PROPERTY_VALUE>5300</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>connectionType</PROPERTY_NAME>
          <PROPERTY_VALUE>TCP</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>swapAxlEncoding</PROPERTY_NAME>
          <PROPERTY_VALUE>UTF-8</PROPERTY_VALUE>
        </SET_PROPERTY>
      </AIMS_SERVER>

</AIMS_SERVERS>
      …
```

4.2.2.6 If multiple ArcIMS servers are required, then simply add a new <AIMS_SERVER> XML tag with all the necessary attributes within the <AIMS_SERVERS> parent XML tag.

4.2.2.7 Example **(Multiple ArcIMS Server Configuration)**

```
<AIMS_SERVERS>

   <AIMS_SERVER
        key="imsserver1"
        type="com.esri.sln.ss.core.aims.Base40AimsServerImpl">
      <SET_PROPERTY>
        <PROPERTY_NAME>host</PROPERTY_NAME>
        <PROPERTY_VALUE>mod.esri.com</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>port</PROPERTY_NAME>
        <PROPERTY_VALUE>5300</PROPERTY_VALUE>
      </SET_PROPERTY>
```

```
            <SET_PROPERTY>
              <PROPERTY_NAME>connectionType</PROPERTY_NAME>
              <PROPERTY_VALUE>TCP</PROPERTY_VALUE>
            </SET_PROPERTY>
            <SET_PROPERTY>
              <PROPERTY_NAME>swapAxlEncoding</PROPERTY_NAME>
              <PROPERTY_VALUE>UTF-8</PROPERTY_VALUE>
            </SET_PROPERTY>
          </AIMS_SERVER>

        <!--Example of regular ArcIMS server config using ArcIMS 3.1
libraries -->
        <AIMS_SERVER key="aims31_1"
type="com.esri.sln.ss.core.aims.Base31AimsServerImpl">
          <SET_PROPERTY>
            <PROPERTY_NAME>swapAxlEncoding</PROPERTY_NAME>
            <PROPERTY_VALUE>UTF-8</PROPERTY_VALUE>
          </SET_PROPERTY>
          <SET_PROPERTY>
            <PROPERTY_NAME>host</PROPERTY_NAME>
            <PROPERTY_VALUE>161.149.220.102</PROPERTY_VALUE>
          </SET_PROPERTY>
          <SET_PROPERTY>
            <PROPERTY_NAME>port</PROPERTY_NAME>
            <PROPERTY_VALUE>5300</PROPERTY_VALUE>
          </SET_PROPERTY>
        </AIMS_SERVER>

      </AIMS_SERVERS>
          …
```

4.2.3 ArcIMS Map Services

4.2.3.1 Search for the `<AIMS_SERVICES>` XML tag that is usually located right after the `<AIMS_SERVERS>` XML tag near the middle of the document.

4.2.3.2 Within this XML tag, you will find `<AIMS_SERVICE key="bar_ov" server_key="imsserver1" service_name="bar_o" is_arcmap="false" request_all_layers="false" unconfigured_layer_visibility="false" />` tags that are associated with ArcIMS Map Services.

4.2.3.3 The `key="bar_ov"` attribute is a reference from the configuration file "*profiles.xml*" to this particular map service. In the *"profiles.xml"*, the `<THEME>` tag attribute `aims_service_key="bar_demo"` must match this attribute

field if the theme configured is part of this configured map service.

4.2.3.4 The `server_key="imsserver1"` attribute is associated with an ArcIMS server configuration, which was configured in Section 4.2.2.

4.2.3.5 The `service_name="bar_o"` attribute is the actual map service name.

4.2.3.6 The `is_arcmap="false"` attribute specifies whether this map service is an ArcMap map service. Valid values are `"true"` or `"false."`

4.2.3.7 The `request_all_layers="false"` attribute specifies whether to request the entire configured map service's layers when doing an AXL request. This request will override whatever is configured in the "*profiles.xml*" file. Valid values are `"true"` or `"false."`

4.2.3.8 The `unconfigured_layer_visibility="false"` attribute tag specifies whether all the layers requested by the `request_all_layers="true"` attribute will be either visible or not. Valid values are `"true"` or `"false."`

4.2.3.9 The attribute `data_frame="dataFrame1"` specifies the source of the ArcMap map service **only** if the map service is an ArcMap map service. This attribute is used in case there are multiple data frames in an MXD. The tag can be any string literal.

4.2.3.10 Example (Single ArcIMS Map Service Configuration)

In ssx_config.xml

```
    …
<AIMS_SERVICES>
     <AIMS_SERVICE
         key="bar_ov"
         server_key="imsserver1"
         service_name="bar_o"
         is_arcmap="false"
         request_all_layers="false"
         unconfigured_layer_visibility="false" />
```

```
</AIMS_SERVICES>
    …
```

In profiles.xml

```
  …
      <THEME can_be_active="N" def_active="N" def_visible="Y"
disp_in_legend="Y" description="County" theme_id="59"
disp_name="County" aims_service_key="bar_ov" aims_layer_id="0"
draw_order="1" avail_cd="QD">
  …
```

4.2.3.11 Example (Multiple ArcIMS Map Service Configuration)

In ssx_config.xml

```
    …
<AIMS_SERVICES>
     <AIMS_SERVICE
         key="bar_ov"
         server_key="imsserver1"
         service_name="bar_o"
         is_arcmap="false"
         request_all_layers="false"
         unconfigured_layer_visibility="false" />
<AIMS_SERVICE
         key="bar_demo"
         server_key="imsserver1"
         service_name="bar_demo_axl"
         is_arcmap="false"
         request_all_layers="false"
         unconfigured_layer_visibility="false" />
</AIMS_SERVICES>
    …
```

In profiles.xml

```
  …
      <THEME can_be_active="N" def_active="N" def_visible="Y"
disp_in_legend="Y" description="County" theme_id="59"
disp_name="County" aims_service_key="bar_ov" aims_layer_id="0"
draw_order="1" avail_cd="QD">
  …
       <THEME can_be_active="Y" def_active="N" def_visible="Y"
disp_in_legend="Y" description="Building_Complex" theme_id="53"
disp_name="Building_Complex" aims_service_key="bar_demo"
aims_layer_id="Building_Complex" draw_order="56" avail_cd="QD">
  …
```

4.2.4 Data Source

4.2.4.1 A data source is a database connection. Here is where any database configurations take place. The JITK supports many types of data sources (MS SQL Server, Oracle, DB2, etc.)

4.2.4.2 Search for the `<DATA_SOURCES>` XML tag that is usually located right after the `<AIMS_SERVICES>` XML tag near the middle of the document.

4.2.4.3 Within the `<DATA_SOURCES>` tag, there is a `<DATA_SOURCE key="bargc" type="com.esri.sln.share.dbms.BasicJdbcDataSource" >` with a series of `<PROPERTY_NAME>` attributes such as "driverName", "user name", and "password."

4.2.4.4 The `key="bargc"` attribute on the `<DATA_SOURCE` tag is a reference from the "*profiles.xml*" file to this database configuration. In the "*profiles.xml*", there is an XML tag `<TABLE_REG>,` which has an attribute `data_source_key="bargc",` which references this database connection as a layer's source.

4.2.4.5 There are two common types of data sources:

4.2.4.5.1 Oracle Data Source

4.2.4.5.1.1 The `<PROPERTY_NAME>driverType</PROPERTY_NAME>` tag is setting the correct Oracle driver type that the JITK will use to connect to the database.

4.2.4.5.1.2 The `<PROPERTY_NAME>serverName</PROPERTY_NAME>` tag is used to configure the server name for the Oracle connection.

4.2.4.5.1.3 The `<PROPERTY_NAME>portNumber</PROPERTY_NAME>` tag is used to configure the port number for the Oracle Connection.

4.2.4.5.1.4 The `<PROPERTY_NAME>databaseName</PROPERTY_NAME>` tag is used to configure the database name for the Oracle Connection.

4.2.4.5.1.5 The `<PROPERTY_NAME>user</PROPERTY_NAME>` tag is used to configure the user name for the Oracle Connection.

4.2.4.5.1.6 The `<PROPERTY_NAME>password</PROPERTY_NAME>` tag is used to configure the password for the Oracle Connection.

4.2.4.5.1.7 Example

In ssx_config.xml

```
...
<DATA_SOURCE key="cpa"
type="oracle.jdbc.pool.OracleConnectionPoolDataSource">
        <SET_PROPERTY>
          <PROPERTY_NAME>driverType</PROPERTY_NAME>
          <PROPERTY_VALUE>thin</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>serverName</PROPERTY_NAME>
          <PROPERTY_VALUE>server</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>portNumber</PROPERTY_NAME>
          <PROPERTY_VALUE>1521</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>databaseName</PROPERTY_NAME>
          <PROPERTY_VALUE>database</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>user</PROPERTY_NAME>
          <PROPERTY_VALUE>user</PROPERTY_VALUE>
        </SET_PROPERTY>
```

```
            <SET_PROPERTY>
              <PROPERTY_NAME>password</PROPERTY_NAME>
              <PROPERTY_VALUE>pass</PROPERTY_VALUE>
            </SET_PROPERTY>
          </DATA_SOURCE>
        …
```

In profiles.xml

```
        …
      <TABLE_REG table_reg_id="7" data_source_key="cpa"
sde_data_source_key="sdecpa" owner="CPA"
table_name="CPA.BUILDING_FOOTPRINT" description=""
definition_query="" allow_jdbc_queries="Y" />
        …
```

4.2.4.5.2 MS SQL Server Data Source

4.2.4.5.2.1

The `<PROPERTY_NAME>driverName</PROPERTY_NAME>` tag is setting the correct JDBC driver to use when connecting to the specified database.

4.2.4.5.2.2

The `<PROPERTY_NAME>jdbcUrl</PROPERTY_NAME>` tag is the JDBC URL needed to connect to the MS SQL Server database.

4.2.4.5.2.2.1

The URL is usually set up with “jdbc:odbc:[ODBC data source name].

4.2.4.5.2.2.2

The ODBC data source name is configured in the "Microsoft ODBC Administrator."

4.2.4.5.2.3

The `<PROPERTY_NAME>user</PROPERTY_NAME>` tag is used to configure the user name for the MS SQL Server connection.

4.2.4.5.2.4

The `<PROPERTY_NAME>password</PROPERTY_NAME>` tag is used to configure the password for the MS SQL Server connection.

4.2.4.5.2.5 Example

In ssx_config.xml

```
  …
 <DATA_SOURCE key="bargc"
type="com.esri.sln.share.dbms.BasicJdbcDataSource">
      <SET_PROPERTY>
        <PROPERTY_NAME>driverName</PROPERTY_NAME>

<PROPERTY_VALUE>sun.jdbc.odbc.JdbcOdbcDriver</PROPERTY_VALU
E>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>jdbcUrl</PROPERTY_NAME>
        <PROPERTY_VALUE>jdbc:odbc:bargc</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>userName</PROPERTY_NAME>
        <PROPERTY_VALUE>user</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>password</PROPERTY_NAME>
        <PROPERTY_VALUE>pass</PROPERTY_VALUE>
      </SET_PROPERTY>
    </DATA_SOURCE>
     …
```

In profiles.xml

```
      …
    <TABLE_REG table_reg_id="6" data_source_key="bargc"
sde_data_source_key="sdebargc" owner="BARGC"
table_name="BARGC.BUILDING_COMPLEX" description=""
definition_query="" allow_jdbc_queries="Y" />
     …
```

4.2.4.6 Multiple data sources can be configured to pull data from multiple databases if needed. They must all be contained within the <DATA_SOURCES> XML tag.

4.2.4.6.1 Example

In ssx_config.xml

```
  …
 <!-- ================== Data Sources ================== --
>
   <DATA_SOURCES>
  <DATA_SOURCE key="cpa"
  type="oracle.jdbc.pool.OracleConnectionPoolDataSource">
```

```
            <SET_PROPERTY>
              <PROPERTY_NAME>driverType</PROPERTY_NAME>
              <PROPERTY_VALUE>thin</PROPERTY_VALUE>
            </SET_PROPERTY>
            <SET_PROPERTY>
              <PROPERTY_NAME>serverName</PROPERTY_NAME>
              <PROPERTY_VALUE>server</PROPERTY_VALUE>
            </SET_PROPERTY>
            <SET_PROPERTY>
              <PROPERTY_NAME>portNumber</PROPERTY_NAME>
              <PROPERTY_VALUE>1521</PROPERTY_VALUE>
            </SET_PROPERTY>
            <SET_PROPERTY>
              <PROPERTY_NAME>databaseName</PROPERTY_NAME>
              <PROPERTY_VALUE>database</PROPERTY_VALUE>
            </SET_PROPERTY>
            <SET_PROPERTY>
              <PROPERTY_NAME>user</PROPERTY_NAME>
              <PROPERTY_VALUE>user</PROPERTY_VALUE>
            </SET_PROPERTY>
            <SET_PROPERTY>
              <PROPERTY_NAME>password</PROPERTY_NAME>
              <PROPERTY_VALUE>pass</PROPERTY_VALUE>
            </SET_PROPERTY>
 </DATA_SOURCE>

 <DATA_SOURCE key="bargc"
type="com.esri.sln.share.dbms.BasicJdbcDataSource">
      <SET_PROPERTY>
        <PROPERTY_NAME>driverName</PROPERTY_NAME>

<PROPERTY_VALUE>sun.jdbc.odbc.JdbcOdbcDriver</PROPERTY_VALU
E>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>jdbcUrl</PROPERTY_NAME>
        <PROPERTY_VALUE>jdbc:odbc:bargc</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>userName</PROPERTY_NAME>
        <PROPERTY_VALUE>user</PROPERTY_VALUE>
      </SET_PROPERTY>
      <SET_PROPERTY>
        <PROPERTY_NAME>password</PROPERTY_NAME>
        <PROPERTY_VALUE>pass</PROPERTY_VALUE>
      </SET_PROPERTY>
    </DATA_SOURCE>
 </DATA_SOURCES>
```

…

In profiles.xml

…

```
<TABLE_REG table_reg_id="7" data_source_key="cpa"
sde_data_source_key="sdecpa" owner="CPA"
table_name="CPA.BUILDING_FOOTPRINT" description=""
definition_query="" allow_jdbc_queries="Y" />
…
<TABLE_REG table_reg_id="6" data_source_key="bargc"
sde_data_source_key="sdebargc" owner="BARGC"
table_name="BARGC.BUILDING_COMPLEX" description=""
definition_query="" allow_jdbc_queries="Y" />
…
```

4.2.5 SDE Data Sources

4.2.5.1 An SDE data source is an SDE connection. Here is where any SDE configurations take place.

4.2.5.2 Search for the `<SDE_DATA_SOURCES>` XML tag that is usually located right after the `<DATA_SOURCES>`XML tag near the middle of the document.

4.2.5.3 Within the `<SDE_DATA_SOURCES>` tag, there is a `<SDE_DATA_SOURCE key="sdebargc" type="com.esri.sln.share.sde.PooledSdeDataSourceImpl">` with a series of `<PROPERTY_NAME>` attributes such as "server", "instance", and "database."

4.2.5.4 The `key="sdebargc"` attribute on the `<SDE_DATA_SOURCE` tag is a reference from the "*profiles.xml*" file to this SDE configuration. In the "*profiles.xml*", there is an XML tag `<TABLE_REG>,` which has an attribute `sde_data_source_key="sdebargc",` which references this SDE connection as a layer's source.

4.2.5.5 The `<PROPERTY_NAME>server</PROPERTY_NAME>` tag is used to configure the server name for the SDE connection.

4.2.5.6 The `<PROPERTY_NAME>instance</PROPERTY_NAME>` tag is used to configure the SDE instance for the SDE connection.

4.2.5.7 The `<PROPERTY_NAME>database</PROPERTY_NAME>` tag is used to configure the database name for the SDE connection.

4.2.5.8 The `<PROPERTY_NAME>userName</PROPERTY_NAME>` tag is used to configure the user name for the SDE connection.

4.2.5.9 The `<PROPERTY_NAME>password</PROPERTY_NAME>` tag is used to configure the password for the SDE connection.

4.2.5.10 The `<PROPERTY_NAME> maxPoolSize</PROPERTY_NAME>` tag is used to configure the maximum Pooled SDE connections.

4.2.5.11 The `<PROPERTY_NAME> timeout</PROPERTY_NAME>` tag is used to configure the maximum time (in seconds) an SDE connection will attempt to connect until it fails.

4.2.5.12 The `<PROPERTY_NAME> validate</PROPERTY_NAME>` tag is used to make sure that the SDE connection is first validated before attempting the actual SDE connection. Valid values are `"true"` or `"false."`

4.2.5.13 Example

In ssx_config.xml

```
        …
<!-- ================== SDE Data Sources ================== -->
    <SDE_DATA_SOURCES>

      <SDE_DATA_SOURCE key="sdebargc"
type="com.esri.sln.share.sde.PooledSdeDataSourceImpl">
        <SET_PROPERTY>
          <PROPERTY_NAME>server</PROPERTY_NAME>
          <PROPERTY_VALUE>host</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>instance</PROPERTY_NAME>
          <PROPERTY_VALUE>5151</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>database</PROPERTY_NAME>
          <PROPERTY_VALUE>bargc</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>userName</PROPERTY_NAME>
          <PROPERTY_VALUE>user</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>password</PROPERTY_NAME>
```

```
            <PROPERTY_VALUE>pass</PROPERTY_VALUE>
          </SET_PROPERTY>
          <SET_PROPERTY>
          <PROPERTY_NAME>maxPoolSize</PROPERTY_NAME>
          <PROPERTY_VALUE>4</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>timeout</PROPERTY_NAME>
          <PROPERTY_VALUE>5000</PROPERTY_VALUE>
        </SET_PROPERTY>
        <SET_PROPERTY>
          <PROPERTY_NAME>validate</PROPERTY_NAME>
          <PROPERTY_VALUE>true</PROPERTY_VALUE>
        </SET_PROPERTY>
        </SDE_DATA_SOURCE>

    </SDE_DATA_SOURCES>
  …
```

In profiles.xml

```
          …
        <TABLE_REG table_reg_id="6" data_source_key="bargc"
sde_data_source_key="sdebargc" owner="BARGC"
table_name="BARGC.BUILDING_COMPLEX" description=""
definition_query="" allow_jdbc_queries="Y" />
          …
```

4.2.6 Image Output Path/Image Output URL Base

4.2.6.1 The image output path and the image output URL base are part of the slew of initialization parameters that follow the SDE data sources configuration section.

4.2.6.2 The image output path specifies where all the images that are created from the JITK image merging process are placed within the file system.

4.2.6.2.1 Search for the `<PARAM_NAME>imageOutputPath</PARAM_NAME>>` XML tag that is located near the bottom of the document. The tag is located with the slew of `<INIT_PARAM>` tags.

4.2.6.2.2 Modify the `<PARAM_VALUE>` for this XML pair mapping so that the image output path is set to directory where the Web server can access it. This should be a virtual directory that can be accessed through the Web.

4.2.6.2.3 If ArcIMS is installed on the same machine as the JITK viewer than using <ArcIMS>\Output directory will do.

4.2.6.2.4 Example

```
…
<INIT_PARAM>
   <PARAM_NAME>imageOutputPath</PARAM_NAME>
   <PARAM_VALUE>C:\\ArcIMS\\Output</PARAM_VALUE>
 </INIT_PARAM>
 <INIT_PARAM>
   <PARAM_NAME>imageOutputUrlBase</PARAM_NAME>
   <PARAM_VALUE>http://myserver/output</PARAM_VALUE>
 </INIT_PARAM>
 …
```

4.2.6.3 The image output URL base is the base URL that is used to form the URL for any images referenced by the JITK viewer. These images are usually the images produced by ArcIMS (map images, legend swatches, etc.).

4.2.6.3.1 Search for the `<PARAM_NAME>imageOutputUrlBase</PARAM_NAME>>` XML tag that is located near the bottom of the document. The tag is located with the slew of `<INIT_PARAM>` tags.

4.2.6.3.2 Modify the `<PARAM_VALUE>` for this XML pair mapping so that the image output URL path is set to a virtual directory URL where the Web server can access it.

4.2.6.3.3 If ArcIMS is installed on the same machine as the JITK viewer than using "*http://<my server>/output*" directory will do.

4.2.6.3.4 Example

```
…
<INIT_PARAM>
   <PARAM_NAME>imageOutputPath</PARAM_NAME>
   <PARAM_VALUE>C:\\ArcIMS\\Output</PARAM_VALUE>
 </INIT_PARAM>
 <INIT_PARAM>
   <PARAM_NAME>imageOutputUrlBase</PARAM_NAME>
   <PARAM_VALUE>http://myserver/output</PARAM_VALUE>
 </INIT_PARAM>
```

…

Image URL from Map Image on viewer

http://myserver/output/myMapService_MYSERVER33682380440.png

4.3 Modifying the *init.properties* file **(BAR-GC Configuration Only)***

4.3.1 DDE (Data Delivery Extension) Configuration

4.3.1.1 Locate the parameter `DDE_HOST` and `DDE_PORT` in the ***init.properties*** file.

4.3.1.2 Modify the `DDE_HOST` and `DDE_PORT` parameters to reflect your current DDE configuration

4.3.1.3 Set the `DDE_HOST` parameter equal to the actual DDE host name.

4.3.1.3.1 Example: `DDE_HOST = myserver`

4.3.1.4 Set the `DDE_PORT` parameter equal to the actual DDE port value.

4.3.1.4.1 Example: `DDE_PORT = 7071`

4.4 Modifying the *event_config.xml* file **(BAR-GC Configuration and Event Functionality Only)***

4.4.1 Event Theme ID

4.4.1.1 Locate the `<PARAM_NAME>themeId</PARAM_NAME>` tag near the top of the document.

4.4.1.2 Modify the `<PARAM_VALUE>` value to reflect the actual theme ID associated with the event layer that is configured in the ***profiles.xml***.

4.4.1.3 Example

In event_config.xml

…

```
<!-- Event Theme Id -->
```

```
    <INIT_PARAM>
      <PARAM_NAME>themeId</PARAM_NAME>
      <PARAM_VALUE>1</PARAM_VALUE>
    </INIT_PARAM>
    …
    In profiles.xml
     …
    <THEME can_be_active="N" def_active="N" def_visible="Y"
disp_in_legend="Y" description="County" theme_id="1"
disp_name="County" aims_service_key="bar_ov" aims_layer_id="0"
draw_order="1" avail_cd="QD">
```

4.4.2 Event SDE Data Source

4.4.2.1 Locate the `<PARAM_NAME>sdeDataSourceKey</PARAM_NAME>` tag near the top of the document.

4.2.2.2 Modify the `<PARAM_VALUE>` value to reflect the actual SDE data source key associated with the event layer that is configured in the ***ssx_config.xml***.

4.4.2.3 Example

```
In event_config.xml
    …
    <!-- Event Sde Data Source Key -->
    <INIT_PARAM>
      <PARAM_NAME>sdeDataSourceKey</PARAM_NAME>
      <PARAM_VALUE>sdebargc</PARAM_VALUE>
    </INIT_PARAM>
    …
    In ssx_config.xml
     …
     <!-- ==================== SDE Data Sources
==================== -->
    <SDE_DATA_SOURCES>

      <SDE_DATA_SOURCE key="sdebargc"
type="com.esri.sln.share.sde.PooledSdeDataSourceImpl">
        <SET_PROPERTY>
          <PROPERTY_NAME>server</PROPERTY_NAME>
          <PROPERTY_VALUE>host</PROPERTY_VALUE>
        </SET_PROPERTY>
      …
```

4.4.3 Event Data Source

4.4.3.1 Locate the `<PARAM_NAME>dbSourceKey</PARAM_NAME>` tag near the top of the document.

4.4.3.2 Modify the `<PARAM_VALUE>` value to reflect the actual data source key associated with the event layer that is configured in the ***ssx_config.xml***.

4.4.3.3 Example

In event_config.xml

```
 …
<!-- Event DB Source Key -->
<INIT_PARAM>
  <PARAM_NAME>dbSourceKey</PARAM_NAME>
  <PARAM_VALUE>bargc</PARAM_VALUE>
</INIT_PARAM>
 …
```

In ssx_config.xml

```
 …
   <!-- ================== Data Sources
================== -->
  <DATA_SOURCES>

    <DATA_SOURCE key="bargc"
type="com.esri.sln.share.dbms.BasicJdbcDataSource">
    <SET_PROPERTY>
      <PROPERTY_NAME>driverName</PROPERTY_NAME>

<PROPERTY_VALUE>sun.jdbc.odbc.JdbcOdbcDriver</PROPERTY_VALUE>
    </SET_PROPERTY>
    …
```

4.4.4 Event Table Name

4.4.4.1 Locate the `<PARAM_NAME>tableName</PARAM_NAME>` tag near the top of the document.

4.4.4.2 Modify the `<PARAM_VALUE>` value to reflect the actual table name of the event layer that is configured in the ***profiles.xml***.

4.4.4.3 Example

In event_config.xml

```
 …
<!-- Event Table Name -->
 <INIT_PARAM>
   <PARAM_NAME>tableName</PARAM_NAME>
   <PARAM_VALUE>BARGC.INCIDENT</PARAM_VALUE>
 </INIT_PARAM>
 …
```

In profiles.xml

```
 …
<TABLE_REG table_reg_id="4" data_source_key="bargc"
sde_data_source_key="bargc" owner="BARGC"
table_name="BARGC.INCIDENT" description="" definition_query=""
allow_jdbc_queries="Y" />
 …
```

4.5 Modifying the *gems_config.xml* file **(BAR-GC Configuration and Event Functionality Only)***

4.5.1 Event Theme ID

4.5.1.1 Locate the `<PARAM_NAME>eventThemeId</PARAM_NAME>` tag near the top of the document.

4.5.1.2 Modify the `<PARAM_VALUE>` value to reflect the actual theme ID associated with the event layer that is configured in the ***profiles.xml***.

4.5.1.3 Example

In gems_config.xml

```
…
 <!-- the id of the theme containing all Events -->
<INIT_PARAM>
  <PARAM_NAME>eventThemeId</PARAM_NAME>
  <PARAM_VALUE>1</PARAM_VALUE>
</INIT_PARAM>    …
```

In profiles.xml

```
 …
<THEME can_be_active="N" def_active="N" def_visible="Y"
disp_in_legend="Y" description="County" theme_id="1"
disp_name="County" aims_service_key="bar_ov" aims_layer_id="0"
draw_order="1" avail_cd="QD">
```

4.5.2 Event SDE Data Source

4.5.2.1 Locate the `<PARAM_NAME>sdeDataSourceKey</PARAM_NAME>` tag near the top of the document.

4.5.2.2 Modify the `<PARAM_VALUE>` value to reflect the actual SDE data source key associated with the event layer that is configured in the ***ssx_config.xml***.

4.5.2.3 Example

In gems_config.xml

```
…
<!-- Event Sde Data Source Key -->
<INIT_PARAM>
  <PARAM_NAME>sdeDataSourceKey</PARAM_NAME>
  <PARAM_VALUE>sdebargc</PARAM_VALUE>
</INIT_PARAM>
…
```

In ssx_config.xml

```
 …
 <!-- =================== SDE Data Sources
=================== -->
<SDE_DATA_SOURCES>

  <SDE_DATA_SOURCE key="sdebargc"
type="com.esri.sln.share.sde.PooledSdeDataSourceImpl">
    <SET_PROPERTY>
      <PROPERTY_NAME>server</PROPERTY_NAME>
      <PROPERTY_VALUE>host</PROPERTY_VALUE>
    </SET_PROPERTY>
  …
```

4.5.3 PIR Servlet URL

4.5.3.1 Locate the `<PARAM_NAME>pirServletURL</PARAM_NAME>` tag near the top of the document.

4.5.3.2 Modify the `<PARAM_VALUE>` value to reflect the URL of the viewer plus adding the "/servlet/pirservlet" to the end of it.

4.5.3.3 Example

In gems_config.xml

```
…
```

```
 <!-- PIRServlet URL -->
<INIT_PARAM>
   <PARAM_NAME>pirServletURL</PARAM_NAME>

<PARAM_VALUE>http://myserver/bargc/servlet/pirservlet</PARA
M_VALUE>
 </INIT_PARAM>
                    …
```

4.6 Final Steps

4.6.1 Save all modified files.

4.6.2 Repackage the Web application.

4.6.3 Deploy WAR file to Servlet Engine.

4.6.3.1 Warning: Different Servlet Engines come configured with different versions of standard Sun and Apache support packages (such as Xerces). Depending on the application server, some additional configurations at the application level or Servlet Engine level may be required.

4.0 Contact Information

Contact Sergio Rodriguez (909-793-2853, extension 2656, or e-mail srodriguez@esri.com), ESRI project manager, for project-related issues.

Appendix L

Glossary

Appendix M—Glossary

NGA Homeland Security Enterprise GIS Support Project Glossary	
Architectural View	A view is a representation of a set of system elements and the relations associated with them. A view documents a particular aspect of the system's architecture while intentionally suppressing others. Different views will highlight different system elements and/or relationships.
Architecture	The organizational structure of a system or component.
Architecture Design	The process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.
Coding	The transforming of logic and data from design specifications (design descriptions) into a programming language.
Coding Standards	Written procedures describing coding (programming) style conventions specifying rules governing the use of individual constructs provided by the programming language and naming, formatting, and documentation requirements, which prevent programming errors, control complexity and promote understandability of the source code.
Conceptual Data Model	Represents the overall logical structure of a database, which is independent of any software or data storage structure. A conceptual model often contains data objects not yet implemented in the physical databases. It gives a formal representation of the data needed to run an enterprise or a business activity.
Configuration Management	A discipline applying technical and administrative direction and surveillance to identify and document the functional and physical characteristics of a configuration item, control changes to those characteristics, record and report change processing and implementation status, and verifying compliance with specified requirements.
Data Collaboration	Allows multi-directional sharing of data, files and services, allowing ideas and thoughts to be communicated interactively.
Data Dictionary	A collection of the names of all data items used in a software system, together with relevant properties of those items (e.g., length of data item, representation, etc.).
Data Flows	Data flows involve moving (and sometimes transforming) data among multiple entities.
Data Model	In a general sense, an abstraction of the real world that incorporates only those properties thought to be relevant to the application at hand. It would normally define specific groups of entities, their attribute values, and the relationships between these. In GIS, it is often used to refer to the mechanistic representation and organization of spatial data, for example, the vector data model and the raster data model. It is independent of a computer system and its associated data structures.

NGA Homeland Security Enterprise GIS Support Project Glossary	
Data Set	A collection of related records.
Data Synthesis	The acquisition and integration of data, files and Web services.
Data Validation	The checking of data for correctness or compliance with applicable standards, rules, and conventions.
Database	A collection of interrelated data, often with controlled redundancy, organized according to a schema to serve one or more applications. The data is stored so that different programs can use it without concern for the data structure or organization. A common approach is used to add new data and to modify and retrieve existing data.
Design	The process of defining the architecture, components, interfaces, and other characteristics of a system or component. See: architectural design, preliminary design, detailed design.
Developer	A person, or group, that designs and/or builds and/or documents and/or configures the hardware and/or software of computerized systems.
Enhancement	A change to a product, which is intended to make it better in some way (e.g., new functions, faster, or occasionally more compatible with other systems).
Flowchart or Flow diagram	A control flow diagram in which suitably annotated geometrical figures are used to represent operations, data, or equipment, and arrows are used to indicate the sequential flow from one to another.
Hardware	Physical equipment, as opposed to programs, procedures, rules, and associated documentation.
Installation	The phase in the system life cycle that includes assembly and testing of the hardware and software of a computerized system. Installation includes installing a new computer system, new software or hardware, or otherwise modifying the current system.
Logical Data Model	This model attempts to describe the data in as much detail as possible, without regard to how the data will be physically implemented in the database. Features of the logical data model include the following: (1) All entities and relationships among them, (2) All attributes for each entity are specified, (3) The primary key for each entity is specified, (4) Foreign keys (keys identifying the relationship between different entities) are specified, and (5) Normalization occurs at this level.
Metadata	Information about data describing a collection of data. Metadata for geographical data may include the source of the data, its creation date and format, its projection, scale, resolution, and accuracy, and its reliability with regard to some standard.
Meta-Model	A data model for metadata
Pilot	Serves as an original model upon which future development will be based.

NGA Homeland Security Enterprise GIS Support Project Glossary	
Prototyping	Using software tools to accelerate the software development process by facilitating the identification of required functionality during analysis and design phases. A limitation of this technique is the identification of system or software problems and hazards.
Quality Assurance	The planned systematic activities necessary to ensure that a component, module, or system conforms to established technical requirements.
Quality Control	The operational techniques and procedures used to achieve quality requirements.
Quality Management System	Storage system for the quality processes, procedures, templates, and other documents used by ESRI's Database Services division. These processes, and their use, are in compliance with ISO 9001.
Release	The formal notification and distribution of an approved version.
Requirements Analysis	(1) The process of studying user needs to arrive at a definition of a system, hardware, or software requirements. (2) The process of studying and refining system, hardware, or software requirements.
Risk Assessment	A comprehensive evaluation of the risk and its associated impact.
Software	Programs, procedures, rules, and any associated documentation pertaining to the operation of a system.
Source Code	The human readable version of the list of instructions (program) that cause a computer to perform a task.
Specification	A document that specifies, in a complete, precise, verifiable manner, the requirements, design, behavior, or other characteristics of a system or component, and often, the procedures for determining whether these provisions have been satisfied.
Spiral Development	A family of software development processes characterized by repeatedly iterating a set of elemental development processes and managing risk so it is actively being reduced.
System Architecture	The architecture for a system is the structure or structures of the system, which comprise components, their externally visible behavior, and the relationships among them.
Test Case	Documentation specifying inputs, predicted results, and a set of execution conditions for a test item.
Test Plan	Documentation specifying the scope, approach, resources, and schedule of intended testing activities. It identifies test items, the features to be tested, the testing tasks, responsibilities, resources, and any risks requiring contingency planning.
Test Report	A document describing the conduct and results of the testing carried out for a system or system component.

NGA Homeland Security Enterprise GIS Support Project Glossary	
Testing	The process of analyzing a software item to detect the differences between existing and required conditions (i.e., bugs), and to evaluate the features of the software items.
User Guide	Documentation that describes how to use a functional unit and that may include a description of the rights and responsibilities of the user, the owner, and the supplier of the unit.
User Interface	The aspects of a computer system or program with which a user can interact and the commands and mechanisms the user uses to control its operation and input data.

Appendix M

Acronyms and Definitions

Appendix N—Acronyms and Definitions

NGA Homeland Security Enterprise GIS Support Project Acronyms and Definitions	
ADL	Alexandria Digital Library
ABCI	Arizona Border Control Initiative
AMOC	Customs and Enforcement Air and Marine Operations Center
AS	Application Server
ASP	Active Server Pages
AV	All-Views
BAR-GC	Bay Area Regional GIS Council
CAD	Computer-Aided Design
CBP	U.S. Customs and Border Protection
CIFS	Common Internet File System
CIGT	Center for Innovative Geospatial Technology
CIP	Critical Information Protection
CLIN	Contract Line Item Number
COMFORCE	Common Enforcement Environment
COP	Common Operating Picture
COTS	Commercial Off-the-Shelf
CPFF	Cost-Plus-Fixed-Fee
CS-W	Catalog Services—Web
DEM	Digital Elevation Model
DGN	Intergraph Design File
DHS	Department of Homeland Security
DoD	Department of Defense
DODAF	Department of Defense Architecture Framework
DS	Data Server
DTC	Defense Technology Center
DWG	Autocad Drawing File
ESRI	Environmental Systems Research Institute, Inc.
ETL	Extraction, Translation, and Loading
FGDC	Federal Geographic Data Committee
FTP	Fire Transfer Protocol
GCCS	Global Command and Control System
GFE	Government-Furnished Equipment
GFI	Government-Furnished Information
GIS	Geographic Information System
HIFLD	Homeland Infrastructure Foundation Level Database
HLD	Homeland Defense
HLS	Homeland Security

NGA Homeland Security Enterprise GIS Support Project Acronyms and Definitions	
HSIP	Homeland Security Infrastructure Program
HSPD-7	Homeland Security Presidential Directive 7
ICMP	Internet Control Message Protocol
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
JWICS	Joint Worldwide Intelligence Communications System
LAN	Local Area Network
LIDAR	Light Detection and Ranging
MEDS	Minimum Essential Data Set
MSOP	Mission Specific Operating Picture
NFS	Network File System
NGA	National Geospatial-Intelligence Agency
NIMA	National Imagery and Mapping Agency
NIPRNET	Unclassified but Sensitive IP Router Network
NOAA	National Oceanic and Atmospheric Administration
NORTHCOM	United States Northern Command
NRE	Nonrecurring Engineering
NSDI	National Spatial Data Infrastructure (U.S.)
NSSE	National Special Security Event
OV	Operational View
OWL	Ontology Web Language
POC	Point of Contact
QA	Quality Assurance
QMS	Quality Management System
RDBMS	Relational Database Management System
RDF	Research Description Framework
RO	Regional Office
SAML	Security Assertion Markup Language
SBU	Sensitive But Unclassified
SDSFIE	Spatial Data Standard for Facilities, Infrastructure, and Environment
SEI	Software Engineering Institute
SIPRNET	Secret Internet Protocol Router Network
SOA	Services-Oriented Architecture
SOAP	Simple Object Access Protocol
SOW	Statement of Work
SSE	Special Security Event
SSL	Secure Sockets Layer

NGA Homeland Security Enterprise GIS Support Project Acronyms and Definitions	
SV	System View
TASC	The Analytical Sciences Corporation
TNM	*The National Map*
TNMC	*The National Map* Commercial
TNMP	*The National Map* Public
TNMS	*The National Map* Sensitive
TV	Technical Standard View
UML	Unified Markup Language
USGS	United States Geological Survey
USOP	Unit Specific Operating Picture
W3C	World Wide Web Consortium
WAN	Wide Area Network
WBS	Work Breakdown Structure
WFS	Web Feature Service
WMS	Web Map Service
WS	Web Server
XCBF	XML Common Biometric Format
XrML	eXtensible rights Markup Language