Implementing a Metadata Catalog Portal in a GIS Network

An ESRI ® Technical Paper • March 2003

ESRI 380 New York St., Redlands, CA 92373-8100, USA • TEL 909-793-2853 • FAX 909-793-5953 • E-MAIL info@esri.com • WEB www.esri.com

Implementing a Metadata Catalog Portal in a GIS Network

An ESRI Technical Paper

Contents

Implementing a Metadata Catalog Portal in a GIS Network

Section 1: What is NSDI?

There is widespread recognition that the data layers and tables in most geographic information system (GIS) databases come from multiple organizations. Each GIS organization will develop some, but not all, of its data content, and other layers will come from external sources. This results in a system in which GIS data management is distributed among many users.

GIS users are hungry for quality geographic information; therefore, there is a fundamental need for users to share their data. Vast resources of information are available today, but the task of finding exactly what you need and knowing the quality and currency of the information can be daunting.

Furthermore, users often duplicate data collection work. While collaborative efforts are often used, massive duplication of GIS datasets is quite common and costly.

Bottom line: GIS users need a mechanism to efficiently catalog and search for available geographic information and to openly share this information.

The NSDI and GSDI vision

The NSDI, or National Spatial Data Infrastructure, is a concept developed by the United States federal government and other state and local governments. The Global Spatial Data Infrastructure (GSDI) is a nonprofit organization composed of individuals, industries, and more than 50 countries whose mission is to support ready global access to geographic information. The NSDI and GSDI promote the vision of a framework for GIS users to openly share geographic information and to collaborate on data development (Figure 1).

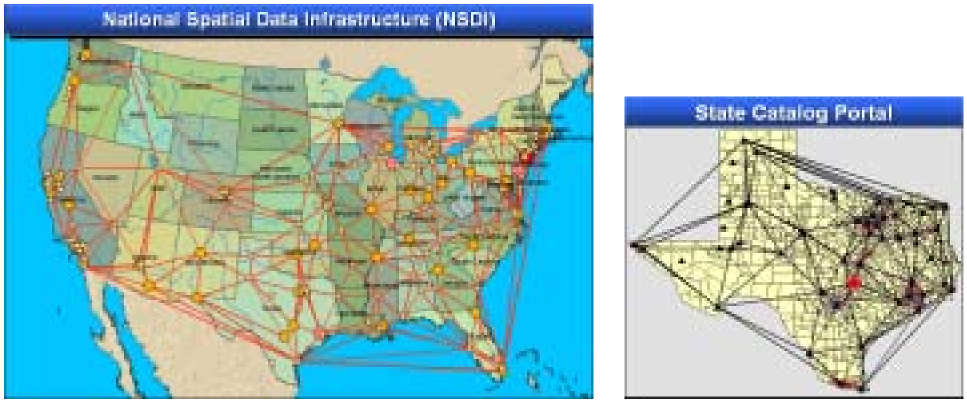

Figure 1. The vision of a National Spatial Data Infrastructure is based on GIS users openly sharing key sets of GIS information and the ability of users to search for and discover information sets across the World Wide Web. Users should also be able to access datasets through a variety of mechanisms—for example, via connections to data streaming services, access to map services, and data downloads in numerous desired formats.

The Spatial Data Infrastructure (SDI) concept describes requirements for computer technologies, policies, and people necessary to promote the sharing of geographic information throughout all levels of government, private industry, Non-Governmental Organizations (NGOs), and the academic community. The SDI interconnects GIS nodes across the Internet—and, in many cases, over secure networks—to share information with one another openly (i.e., based on the best available set of working, widely adopted practices and methods). These and other types of GIS portals are being built at national, state, and local levels for geographic information access and sharing. It is expected that many of these portals will be central sites that users can readily access and search.

A recent Federal Geographic Data Committee (FGDC) slide summarized many of these key visions:

- One-stop shop search system for geospatial data.
- Data is described through metadata and made searchable through many local catalogs known as Clearinghouse Nodes.
- Each catalog is registered with the FGDC.
- Ability to search any node.
- Includes free and for-fee data.
- An open, global network for GIS users, government, business, NGOs, etc.

How does an SDI work?

At its most basic level, an SDI is realized through catalogs holding documentation (metadata) about available GIS datasets and Web services. This is similar in function to the role of the card catalog in a library that references and organizes all library holdings. For example, GIS holdings can include static and interactive maps, map layers, applications, analytical models, reports, datasets and databases, and GIS Web services. Each of these documents can be documented using standards-based metadata.

Figure 2 shows three key activities of an SDI:

1. GIS data publishers document and register their datasets in a metadata catalog;
2. Users search for data at GIS catalog portals to find potential datasets for use. This is similar in concept to a library catalog search; and
3. Users directly connect to a GIS data service or download datasets for use in their application. Intensive data investigation (via live GIS data services) is often a crucial step before advanced GIS use.

Figure 2. The building blocks of the NSDI and GSDI architecture, containing the GIS catalog portal and GIS network concepts.

About this document

A Spatial Data Infrastructure requires:

- Catalogs of GIS data holdings (the building and hosting of a series of GIS catalog search portals)
- Content standards (adoption by user communities of common, multipurpose GIS data models)
- A commitment by data publishers to ensure a persistent trusted source of GIS information and services at their Web sites (the “geographic dial tone”)

Portals require much more than a metadata catalog that can be searched on the World Wide Web and a data delivery framework. This technical paper focuses on the first requirement: GIS catalog portals. As presented in this paper, catalog portals require a complete information system for geography, tools for data import and export, tools to validate data, tools to serve metadata catalogs and data, search and discovery tools, tools to explore and use data, and tools to build and manage GIS data.

In this paper, a framework for building open, interoperable GIS catalog portals will be presented. It covers the capabilities of the ArcGIS™ software currently available from ESRI and how these tools can be used in the critical aspects of building, serving, and using GIS catalogs in an SDI.

This document will also share some thoughts on practical methods that might be applied across the GIS profession to build such portals. The topics presented here are based on the results of a series of case studies developed through implementing the Geography Network℠ and a number of GIS catalog portals at user sites.

Section 2: What is a GIS catalog portal?

A metadata catalog for a GIS plays the same role as a card catalog in a library. The card catalog in a library has a record for each holding. All catalog records are compiled into one comprehensive catalog and indexed by author, subject, and title (and other keywords) so that the catalog can be searched. In a GIS, each dataset has a metadata document. The concept behind a GIS catalog portal is that these metadata documents are collected into a comprehensive metadata catalog that can be indexed by various means, such as by geographic location. The catalog portal can then be searched to find candidate datasets for use. Just as in a library search, each metadata record in a GIS catalog portal search can be viewed. Optionally, users can connect to and explore the actual dataset to make a decision about its potential use. This is a critical function necessary for advanced GIS applications.

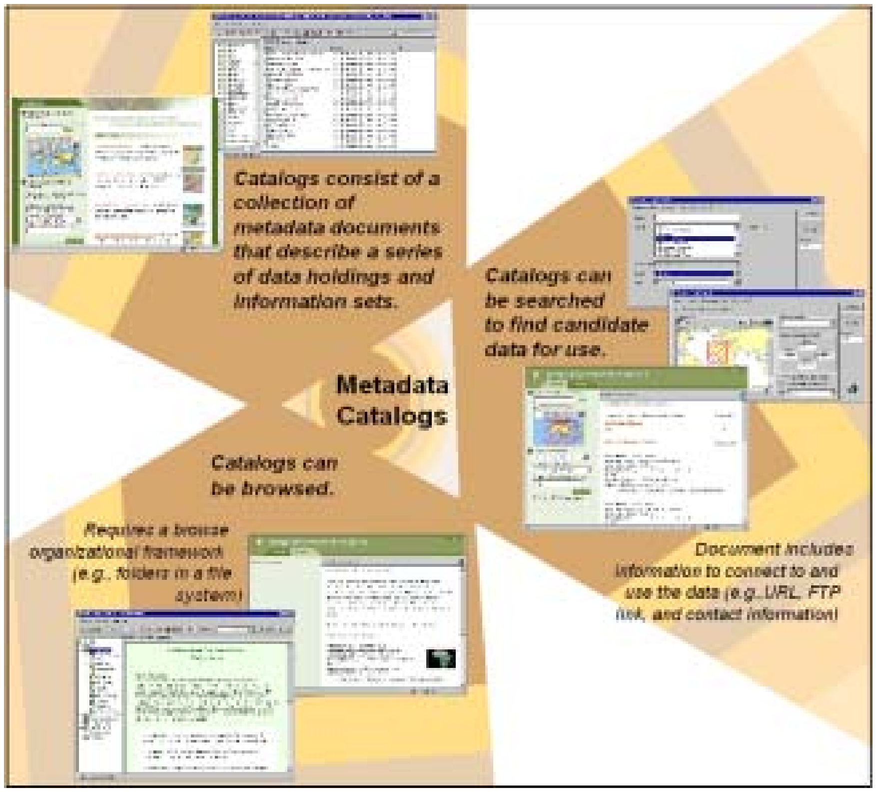

Figure 3. A metadata catalog contains data that can be searched or browsed.

It is also important to note that a library catalog references books as well as many other types of holdings. In turn, the information holdings in a GIS should include more than datasets, and each should be documented, indexed, and cataloged for searching and sharing. For example, GIS holdings can include:

- Static and interactive maps
- Map layers
- Three-dimensional scenes
- Applications
- Analytical models (e.g., geoprocessing scripts)
- Reports
- Datasets and databases
- Tables
- GIS Web services
- Activities and events
- And much more

The GIS metadata catalog can be thought of as the organizing framework for a GIS just as the card catalog reflects the organization of the library as a system, such as the organization of its shelves. Similarly, a GIS network and its catalog portals can be thought of as the organizing framework for a multiorganization GIS (e.g., a state or national GIS). Each GIS requires a catalog that documents its data holdings and provides fast access to them. The goal of metadata must be to create *consistent documentation* for these holdings so that searches work across GIS systems.

Example catalog search portals

There are already many examples of catalog portals—many of which provide GIS data—that support information search and discovery. A short review of a few very typical Web portals provides a helpful context to gain a broader understanding of what's available, how they work, and some of the key characteristics that will be important to build a working, practical Spatial Data Infrastructure.

General Web Searches	
www.google.com www.yahoo.com www.alltheweb.com www.lycos.com www.askjeeves.com	Commonly used World Wide Web search engines for Web content. Web searches are ad hoc—quite free form and designed to support many unforeseen purposes. As everyone has experienced, keyword searches for Web sites are hit and miss.
Library Catalog Searches	
http://catalog.loc.gov/	The Library of Congress Catalog search Web site. Library catalog searches are based on well-developed cataloging principles from the field of library science (e.g., standardized keyword lists, detailed cataloging guidelines, and ability to copy MARC records from publishers into each library catalog). This paper discusses the potential use of library science methods for building consistent GIS catalogs.
GIS Portals	
http://130.11.52.184/	The URL of the FGDC Clearinghouse, which is a collection of GIS sites, known as Clearinghouse Nodes. To search many catalogs, you perform a distributed search across selected nodes. The distributed search method has not worked well in practice for numerous reasons (e.g., search sites may not be online so no guaranteed results; each site takes a performance hit for every search; there is one result set to investigate for each search node, making comparisons difficult; performance is dismal). Recent studies indicate that search users connect directly to selected nodes and *browse* for data, which is also quite inefficient.

www.geographynetwork.com	ESRI's central catalog portal, where users can search for, discover, and directly connect to numerous GIS Web services and Data Download sites. The Geography Network references datasets and services for volunteer GIS participants. Additional sites have been added over time to help ESRI learn about best practices and methods. ESRI also uses this to better understand user requirements and better design capabilities in its commercial software.
www.geocommunicator.gov	The GIS catalog portal for the U.S. Bureau of Land Management and the U.S. Forest Service is built using the same technology as the Geography Network. Its goal is to provide an access portal to Federal Lands information including the Public Land Survey System (PLSS).
http://deli.dnr.state.mn.us/	Browsing each Web site with its own unique layout and design is the most typical access method used today for finding GIS data. This is the State of Minnesota's site for GIS data search and acquisition. Most GIS use at such sites is to search for data by browsing the Web pages at this site.
http://www.datamil.udel.edu/nationalmappilot/ http://www.tnris.org/explorer/index.htm	The State of Delaware's DataMIL site is an example Web site for searching and delivering data layers as part of the USGS *National Map*. Also listed is the State of Texas Geography Network portal for searching and finding GIS data State-wide. These are two in a series of GIS sites that provide modern GIS catalog services for searching.

UDDI Web Services Portals	
http://uddi.microsoft.com/default.aspx http://www-3.ibm.com/services/uddi/ http://uddi.sap.com/	The links listed here are examples of UDDI registries where Microsoft®, IBM®, and SAP provide search portals for SOAP-based Web services. UDDI, or Universal Description, Discovery and Integration, is a standards-based specification for registering and discovering SOAP-based Web services on the Web. UDDI is similar in concept to NSDI. UDDI is part of a computing industry initiative to create a platform-independent, open framework for describing services, discovering businesses, and integrating business services using the Internet. Each registry contains descriptions of businesses and Web services and will be used by entities to make information available about businesses and their service offerings. Generally speaking, UDDI registries are immature compared to GIS industry work on catalogs. However, the catalog of businesses and entities at these Web sites is worthy of continual review as Web services mature over time.
Commercial Search Portals	
http://www.amazon.com/	A nice example of a search and discovery Web site that provides useful overview information and recommendations. It's easy to find what you need and to discover more candidates based on your use history (e.g., "The page you made", "Customers who bought this title also bought").

Characteristics of successful GIS catalog portals

Ideally, GIS catalogs should support a search and discovery experience that more closely resembles a library search than an ad hoc Web search mechanism in broad use today and the current GIS data Web pages.

For an SDI to be practical, searches must work consistently and efficiently from site to site, datasets must be easy to document and the metadata documentation must be consistent, and catalogs must be easy to build and maintain and must support widely used GIS datasets and formats. Catalog records must communicate clearly and be easy to understand. GIS catalog portals must also be built using widely adopted information technology (IT) and GIS technologies (e.g., database management system [DBMS], Web services, HTML, XML, and GIS formats).

Here are some useful criteria to think about when building a GIS catalog portal:

- Datasets should be consistently documented and discovered through standard searches from site to site. This requires that metadata documents be consistent from site to site. Notion: In addition to adhering to metadata content standards, catalog records (i.e., metadata) should be built using widely adopted library science principles.
- Portals should support searching against a single server, not many servers via distributed searches.
- Searches must be fast and efficient.
- Servers should control access levels to each catalog and its contents. Sensitive information must be private and secure. Accepted IT strategies must be employed to manage security, access levels of availability, and keep track of to whom each dataset is accessible.
- There should not be a big load (a “tax”) on servers that host a participating metadata catalog in an SDI. Searches should only hit the targeted server.
- Portals should allow you to view metadata records to determine if a candidate dataset is suitable for the intended use.
- Portals should allow you to access and view geographic data and Web services directly through a GIS catalog portal when available. Maps should be viewable in the user’s Web browser or with GIS software to display multiple data sources along with data from local sources.
- Search results should include up-to-date quality and summary information, which should be held in the catalog as standard metadata.
- Portals should be built using widely adopted information technology (e.g., DBMS, works through firewalls, supports Web service standards, and utilizes commercial off-the-shelf [COTS] GIS tools).
- Portals should always be available for reliable access and searching.
- Portals should support key IT and GIS standards (e.g., Web services frameworks, International Organization for Standardization [ISO]/FGDC metadata, most commonly used GIS file formats and Web application program interfaces).
- Portals should be practical to implement and maintain.
- Portals require more than publish and search tools. They must also provide exploration tools to learn more about candidate datasets. Users need to explore more details of many datasets to determine the suitability of the dataset for their specific use.
- Portals should support a range of methods for online data delivery (e.g., live data streaming, commonly used data formats, FTP download, and CD–ROM). Long-term data delivery should be through guaranteed highly available and high-performance data streaming services.
- Common data representations and essential content standards are important for collaborative efforts to assemble data from many federal, state, and local partners. Data content should employ generic types (e.g., OGC-compliant simple features) whenever practical. Data content must be multipurpose, designed to support most common GIS applications.
- Portals should support the ability to “Register for Notification” when new or updated data, maps, activities, references, and so forth, have been added to a favorite search area.

- Some central portals—often referred to as Clearinghouse Nodes—should enable selected users the ability to publish (register) map services, images, geographic datasets, geoservices, spatial solutions, geographic and land reference material, and geographic activities or events to share with others through submission of online forms. Data providers should have the ability to update their metadata submissions to reflect edits and updates over time.

Section 3: How to build a GIS catalog portal

GIS catalog portals require much more than a Web page and a metadata standard. They also require complete tools for a GIS and the best methods and practices currently available to apply these GIS tools.

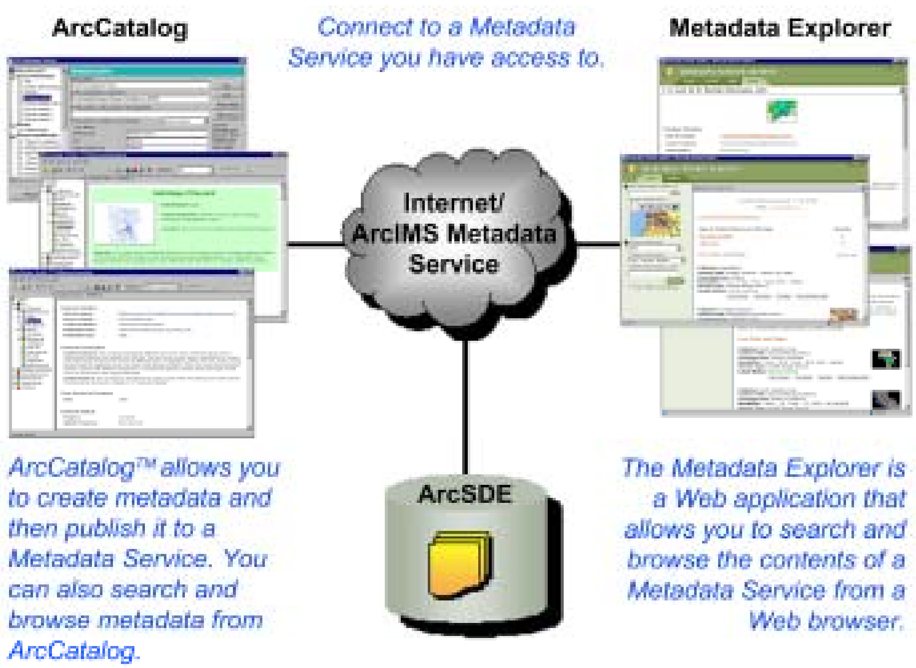

Figure 4. Overview of how ArcGIS is used to build, serve, and use a GIS catalog portal.

GIS users will perform the following tasks to build a GIS portal:

1. Create GIS data holdings and agree to make selected datasets available to other users.
2. Document these datasets with standards-based metadata.
3. Compile the individual metadata documents into a metadata catalog.
4. Publish the metadata catalog as a Web Service that users can search.
5. Build a custom gazetteer for geographies covered by the GIS database so that meaningful geographic searches can be performed against local place names.
6. Build a Search Tool, most typically a Web browser application for searching and browsing the metadata catalog.

Building a central GIS catalog portal

Many have a vision for central portals at state, local, and national levels. There is a common vision for one central GIS portal in each state and one in the federal government. Many believe that the central federal portal is the role of GeoSpatial One-Stop, which will be further explained in Section 5. Many expect that numerous large local governments (such as a regional council of governments [COG]) will also host central GIS portals.

Two common strategies for building central portals include:

- Perform *distributed* searches from a centralized Web site (e.g., the current FGDC Clearinghouse implementation); or
- Harvest catalogs into one central catalog database for focused search and access.

Distributed search

Through a series of pilot projects and prototypes, ESRI discovered numerous challenges with each approach but determined that the distributed search method was problematic. Some of the problems encountered with distributed search portals included:

- Performance hit by each GIS portal for every search through the distributed framework.
- Bad performance.
- Result sets were very difficult to interpret.
- Results were dependent on online available services. Many portals were frequently down or offline.

A recent study on the use of the FGDC Clearinghouse discovered that most users do not perform distributed searches at this site due to such problems. Instead, users went to specific GIS portals and browsed their Web site to find datasets relevant to their GIS projects.

Harvesting

The second strategy for central GIS catalog portals is one of harvesting. The concept is that a number of collaborating GIS sites would periodically share their catalogs with one another. One central site would harvest these catalogs and compile them into one comprehensive catalog that could be searched. GIS users would connect to and search these central GIS catalog portals. Since GIS users already share GIS datasets, harvesting appears to be a practical solution that will work (Figure 4).

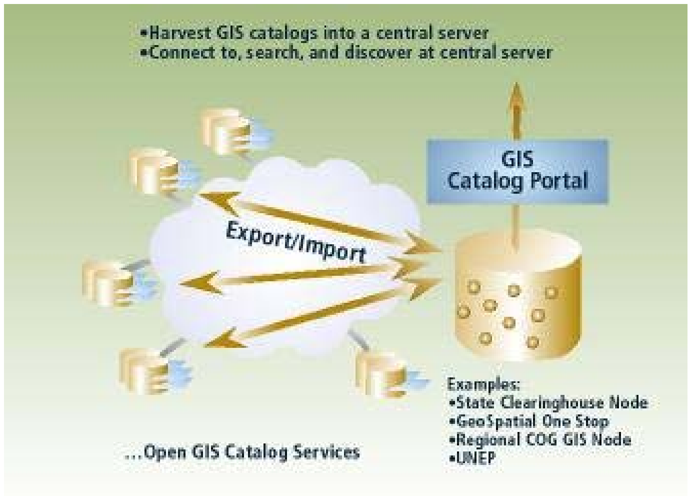

Figure 5. The concept of a central catalog portal where GIS users can search for and find GIS information relevant for their use. Periodically, the central portal site harvests catalogs from a collection of participating sites to publish one central GIS Catalog. The GIS Catalog can reference data holdings contained at its site as well as at other sites.

In cases where central portals are to be built, harvesting and sharing tasks include:

- Harvest catalogs of collaborating organizations to allow for a single centralized GIS portal.
- Export catalogs to include in other central GIS portals.
- Scheduled synchronization of shared metadata repositories (Figures 6 and 7).

This requires a formalized structure for multiorganizational collaboration.

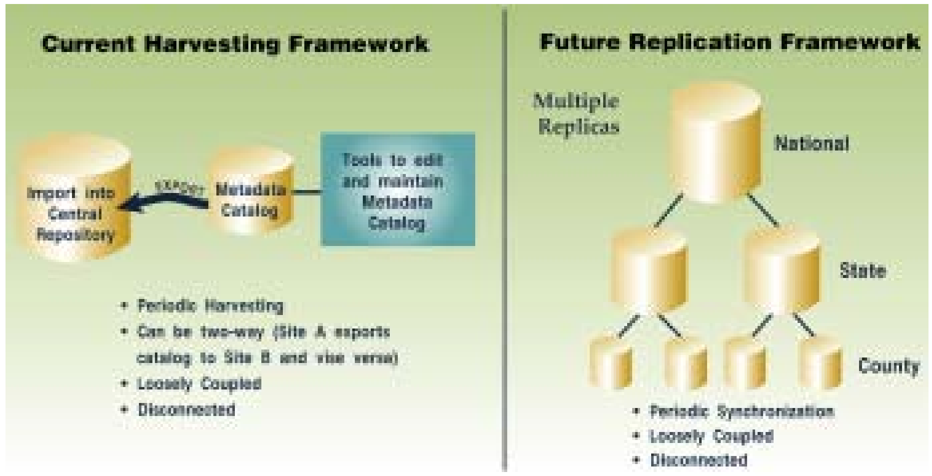

Figure 6. Harvesting frameworks for metadata catalogs will require high-end database sharing tools and methods. A major challenge will be to ensure that each metadata catalog is synchronized with the others. Current GIS tools support capabilities to import and export entire catalogs. This can be managed on a regular schedule to maintain synchronicity. Future methods will include loosely coupled replication, allowing for efficient, periodic change-only updates to be shared.

Figure 7. Each GIS catalog portal must support standard searches. Portals often contain support for various search mechanisms, including standard Web portal searches as well as Z39.50.

How to use a GIS catalog portal

Users need to find GIS data for use in their applications. They need to explore datasets and often perform further geoprocessing on the data to make it usable (and often simply just to certify that the dataset is appropriate for use in their applications).

Users perform the following tasks to discover available information at GIS catalog portals:

1. Connect to and search specific portals (e.g., connect to Delaware DataMIL).
2. Discover and investigate candidate data to find appropriate datasets for specific purposes. This may be as simple as reviewing the metadata record for a dataset or as involved as further geoprocessing on the data for validation. Advanced exploration typically occurs in concert with steps 3 and 4.
3. Connect to live data services (e.g., a GIS Web service), download data (e.g., from an FTP node), or contact the provider for the data.
4. Integrate and use the selected data in the application.

Section 4: How is ArcGIS used for SDI?

ArcGIS tools can be used for SDI in the following ways:

- Client access. ArcIMS® includes a range of client applications to connect to and use standards-based Metadata Services and GIS portals. ArcIMS clients used in a portal can include the Metadata Explorer, various HTML- and Java™-based Web-mapping clients, and other GIS desktops. Additional GIS desktops from ESRI include ArcReader™, ArcView®, ArcEditor™, ArcInfo™, MapObjects® for Java, and ArcPad®.
- Create, manage, and serve metadata. With ArcGIS, users can create and update FGDC- and ISO-compliant metadata (Figures 8 and 9). Using the standards-based ArcIMS Metadata Service and ArcSDE®, users can manage and serve metadata catalogs on a local network, a secure network, or the World Wide Web. The ArcIMS Metadata Service can be accessed through standards-based ArcIMS services and through Open GIS protocols, such as Z39.50.
- Create, manage, and serve GIS information. GIS users build and manage their geographic data and information using ArcGIS and ArcSDE. They can openly serve data, metadata, online maps, and other information using ArcIMS. ArcIMS services can be accessed with a wide range of clients via GIS and XML-based Web services standards.

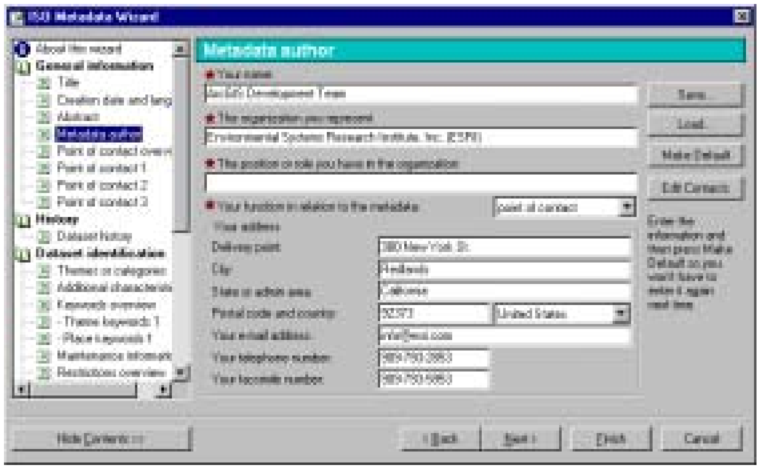

Figure 8. A complete metadata document can be created for datasets using the ArcCatalog interview-style metadata editor. This helps users create and maintain ISO-compliant metadata documentation. Required items are tagged with a red star to help you complete the necessary fields.

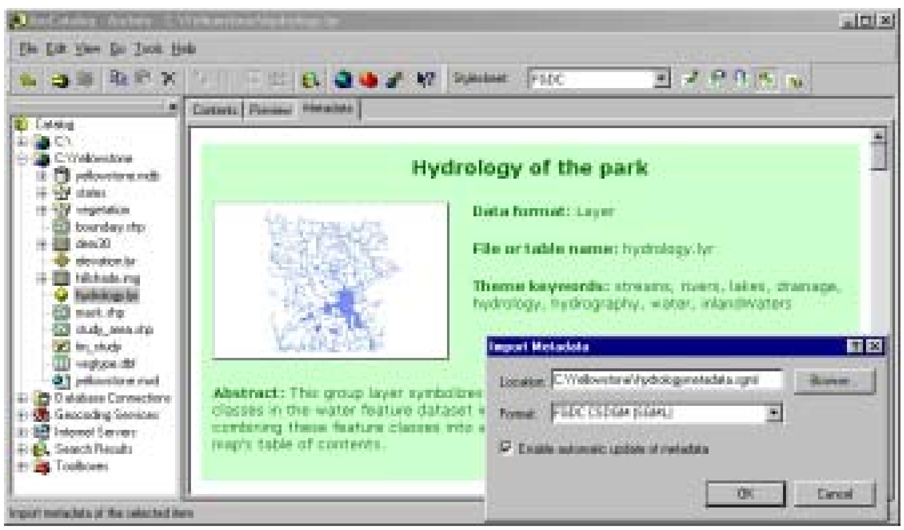

Figure 9. ArcCatalog automates much of the work in documenting data. Simply point ArcCatalog at the data to automatically populate its inherent metadata properties (e.g., coordinate system, feature class, and attribute field names).

While a broad collection of new GIS capabilities is required to meet the specialized needs for information cataloging and searching, all the key tools used for building a complete GIS are also necessary for building and maintaining a catalog portal. ESRI provides full support for these in its ArcGIS product line:

- Practical database designs implemented with full-function GIS
  - A formal, standards-compliant data model for GIS data. The geodatabase is an object relational data model that manages information in any number of DBMS

architectures and adheres to the OGC and ISO simple features specification for multipurpose GIS data.
  - A formal, standards-compliant design for metadata documentation that supports discovery methods that, ideally, are consistent from site to site. Searches must be more like a library search than a World Wide Web search (i.e., GIS cataloging methods should be based on widely adopted library science principles and methods).
  - A metadata catalog database for all published data holdings.
  - Common content standards. ESRI publishes essential data models used in practical real-world GIS implementations. See http://support.esri.com/datamodels for more details.

- ArcEditor and ArcInfo. These advanced GIS seats include database construction tools used to assemble and maintain multipurpose, generic GIS datasets.
  - Comprehensive editing
  - Data conversion
  - Data automation
  - Geoprocessing and validation

- Geoprocessing tools. A comprehensive collection of geoprocessing tools must be available that work on all GIS data types to support format conversions, topology construction, coordinate management, tabular and attribute manipulation, generalization, GIS scripts for repeatable validation and quality assessment procedures, analysis to derive new information, and many data manipulation tasks.

- Efficient metadata documentation tools. The ArcCatalog application that comes with ArcView, ArcEditor, and ArcInfo has a full suite of documentation (metadata) tools.

  ArcCatalog is a standards-based tool for documenting each dataset. It is highly efficient and automates much of the work by sensing all the properties inherent in a dataset. It will automatically populate these metadata fields for you. ArcCatalog also supports full import, export, and metadata format conversion. Users can start with standard metadata abstracts so that many organizations catalog their datasets consistently.

- Tools for assembling a metadata catalog, including tools to build, manage, update, and replicate the GIS metadata catalog. Three technologies—ArcGIS, ArcIMS, and ArcSDE—support these functions. ArcCatalog, along with ArcSDE, provides support for building a metadata catalog in any commercial DBMS. ArcIMS provides a Metadata Catalog Service to orchestrate this effort among collaborating organizations.

These tools help you easily build a metadata catalog. It's much like building a directory of files and folders. This also includes the following:

- Tools to build, maintain, and update the metadata catalog database.
- Tools to manage catalogs, associate documents with one another, organize catalogs for easier browsing and searching, and so forth.
- Tools to harvest and share metadata catalogs.
- Tools to replicate catalogs (coming in ArcGIS 9).

■ Gazetteer tools
- Tools to build and maintain a custom Gazetteer database. You can add place names and locations from any feature class for your local geographies. Your gazetteer can be fully integrated with the geographic features in your GIS database. For example, your Gazetteer tables can be managed as a comprehensive names database that is fully integrated with all of the feature classes in your database.
- Gazetteer Search tools. For example, find a place name and see an ordered list of matches.

■ The ArcIMS Metadata Service. Web tools for publishing and searching the metadata catalog.
- Tools to publish GIS catalogs as a set of search and browse services.
- Tools to serve data and map services for published datasets (optional).
- Web site management tools to maintain high availability to both the catalog and data services.
- Tools that support all modern catalog services standards.

■ Metadata Explorer. Web-based search and discovery tools.
- Metadata Explorer is an HTML, browser-based application that includes a set of discovery and search tools that enable users to connect to and search and browse the GIS catalog. This tool can also be used to search any Z39.50-compliant portal.

■ Data dissemination tools in ArcIMS.
- Tools to directly connect to and use GIS data.
- Tools to download data in any number of common GIS formats (e.g., clip-zip-ship or FTP nodes).

■ Database management tools to allow transactions and updates against all information sets, including metadata documentation.

Section 5: Key metadata concepts and methods

The role of metadata is the key to the SDI vision. Thus, it is critical to recognize the appropriate role of metadata within this context. It is important to simplify the metadata discussion to fundamental goals about data sharing.

Metadata is documentation about a GIS dataset and is used to:

■ Support catalog searches.

- Provide simple high-level descriptions.
- Provide details to determine appropriate uses for a dataset—especially when the user cannot directly connect to and browse the dataset to infer those details.

A number of metadata content standards exist. Each provides an overwhelming list of required and optional fields, which tends to cloud people's perceptions of the work involved. As presented in previous sections, COTS software tools exist that simplify the work involved.

Strategies for cataloging GIS data

Numerous metadata content standards have been adopted worldwide, including FGDC, ISO, and Comité Européen de Normalisation. Most have a huge number of properties that can be used to document each GIS dataset. What are all of these fields, and why are they there? There are four primary roles for this content:

- Fields that help to describe the GIS dataset in much the same way that a card catalog record describes a library holding. What is this thing, what is it about, etc. These are the fields that enable consistent searches.
- Fields that describe some of the dataset properties that are inherent in the data (its map projection, its feature presentation, its attribute fields and types, how many features, etc.). Recall that these can be automatically sensed and recorded for each dataset using ArcCatalog.
- User-defined information that expresses accuracy, use, relevant scales, data sources, currency, contact information, and so forth, of each dataset. These are often very critical details that help other users determine the suitability of each dataset for their use.
- Consistent catalog and search information, such as keywords (based on library science methods and practices).

While each metadata record seemingly contains an unmanageable amount of information on each of these four topics, the task of documenting each dataset according to standards needn't be complex or manually intensive. The following table presents some strategies for compiling consistent, searchable metadata documentation across a wide range of GIS databases.

Metadata Contents	Available Tools	Responsibility	Strategy
Standard content descriptions and catalog properties, common keyword lists	ArcGIS data models will contain examples of shared metadata templates for selected data models, including keyword lists.	For the layers in standard data models, state and federal librarians working with data model authors could write one consistent set of records that can be shared.	Write one description for content (e.g., national map layers, census layers, etc.).[1] Share this as a MARC-like abstract as part of a common data model. Users can import standard descriptions into their metadata documents.
Properties inherent in each GIS dataset	ArcCatalog	Dataset author simply points ArcCatalog at the dataset to capture as well as to update and maintain this information.	Use automated tools to capture and maintain this information. ArcCatalog is a great productivity tool.
Use, currency, data sources, contact information	ArcCatalog	Dataset author	Follow content standard guidelines to record these properties.

[1] This is similar in concept to the Machine Readable Cataloging (MARC) record for sharing digital catalog records in the library community. Book publishers write a MARC record for each new publication that libraries download into their individual catalogs, ensuring a consistent catalog record for each publication.

Cataloging of GIS datasets is different than cataloging books in a library. MARC records are completely self-contained card catalog records for library holdings, while each GIS dataset has some properties unique to each instance. GIS datasets differ from site to site (e.g., in their area of coverage, attributes). Yet each dataset that is based on a common data model will share a number of common properties with other datasets.

The common data model should include a common metadata abstract and template for that data model as the starting point to document each instance of a dataset. Library scientists could help GIS users prepare useful abstracts that will support more consistent and reliable searches if adopted by the GIS industry.

Section 6: A possible vision for building a national GIS network

Most state and federal GIS organizations are interested in Web dissemination of GIS information using portal sites. Each major government agency responsible for GIS data content should provide access to this information and in many cases is required to do so. Further, there is a shared vision that each state and federal agency will host sites that act as GIS repositories for sharing and maintaining key geographic datasets for access by other users.

Each repository would house GIS information as well as a metadata catalog that describes the data holdings available at the repository. Optionally, each catalog could also reference data holdings on external servers. In the 1990s, such portal sites were often referred to as Clearinghouse Nodes.

There is broad acceptance and excitement about this vision, and it is being widely adopted. For example, this concept forms the basis of the USGS National Map; the FGDC's Geospatial One-Stop; and the U.S. Bureau of Land Management's Geo Communicator Web portal. Most states are developing similar portals, such as the Delaware DataMIL, the State of Texas Strat Map, and the State of Utah's State-wide Geographic Information Database. Many other GIS portals exist or are in development across the country and around the globe. There is no doubt that numerous portals will come online in the coming months that allow easy access to their GIS data holdings.

Yet if the GIS community does not quickly come to terms with common metadata descriptions for each dataset instance, GIS catalog searches will be inconsistent across nodes.

What organizations should participate in assembling the National Spatial Data Infrastructure for the United States?

The U.S. FGDC has a concept called the GeoSpatial One-Stop for organizing information access for the key federal agencies that are responsible for various layers of GeoSpatial information.

Each organization that creates and shares key GIS datasets with other users should participate. This includes strategic sites in each state as well as federal partners. Each of the 50 states that comprise the United States as well as key federal agencies who own the responsibility for key thematic layers of GeoSpatial information should participate in building GeoSpatial One-Stop portals for the country.

The NSDI vision for the United States is to interconnect the constellation of state and federal Web Portals as well as the set of focused sites referenced within each catalog portal. Many states are moving ahead on their plans for GIS catalog portals. In the federal government, a critical part of the NSDI vision is the deployment of a federal GeoSpatial One-Stop—one comprehensive search portal for federal geographic data.

Why will this idea work? The United States works very well for this concept of distributed responsibilities. The 50 states and six territories are a manageable collection of government bodies. Each state is well known and easy to remember. Many governing functions are the responsibility of—that is, distributed to—each state. Most fully manage their GIS data stewardship responsibilities and coordinate GIS activities with their neighboring states, local governments, and collaborating federal agencies. In many, if not most, states, federal GIS organizations have begun to define their main connect point to find GeoSpatial information for the state (e.g., the Utah Automated Geographic Reference Center, Delaware DataMIL, Texas Geography Network).

Since the name of each state is known, every GIS user can readily identify each state and thus should be able to visit the state's main GIS catalog portal. From this portal, it should be very easy to find data as well as to jump to other key portals within the state, such as a regional COG or a municipal GIS portal. Perhaps we should promote the use of common URL conventions to be used by GIS "clearinghouse" catalog portals, such as www.nsdi.tx.gov for Texas and www.nsdi.ca.gov for California.

GeoSpatial One-Stop

The federal government has a similar vision for a common one-stop portal. Just as a user could visit any state site as a jumping off point, GeoSpatial One-Stop can be the jump point for all federal GIS datasets.

This collective vision means that there would be 50 state portals and one main federal portal as initial access points for all government GIS datasets in the United States. The federal one-stop portal should also provide access to the major database stewards in federal agencies, such as the U.S. Bureau of Land Management's portal (http://www.geocommunicator.gov/).

There is broad support for the GeoSpatial One-Stop and the GIS community insists that this portal is crucial to build. Most professionals recognize that this can be implemented today using available commercial GIS tools.

The current technical vision for the GeoSpatial One-Stop has been focused on interoperability research for future use and by theoretical debates on what's needed. The United States should focus on real implementations using today's widely adopted GIS technology and evolve its implementation over time based on practical implementation experience.

It's important to recognize that this work is achievable now; it has already been implemented in numerous organizations at all government levels. Building the GeoSpatial One-Stop should use best-known practices and methods. Implementation would also lead to better knowledge of the appropriate practices and methods and a practical vision for its future evolution.

Potential collaboration with librarians

While the FGDC and ISO metadata content standards are a great start, they did not go far enough in establishing a uniform approach to information cataloging. Library science professionals need to be used to establish uniform methods for documenting metadata content so it is searchable and consistent across all implementations. Otherwise, searches at GIS catalog portals will be more like Web searches and less like library catalog searches. The NSDI can provide a vision for federal and state agencies to collaborate with librarians on the development of these methods.

- A framework to work with librarians (e.g., in each state and at the federal level) to help the GIS profession develop a common methodology for metadata documentation and indexing.
- Common metadata records for key GIS datasets—the GIS MARC records. These records can be a standard part of the common data model efforts under way. They could be imported into the metadata document for each GIS dataset at user sites. This would support more consistent searching across all GIS catalog portals.

Additional topics

Below is a list of topics that are worthy of further exploration. They are outside the scope of this document but are also important topics for further work. Most could be debatable in a theoretical setting but are perhaps best answered by performing real implementations.

- What is the role of common content standards (i.e., standard data models) in NSDI? What comprises a common data model? Can a common data model that works across GIS systems be defined?

- What role do standard schemas play? What are the data model schemas, and how are they shared?

- What are some key goals for data content standards?
  - Build data using multipurpose designs.
  - Follow common content standards, best/accepted practices.

- Why is the use of COTS software superior to exotic one-of-a-kind implementations?

- What should the methodology be to build consistent searchable metadata catalogs? And how can library scientists help the GIS community?

- What are the data sharing mechanisms that will be used to download and disseminate GIS datasets?

- How should states organize their work?

- How should federal agencies organize their work?