IMPORTANCE OF OPEN SPATIAL DATA INFRASTRUCTURE FOR DATA SHARING

E.Bank

ISLEM GIS 13.cad. No 14 06530 Beysukent Ankara ebank@islem.com.tr

Commission TS-PS-WG IV/4

KEY WORDS: GIS, Infrastructure, Spatial Infrastructures, Interoperability, Standarts, Open Systems, Metadata, Inter-Operability

ABSTRACT:

An exciting trend and vision in Geographic Information System (GIS) is the development of National, Continental, and Global Spatial Data Infrastructures (SDI). Assembling an SDI portal involves three key parts:

1. The Metadata Registry and Search Portal
2. Dynamic mapping and data fusion on the Internet
3. Data modeling and data download to acquire data in a useable and well-documented format for use.

By necessity, these are quickly becoming a key part of many GIS implementations.

This presentation will describe a practical approach to building specifications, standards, and reference data models for SDI programs.The presentation will also presents ESRI's GIS software platform for open SDI solutions that can enable local, regional, national, and pan-European data sharing.

The presentation will include;

- Spatial Data Infrastructure
- Information Modeling for Spatial Data Infrastructure
- Building National and Continental Data Models for SDI
- Best Practices for Implementing SDI
- Using ESRI software in Collaboration with Other GIS's
- The role of standards for SDI and key content and metadata specifications

1. Spatial Data Infrastructure

a. What is GIS technology ?

GIS technology is making a difference. GIS is more than software or application, it is a world of digital data, online portals, web services and location services that enable access to information when and where needed. As an enabling technology GIS is making a significant difference tool of good governing, better buiness and a healty planet (ESRI. 2004). GIS is an Information System for creating, maintaining, managing and using Geographic knowledge. GIS is a complete platform for working (editing, data management, mapping, spatial analysis and visualization) with Geographic Information. GIS integrates many types of disciplines/organizations like; *social factors, roads/infrastructure, land use/land cover, imagery, environment, base maps,* activities like; *georeferencing, digital processing, map overlay, spatial analysis, visualization*, and data like; vector and raster. GIS was a tool for individual projects and departments in the past but in recent years GIS has moved very fastly from being a tool to become the framework for sharing information among organizations and across society. GIS illustrates relationships, connections and patterns in data. The unique integration capabilities of GIS bring together data from many sources for visual display and analysis, creating a more complete picture of a situation than would otherwise be possible. This enables organizations to make better decisions based on all relevant factors. For this reason GIS technology has become an integral part of the information infrastructure in many organizations. GIS has long been recognized as an integrating technology that can bring highly disparate data into a common map display. Overlaying layers of different data in a common geographic space allows you to integrate and derive new information to help you solve problems and to work more efficiently and effectively.

b. What is SDI technology ?

Spatial data is the locations and shapes of geographic features with descriptions of each feature and differs from normal data with it's volume and structure. Spatial datasets are interrelated and often very large. Unlike business queries that may only return a few records, spatial queries for generating maps are routinely run against many thousands of features. Spatial data usualy comes from across an organization and increasingly from outside sources. Because spatial datasets are fundamentally distributed and dynamic, standarts and interoperability have always been a fundamental part of spatial data. There is widespread recognition that the data layers and tables in most GIS implementations come from multiple organizations. An organization typically develops some, but not all, of its own spatial data content. At least some of the layers will be from external sources. Thus GIS data management, by its very nature, is distributed among many users (ESRI. March 2003). GIS requires a distributed information system concept to manage and share spatial data. Because GIS users are hungry for quality geographic information, there is a fundamental need for users to share their data. Today, thousands of organizations worldwide invest billions of dollars annually automating and integrating map information for their focused GIS projects. Their need for up-to-date geographic information drives these efforts. SDI has become an essential tool when information integration is critical. It works with geographic information from many sources to support a broad range of applications. SDI can repurpose information for new applications that go beyond the data's original intent. For example, information compiled for land records management can be used for environmental applications, utilities, emergency response, homeland security, and many other applications. Spatial data management includes not only data collection but also data dissemination. Interoperability enables data sharing between organizations and across applications and industries, which results in the generation and sharing of more useful information.

2. Information Modeling for Spatial Data Infrastructure

A Spatial Data Infrastructure interconnects GIS nodes across the World Wide Web to promote information sharing and access. The SDI concept binds many GIS nodes into loosely coupled information networks. There are three functional node in information model of SDI.

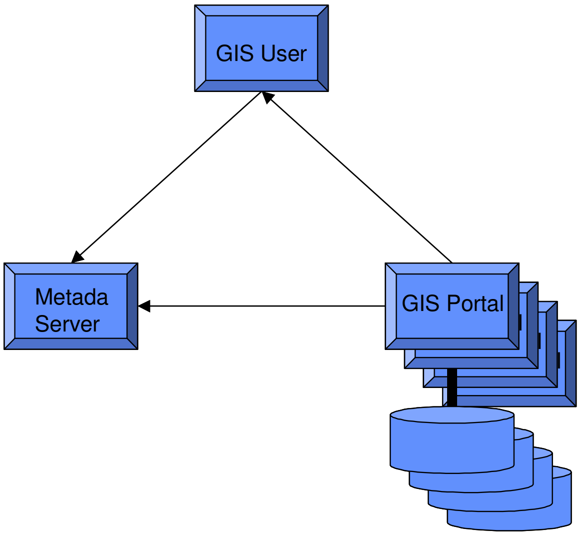


Figure_1 : Building Blocks of SDI

GIS Users--A wide range of users who connect to the catalogs, search for useful geographic information, and then connect to GIS portals that provide the information access. GIS Users can find the data they need, quickly review and display its contents, and read or create metadata. GIS Users can also manage spatial data stored in folders on local disks or in relational databases that are available on their network.

Metadata Catalogs--A series of well-organized catalogs that describe and reference geographic information sets at each node. Metadata is information about GIS 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.

GIS Portal--A GIS portal that provides access to FTP download sites, maps, reports, Internet GIS services, and so on. GIS Portal publishes GIS Data & Services & Documents. A broad range of GIS users access metadata and services with any client (wireless, html, java,pocket PC, windows).

In practice, each node in an SDI would be built to provide one or more of these three capabilities. For example, a city might provide a catalog referencing the available data and services at its Web site along with a series of data and mapping services. A small GIS shop would be able to access and use those services in their Software seat across the Internet. In another case, a state government might provide the metadata catalog describing its services as well as data and services at other GIS portals within the state. The state might also publish GIS Web services with a set of HTML clients to access various Internet map and data services. At its most basic level, a Spatial Data Infrastructure is realized through a catalog holding metadata about available data and services. As the number of participating organizations and users grow, so will the numbers of entries and complexity of the metadata catalog. The role of a metadata server and search tools becomes more critical. Metadata server makes accessing and managing geographic data simple.

3. Building National and Global Data Models for SDI

GIS professionals are learning about the benefits of a Spatial Data Infrastructure (SDI). The SDI, developed by national, state, and local governments, promotes the vision of a framework for GIS users to openly share geographic information with one another. The SDI is meant to address the needs for users to interconnect their existing GIS nodes across the Internet (and in many cases over secure networks) in order to share information with one another openly (i.e., based on standards). Many people are familiar with the term NSDI, or National Spatial Data Infrastructure, a concept defined as the technologies, policies, and people necessary to promote the sharing of geospatial data throughout all levels of government, the private and nonprofit sectors, and the academic community. Recent discussions have revolved around the GSDI, or Global Spatial Data Infrastructure, which describes a framework for data sharing at the global level. Many GIS users want to provide information portals to publish a metadata catalog for their collection of GIS data and services. A few clearinghouse sites want to provide a catalog of information sets published by many users. Sometimes the data and services will be published for a user's own GIS node, while at other times the catalog will include data and services at many other GIS sites. Access to such a clearinghouse node could be open on the World Wide Web, while at other times it can be restricted and secure for a select audience. 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 (ESRI. March 2003). 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.

Catalog portals require a complete information system for geography, tools for data import and export, tools to validate data, tools to serve catalogs and data, search and discovery tools, tools to explore and use data, and tools to build and manage GIS data. A GIS metadata catalog plays the same role as a card catalog in a library (ESRI. March 2003). 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. Usually GIS experts intend more detailed metdata and General Public intend minimal metadata.

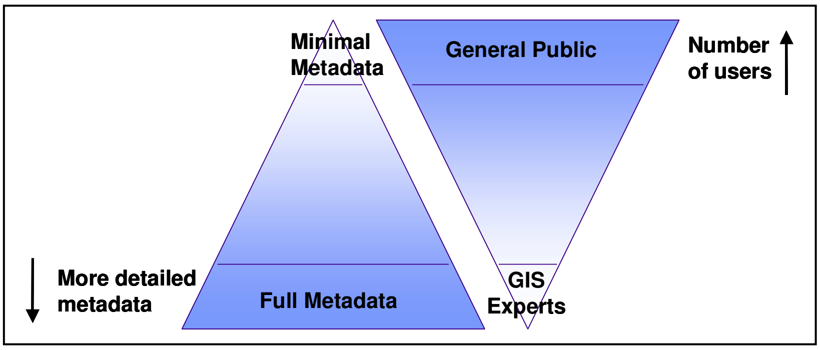
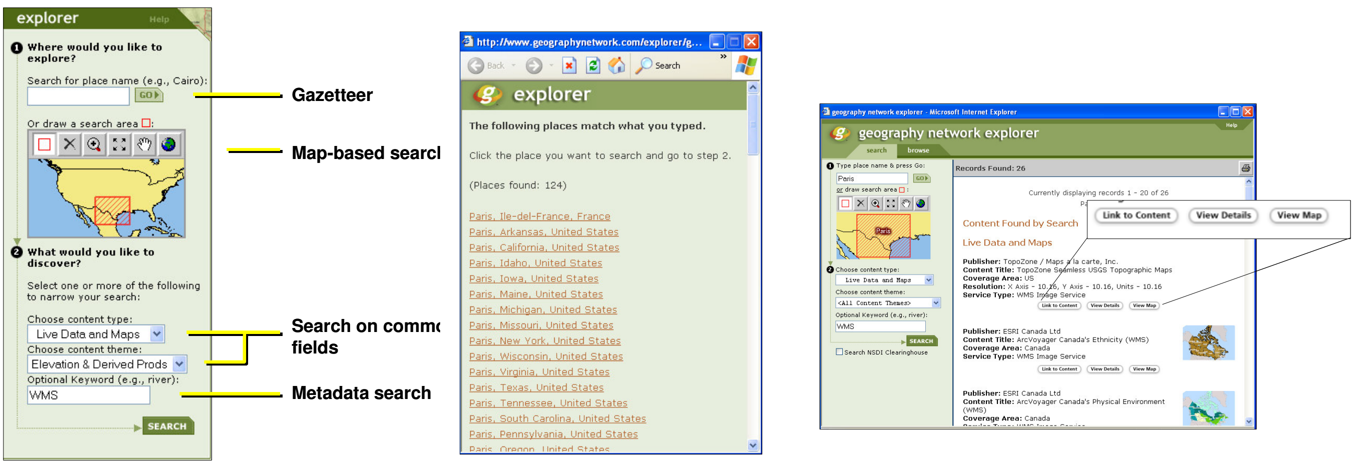


Figure_2 : Metadata Completeness (Gerco Hoogeweg.2004) .

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 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. In addition, 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.



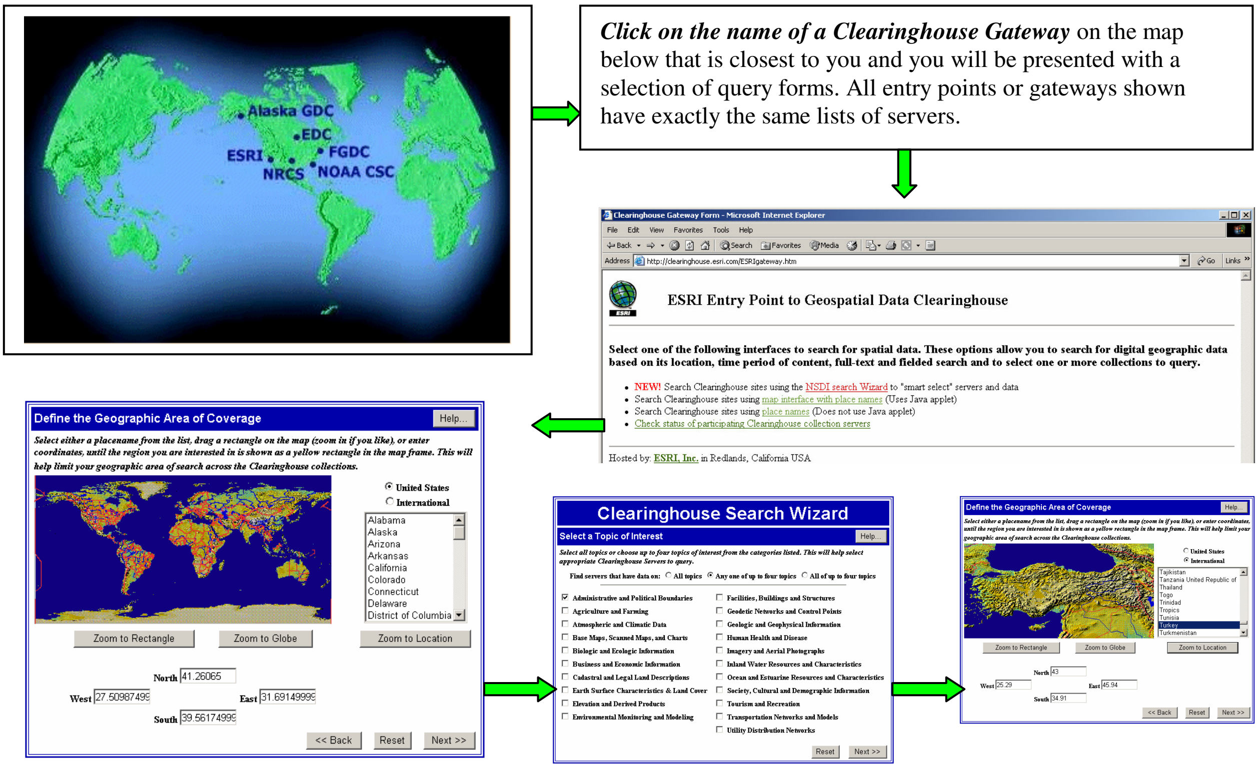
(a) (b) (c)

Figure_3 : GIS Catalaog (a): Catalog search, (b): Gazetteer, (c): Search result

4. Best Practices for Implementing SDI

There are already many examples of catalog portals—many of which are GIS portals—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 (ESRI. March 2003). **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.

a. ***The Geospatial Data Clearinghouse*** is a collection of over 250 spatial data servers, that have digital geographic data primarily for use in Geographic Information Systems (GIS), image processing systems, and other modelling software. These data collections can be searched through a single interface based on their descriptions, or "metadata."



b. 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.

c. 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)

d. 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

e. 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.

5. Using ESRI software in Collaboration with Other GIS's

In 2001, ESRI began to discuss g.net—a new architecture for sharing and using geographic information system (GIS) information from distributed sources. g.net is an architectural vision for how ESRI users build parts of and participate in a Spatial Data Infrastructure. g.net describes how ArcGIS, ArcIMS, and other ESRI software technology are used. ArcGIS, ArcIMS, ArcSDE, and other ESRI software can be used to implement all of the functional nodes in a Spatial Data Infrastructure. This architecture is known as g.net. ArcGIS tools can be used for SDI in the following ways (ESRI. March 2003):

Client access - Any GIS user wishing to access information and services remotely can connect to and use metadata servers and GIS portals.ArcIMS includes a range of client applications to connect to and use standards-based Metadata Servers and GIS portals. ArcIMS clients can include the Metadata Explorer, various HTML- and Java™-based Web-mapping clients, and other GIS desktops. These 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. Using the standards-based ArcIMS Metadata Server and ArcSDE®, users can manage and serve metadata catalogs on a local network, a secure network, or the World Wide Web. The ArcIMS Metadata Server 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.

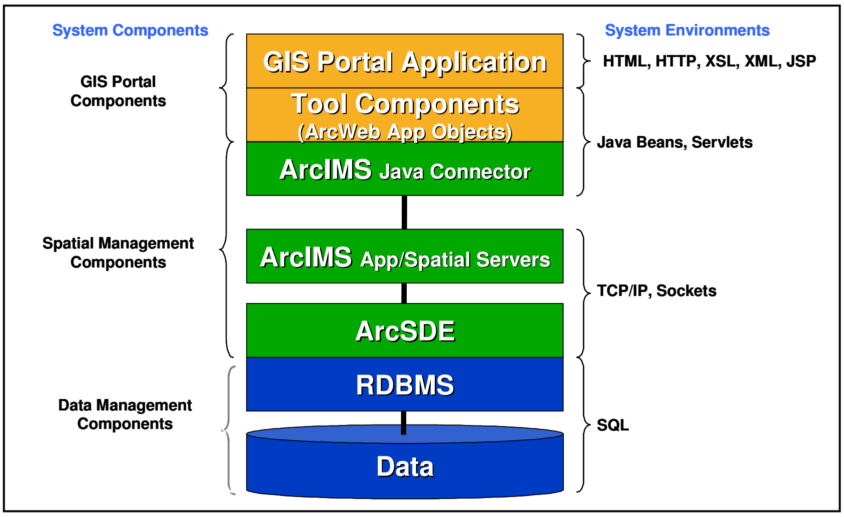


Figure_4: ESRI GIS Portal Architecture (Gerco Hoogeweg.2004)

6. The role of standards for SDI and key content and metadata specifications

Until the mid-'90s, organizations purchased geographic information systems that closely tied applications to a native, proprietary spatial data model. These early nonrelational file structures were highly optimized for fast access to data and, being file based, were relatively easy to distribute between sites using the same GIS vendor software. However, the ability to share data among users within an organization was limited by network protocols such as network file system (NFS). Data sharing between organizations with different GIS vendor systems was limited to data converters, transfer standards, and later open file formats. Sharing spatial data with other core business applications was rarely achieved. Gradually, GIS models evolved into georelational structures where related attribute data could be stored in a relational database that was linked to the file-based spatial features. However, the georelational format had limited scalability, and the dual data structure (spatial features stored in proprietary file-based format with attributes stored in a relational database) meant that the GIS could not take full advantage of relational database features such as backup and recovery, replication, and fail-over. In addition, supporting large data layers required the use of complex tiling structures to maintain performance, and sharing spatial information with other core business applications was still not possible.

In the mid-'90s, new technology emerged that enabled spatial data to be stored in relational databases (often referred to as spatially enabling the database), opening a new era of broad scalability and the support of large, nontiled, continuous data layers. When
the new spatially enabled databases were combined with client development environments that could be embedded within core business applications, the sharing of spatial features with core business applications, such as customer management systems, became possible. In addition, these spatially enabled databases allowed organizations to take the first steps toward enterprise GIS and the elimination of

organizational "spatial data islands." Perhaps not coincidently, the open GIS movement was spawned shortly after the arrival of the first all-relational models capable of storing both spatial and attribute data in a relational database when standards organizations, such as the Open GIS Consortium (OGC), the International Organization for Standardization, and the U.S. Federal Geographic Data Committee, began promoting the idea of data sharing through spatial data standards. The early work of these organizations was focused on sharing simple spatial features in a relational database, thereby enabling interoperability between the commercial GIS vendors. OGC, an international industry consortium of private companies, government agencies, and universities, published an open spatial standard called the Simple Features Specification. To fully realize the capability and benefits of geographic information and GIS technology, spatial data needs to be shared and systems need to be interoperable. GIS technology provides the framework for a shared spatial data infrastructure and a distributed architecture. It is very important that a GIS product has developed its products based on open standards to ensure a high level of interoperability across platforms, databases, development languages, and applications. The GIS community in general has been pursuing open standards and interoperability for many years. Approachs listed below play a significant role in GIS interoperability (ESRI. Summer 2003).

- Data Converters (DLG, TIGER, MOSS, GIRAS, IGDS)
- Standart Interchange Formats (SDTS, DXF,GML)
- Open File Formats (VPF, Shapefiles,DGN)
- Direct read application programming interface (ArcSDE API, CAD reader, ArcSDE CAD Client)
- Common feature in a DBMS (OGC Simple Feature Specification for SQL)
- Integration of standardized GIS Web services (OGC WMS (Web Map Server), OGC WFS (Web Feature Server))

7. Conclusion

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 (ESRI. March 2003). 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. In early years, the constraints of computational speed and cost caused GIS vendors to focus on practical solutions such as direct file conversion. Data sharing between organizations with different GIS vendor systems was limited to data converters, transfer standards, and later open file formats. Sharing spatial data with other core business applications was rarely achieved. Today, most GIS products directly read and sometimes dynamically transform data with minimal time delay. The point here is that the GIS community has been pursuing open interoperability for many years, and the solutions to achieving this goal have changed with the development of new technologies. In the early days of GIS, the focus, with rare exceptions, was on individual, isolated projects. Today the focus is on the integration of spatial data and analysis in the mission-critical business processes and work flows of the enterprise and on increasing the return on investment (ROI) in GIS technology and databases by improving interoperability, decision making, and service delivery. A GIS must produce useful information products that can be shared among multiple users, while at the same time provide a consistent infrastructure to ensure data integrity. It is important not to get caught up in the technology and forget this basic principle. Interoperability enables the integration of data between organizations and across applications and industries, resulting in the generation and sharing of more useful information.

8. References

ESRI. Summer 2003 Standarts and Interoperability
ESRI. 2004 International User Conference
ESRI. March 2003 Implementing a metadata catalog portal in a GIS network
Gerco Hoogeweg.2004 ESRI GIS Portal Toolkit presentation

9. Author Contact Adress

Dr. Emin BANK
Surveying Engineer, Ms.Sc.
ISLEM Geographic Information Systems&Engineering Co.Ltd.
13. Cad No: 14, Beysukent
06530 ANKARA, TURKEY
Tel: +90 312 235 64 90
Fax: +90 312 235 56 82
Mail: ebank@islem.com.tr