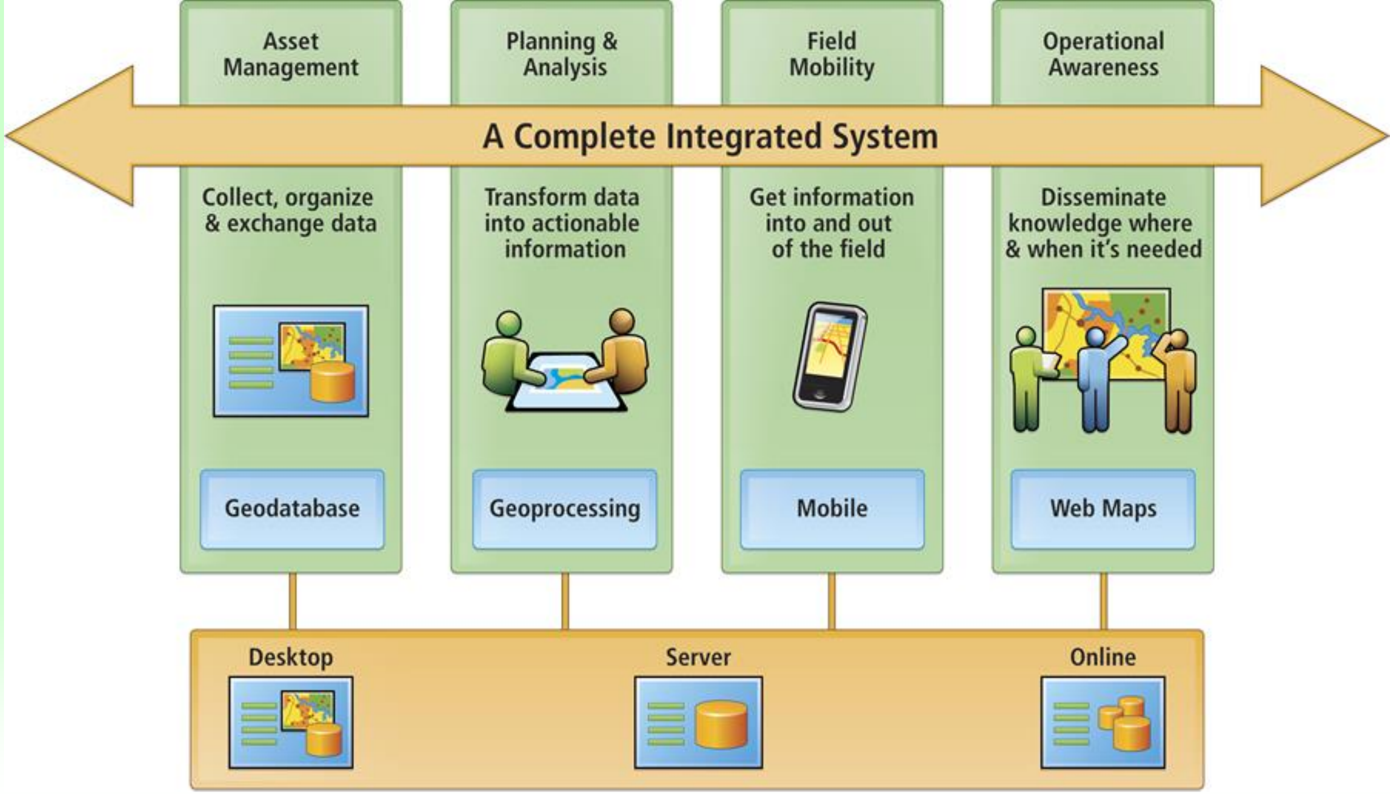


Enterprise GIS Framework

GEOG 482/582 : GIS Data Management

Lesson 10: Enterprise GIS Data Management Strategies

Overview

Learning Objective Questions:

1. What are challenges for multi-user database environments?
2. What is Enterprise GIS?
3. What are the major issues with multi-user data sharing?
4. What are approaches to geospatial data sharing?
5. What are three approaches to enterprise data management and how can we characterize them?
6. What is data warehousing?
7. What is a federated database system?

Lesson Preview

Learning objective questions act as the lesson outline.

Questions beg answers.

1. What are challenges for the multi-user database environments?

Multiple representations of data – one of the most complex challenges in a database environment; consider scale, projection, geodetic datum, data format, and symbology for a map; extend that perspective to many maps that emerge out of a personal geodatabase; then extend that idea to many databases supported by a organization all of which emerge from an integrated database environment.

Common user interface – provide same look and feel and reduce cost of maintenance; but how many different levels of responsibility are there in an organization, each that could likely use a different user interface because their need for information varies by responsibility to use information.

Concurrent access and security – database integrity and address data inconsistency among multiple users

Standards – hardware, software, communications protocols are needed to create a harmonious environment

Metadata support – needed to determine fitness for use

Systems Architecture – Use Warehouse or Federated Database Approach

Multi-user GIS data management strategies

Three contexts for multi-user data management…

- Workgroup data management– file geodatabase
  - Department unit GIS
- Enterprise data management – SDE geodatabase
  - Organization-wide GIS
- Consortium data management – SDE and/or spatial data infrastructure
  - Multi-organization GIS, each with their own enterprise approach

Focus of Enterprise DBMS (focal scale as a meso-scale, Workgroup (scale below as a micro-scale) and Consortium (scale above as a macro-scale) described in relation to Enterprise

Key terms
Workgroup
Enterprise
Consortium

2. What is Enterprise GIS?

Enterprise GIS - an environment combining data, software, hardware, procedures and people for delivering organization-wide geospatial capabilities while improving access to geographic information and extending geospatial capabilities to diverse users of GIS. Cooperation among workgroup (Dept unit) GIS information needs and capabilities, that could lead to coordination of those needs and capabilties.

Advantages of deploying an enterprise GIS include:

- **Using a common infrastructure** for building and deploying GIS solutions
- **Extending geospatial capabilities** to an enterprise community
- **Improving capabilities of other non-GIS enterprise systems** by leveraging the value of geographic information (remember the foundation of a space-time perspective)
- **Increasing overall operating efficiency** using GIS across an organization

Goal for enterprise GIS

Important to have a clear vision for "Why Enterprise" approach

Goal should be agreed to by all (many units) across organization

Goal: getting capabilities out to the organization users so that geographic information efficiency, effectiveness, and equity (knowledge perspectives) can flourish.

Sometimes information custodians will be hesitant due to fear of data misuse or abuse

Those issues need to be addressed in light of the goal

Some questions about the need for enterprise GIS

- Is there a need for a coordinated repository for enterprise data?
- Is there a central unit available to the enterprise for serving data and data storage?
- Is the system, both network and hardware, adequate to support enterprise GIS data traffic?
- Is the current software sufficiently robust for the enterprise?

What is an Enterprise GIS Framework?

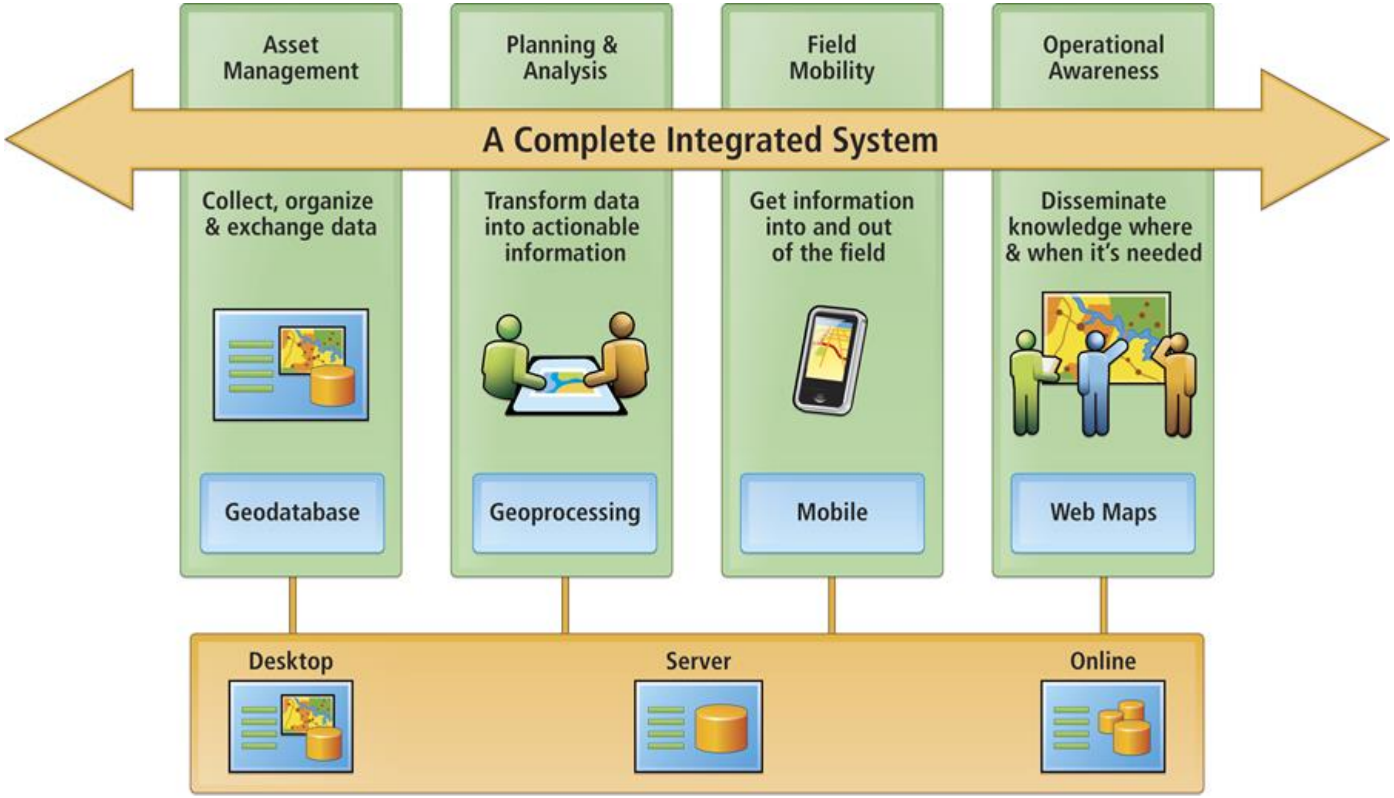
An **enterprise GIS framework** improves organization workflows since it applies a 'geographic approach' to relate legacy and new information for improving decision making; greater efficiency with money, time, and resources; and more effective communication. (see graphic next slide)

Framework has the following characteristics…

- Scalable, extensible, reliable, and secure
- Open, interoperable, and standards based
- Capable of being effectively integrated within the enterprise
- May be complex to implement; requires significant planning and support
- Delivers a high return on investment

Enterprise GIS Framework

Enterprise GIS Framework fosters collaboration across organization and functions through application and data integration leading to synthesis, i.e., making everyone's job easier, and more effective



Enterprise ArcGIS Application Architecture

Applications deliver information value to users and organizations, e.g., planning, programming, and project implementation activities for land resource, transportation, and water resources topics. That would be nine applications (3 activities by 3 topics). Of course, many more activities and topics exist.

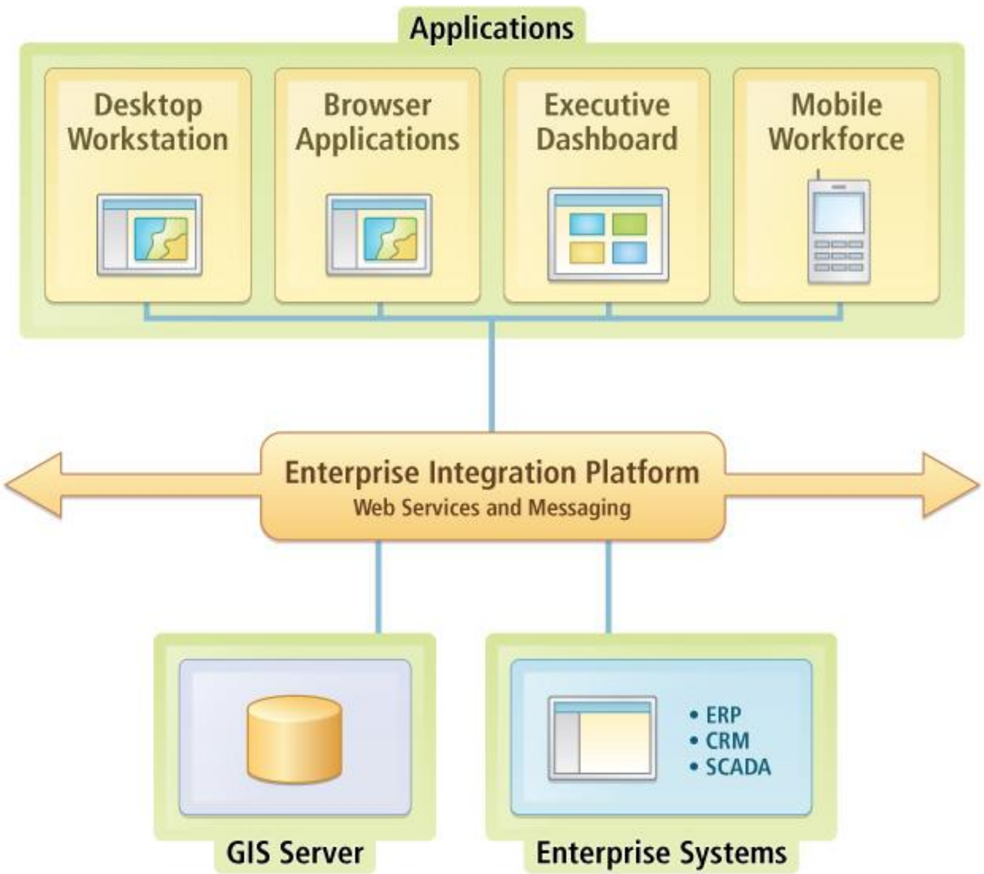
How do we organize information delivery generically?

Four types of application delivery form a GIS application architecture.

- Desktop/workstation – high performance apps
- Browser – distribute displays to many
- Executive dashboard – preset information formats
- Mobile Workforce – field data entry/display

(see next slide for graphic)

Enterprise application architecture



Key terms

Integration Platform – supports information integration

Desktop workstation – high performance apps

Browser – distribute displays to many

Executive dashboard – preset information formats

Mobile workforce – field data entry/display

3. What are the major issues with multi-user data sharing?

- Multi-user databases involving substantial data sharing have been a **long-term goal** for organizations and users for decades
  - New technologies are making it easier
  - Public, private, and not-for-profit sector are all involved
  - Organizational information technology strategy
- To accomplish goal, take advantage of many benefits that exist
- But, there are also significant barriers

Benefits of Geospatial Data Sharing

- Encourage local autonomy, but promote wide collaboration
- Reduce technical costs by minimizing data collection/conversion
- Reduce labor costs through distributing effort across more units
- Increase data quality, thus reduce uncertainty in decision making
- Reduce software costs by sharing applications that manage data
- Expedite application development and deployment
- Reduce risk of “vendor lock-in” and “stranded” technologies
- Increase opportunities and reduce effort for data integration

Barriers to Geospatial Data Sharing

- Inherently complex, e.g. scale, spatial referencing, space-time resolution, feature coding/classification, data models and formats
- People use different terms to mean the same thing, and same term to mean different things
- Differences in data policies, user access protocols, system security measures, and standards among organizations.
- Restricted availability of data, lack of user knowledge about datasets, lack of inability to evaluate usability of datasets
- Organizations unwilling to share data, infringement of copyright and intellectual property, legal liability, fear of losing control
- Restrictions on releasing data due to public regulatory factors such as national security, protection of privacy and archiving requirements
- Coordination among spatial data collectors (local, state, federal), data needs, content, encoding, coverage, revision cycles
- Lack of supporting data discovery and delivery infrastructure, sufficient network bandwidth, organization protocols

Accomplishing data sharing through data integration

- Data integration helps us to understand data similarity and/or differences
- When same, we want to reduce redundancy if possible
- When different, we want to preserve variety, and perhaps even foster it to provide greater insight into topics.
- Promote data management efficiency, effectiveness and equity

Data integration is key to enterprise applications

Data integration…

- Brings data elements together from multiple organizational units
- Enables creation/uncovering geospatial relationships as the foundation of information derivation
- Fosters information insight through information synthesis (identify, integrate, and simplify information content and relationships)
- Provides additional data processing context for information integration
- Enhances opportunity for efficiency and effectiveness in applications
- Clarifies data consistency where data might be redundant

To accomplish Enterprise GIS data integration…

1. Review and evaluate existing datasets…

- native GIS layers,
- GIS-enabled spatial databases,
- associated location-based databases, and
- other enterprise data

… as the first step is essential to understand the assets.

2. Catalog the data from that review

From a User Needs Assessment:

Data should be cataloged

Data should be graded as part of the evaluation

Data should be prioritized for enterprise

- readiness,
- update, and/or
- elimination.

3. Evaluate basic data needs & requirements

Including:

- foundational needs & requirements – data required by other organizational units
- Inter-departmental needs & requirements – data required by two units
- Intra-departmental needs & requirements – data required by single unit

Data review assessment can include

- Completeness in coverage and scope
- Detail of attribute data contained within
- Spatial accuracy and precision of the data
- Attribute accuracy of the information contained
- Spatial integrity
- Applicability for the enterprise

The more users there are for each dataset, the higher the potential value to organization

4. Identify and Grade Data Redundancy

Investigation into enterprise GIS will most likely find

- multiple redundancies in data storage,
- data maintenance, and
- other areas that exist in the current environment.

Redundancies can be graded

- data that is necessary and required,
- data that is not necessary, and
- data that can potentially be eliminated

5. Address data redundancies

Controlled, reduced and/or eliminated through data integration.

- Necessary (controlled) redundancy
- Some redundant data may have to persist.

Redundancies in data

- may be required by local, state, or federal statutes.
- not viewed as an impediment for implementing enterprise GIS.

Should be noted and implementation plans modified to include redundant data through “replication”.

Integration strategy…
Ontology framework fosters shared understanding

Y&H Figure 6.10 Creation and Use of Ontology

Ontology: concepts and their relations within a topical domain

Develop a **controlled vocabulary** (terms associated with concepts)…

- Shared understanding through meaning of data elements
- Establish correspondence among different domains of entities and relations; focus on relationships
- Improve communication among developers, managers, users
- Enable user-centered approach to meaningful data
- Provide underlying concept and technology for interoperable database systems
- Designing spatial databases from an entity perspective

- Local ontology – single database
- Global ontology – across all databases

Integration strategy… Develop a Shared Meaning through Information Mediation

Queries are rewritten to correspond to a common meaning

Queries unpacked into component parts and then matched to data

Database interoperability strategy where queries against multiple heterogeneous data sources are communicated through middleware

- Collection of software components
- Database access optimization rules
- Catalog of information about data sources

See Y&H Figure 6.11 Information Mediation

Key terms
mediation

4. What are approaches to geospatial data sharing?

Data sharing approaches	**Data Sharing Characteristics**			
	Computing Environment	**System Architecture**	**Procedure**	**Purpose and Applications**
Spatial Data Infrastructure	Open Internet standards for distributed data	Global communication network	Global data mediation and info brokering	Societal-wide interoperability and integration
Enterprise	Open Internet standards	Federated data warehouse network	Inter-dept access through mediation and info brokering	Cross-unit data transaction and analytical processing
Domain	TCP/IP, HTTP distributed open database connectivity	Three-tiered client/server in wide area net	Shared databases with collaboration among users	Sector-based data mgt and modeling applications
Functional	TCP/IP, HTTP distributed open DB connectivity	Two-tiered client/server in local area net	Heterogeneous data exchange	Spatial data visualization and overlay analysis
Connected	Peer-to-peer proprietary network	Desktop computer with simple network	Homogeneous data exchange	Electronic exchange of files in same format
Ad Hoc	Standalone computers and independent files	Independent desktop	Manual data exchange	Occasional exchange or sale of data

5. What are three approaches to enterprise data management and how can we characterize them?

Operational databases – support transactional/operational applications for current detailed perspective

- Databases that are used everyday
- Data editing occurs quite often

Data Warehousing – support executive 'overview' applications for broad current perspective

- merge data physically from several sources
- data might be replicated, not necessarily physically moved
- Replication makes a copy for faster performance

Data Federation – support analytic applications for deep perspective across organizational units

- simultaneous online access to multiple sources
- real-time access to distributed sources
- improves maintenance if data copies provide a problem

6. What is data warehousing?

- Subject-oriented – major applications of an organization, e.g. either strategic (long-term) management or operational management (day-to-day) as READ-ONLY easy-to-use browser applications
- Integrated – built from integrating data from multiple sources
- Time variant – all elements are time-stamped
- Non-volatile – applications are usually read-only that do not change the content; key distinction from data federation

Key terms
Data warehouse

Architecture of a data warehouse

Data warehouse server – extracts data from operational databases at set periods; data is cleaned and then stored as a legacy in the data warehouse

OLAP (online application processing) for broadening access to data

OLAP server – maps user queries using

- extended relational OLAP – extensions to standard relational queries
- multi-dimensional OLAP – mixed data model queries

Client applications – front-end interface with tools for

- Query, reporting, data analysis, data mining (but simple easy to use)

Key concept – multi-dimensional data model

Y&H See Figure 6.12

Key terms
OLAP

7. What is a federated database system?

- Create a database architecture that provides uniform and simultaneous access to several heterogeneous data sources

Approaches to Database Federation

- Tight database federation – uses unified schema, also called an integrated or federated schema, as access interface to member data sources of the federation

- Loose database federation – No uniform schema, but a uniform query language to access data from multiple sources.

- Mediated database federation – based on principles and techniques of information mediation, the federated database system protects users from the differences in physical representations of multiple database systems.

Key terms
Federated Database

Architecture of a Federated Database System

- A Federated database makes use of standard database technology, but works with multiple nodes communicating with a connectivity interface such as ODBC, JDBC, and/or DCOM.
- Data sources – structured data in relational or object-oriented databases, geometric and attribute data in spatial databases
- Client applications – can be read only or read/write applications, unlike in the data warehouse where they are read only.
- Smaller datasets are usually involved because more sophisticated and exploratory data analysis is performed. Applications are exploratory because we are not sure what data will really be needed (on the fly processing).

Comparing data warehouses and federated DB systems

Characteristics	Data Warehouses	Federated Database Systems
General description	Collection of subject-oriented data in well-defined and tightly structured repository	Distributed, autonomous, and heterogeneous data using schema integration or information mediation
Typical Systems Architecture	Central data server with distributed clients	Distributed data servers with distributed clients
Data Processing Characteristics	Subject-oriented Easy to control 100 GB to TB High performance	Application-oriented Hard to control 100 MB to GB High availability
Network Requirements	Generally high	Generally low
Interoperability Strategy	Pre-computed, data-oriented to merge data physically from several data sources	On-line applications enable queries on several on-line data sources
Applicability Scenarios	Small number of structured core datasets	Many distributed and heterogeneous data sources
Application Focus	Subject-oriented OLAP for data mining and decision making	Application-oriented OLTP for business operations
Spatial Database Application	Global, national, state and local data for multi-dimensional (space, time, attribute) analysis	Collaborative spatial data analysis using multi-format and multi-media data for data integration

Summary

In this lesson, you learned about…

1. Challenges for the multi-user database environments
2. The nature of Enterprise GIS
3. Major goals to multi-user data sharing
4. Approaches to geospatial data sharing
5. Three approaches to Enterprise GIS Data Management
6. Data warehousing
7. Federated database system

Contact me at nyerges@uw.edu if you have questions or comments about this lesson.

GEOG 482/582: GIS Data Management

END Lesson 10: Enterprise GIS Data Management Strategies