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 goals to 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?
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

**Common user interface** – provide same look and feel and reduce cost of maintenance

**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
  • Multi-organization GIS

Focus of Enterprise, Workgroup and Consortium described in relation to Enterprise
2. What is Enterprise GIS?

Enterprise GIS is a platform for delivering organization-wide geospatial capabilities while improving access to geographic information and extending geospatial capabilities to diverse users of GIS.

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 enterprise systems** by leveraging the value of geographic information
- **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) across organization

**Goal**: getting capabilities out to the organization so that efficiency and effectiveness can flourish.

Obviously, some information custodians will be hesitant because of fear of data misuse / abuse

Those issues need to be addressed in light of the goal
Answer questions about the need for enterprise GIS

• Is there a need for a centralized repository for enterprise data?

• Is there a central point 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 better decision making; greater efficiency with money, time, and resources; and more effective communication.

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

Framework fosters collaboration across organization functions through application and data integration

- Asset Management: Collect, organize & exchange data
- Planning & Analysis: Transform data into actionable information
- Field Mobility: Get information into and out of the field
- Operational Awareness: Disseminate knowledge where & when it's needed

A Complete Integrated System

Desktop | Server | Online
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 applications 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)
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

Enterprise Integration Platform
Web Services and Messaging

Applications

Desktop Workstation
Browser Applications
Executive Dashboard
Mobile Workforce

GIS Server
Enterprise Systems

GEOG 482/582 / My Course / University of Washington
3. What are the major goals to multi-user data sharing?

- Multi-user databases that involve 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
Benefits of Geospatial Data Sharing

• Encourage local autonomy, but promote wide collaboration

• Reduce cost by minimizing need to collect and/or convert data

• Reduce cost 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
Data sharing through data integration

- Data integration helps us to understand how are data are similar/different

- When same, we want to reduce redundancy if possible

- When different, we want to preserve variety

- 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 integration
- Provides additional data context for information integration
- Enhances opportunity for efficiency in applications
- Clarifies data consistency where data might be redundant
To get started with Enterprise GIS data integration...

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.
Catalog the data from 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.
Evaluate basic data requirements

Including:

- foundational requirements – data required by other organizational units
- intradepartmental requirements – data required by two units
- strictly departmental 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
- Precision
- Spatial integrity
- Applicability for the enterprise

The more users for each dataset, the higher the potential value
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
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

Develop a controlled vocabulary...
- 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...
Information Mediation through Common Meaning

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
4. What are approaches to geospatial data sharing?

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

**Operational databases** – support operational applications
- Databases that are used everyday
- Data editing occurs quite often

**Data Warehousing** – support executive applications
- merge data physically from several sources
- data might be replicated, not necessarily physically moved
- Replication makes a copy for faster performance

**Data Federation** – support applications 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
- 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 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
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

Key concept – multi-dimensional data model
Y&H See Figure 6.12
7. What is 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.
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, CORBA, and 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.
## Comparing data warehouses and federated systems

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