Database Management Systems and GIS Design

Overview

This presentation deals with the relationship between database processing requirements, database management systems, and GIS database design theory.

Data, Data Models and Databases

Data has been defined as a collection of facts, concepts, or instructions in a formalized manner suitable for communication or processing by humans (McDonnell and Kemp 1995). It has come to be used as a generic term for any information used by a digital computer.

In this class we will frequently use a related term ‘data model’. In the broad sense this refers to an abstraction of the real world which incorporates only those properties thought to be relevant to the application at hand. In specific, IS, terms Worboys (1995) has defined it as a model of the structure of the information system, independent of implementation details. McDonnell and Kemp (1995) have referred to this as the mechanistic representation and organization of spatial data.

In a digital computer environment, we often interact with data that has been compiled into a database. McDonnell and Kemp (1995) define this as a collection of data organized according to a conceptual schema. A schema is a structured framework.

<table>
<thead>
<tr>
<th>City</th>
<th>Size</th>
<th>Income Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>8</td>
<td>548</td>
</tr>
<tr>
<td>Madrid</td>
<td>4</td>
<td>236</td>
</tr>
<tr>
<td>Moscow</td>
<td>6</td>
<td>126</td>
</tr>
<tr>
<td>N.Y.</td>
<td>11</td>
<td>890</td>
</tr>
<tr>
<td>Paris</td>
<td>10</td>
<td>490</td>
</tr>
</tbody>
</table>

The origins and forms of data

According to Laurini and Thompson (1992), data may have among the following origins:

• Captured – recorded by devices
• Interpreted – some kind of human intervention
  — Field sketches
  — Surveys and questionnaires
• Encoded – as in maps or digital data
• Structured – organized in some way

Requirements of digital databases

In the process of providing functions, including the derivation of information from structured (organized) data, certain requirements have been identified in information science. These include:

• Security – access control and backup
• Reliability – availability of the system providing data
• Integrity – maintenance of linkages between applications and data
• Providing user views of data – access by class of user
• Being independent of the underlying data – access without precise knowledge of implementation details
• Providing support for metadata – data that adds information to the database
• High performance – access and retrieval
• Providing concurrency to users – simultaneous access to data

All of these requirements need to be managed by some type of software (application). Those types of systems responding to these requirements are called database management systems (DBMS).

A DBMS is literally a software system that manages a database
A system providing the functions:

• Data definition
• Manipulation
• Retrieval
• Transaction Management
• Backup/Recovery
The ANSI-SPARC Method

The operation of a DBMS is largely dependent on the structure of the data(base) that it accesses. What is required is a method for designing databases that support the structural requirements of access. To this end the American National Standards Institute (ANSI) Systems Planning and Requirements Committee (SPARC) has developed formalism for dealing with database design issues. The method identifies four levels of data modeling, as a method of tackling the problem of database design.

This method is employed as a critical part of the design of information systems, such as the Life-Cycle approach. There are many varieties of the approach, but one common one is the so-called waterfall model.
To place topics of database and DBMS design in a context for GIS, Worboys (1995) has developed a model for GIS database design. The model responds to the requirements of all information systems and extends it at the beginning to understand the area of use or application domain at hand. The analysis phase has been broken down into an application domain model. This model is, in turn, transformed into a conceptual computational model. This is said to be computationally tractable but independent of any specific approach. The logical model is imbedded in a particular paradigm, for example, a relational approach. The physical model is the representation of the real world as computational processes.
Citations

