UNIT 20 - GIS AS ARCHIVES

UNIT 20 - GIS AS ARCHIVES

- NOTES

This unit introduces a very special type of GIS. Although the examples discussed are older systems, the idea of archival databases is very relevant today as government agencies develop large spatial databases for public use.

UNIT 20 - GIS AS ARCHIVES

A. INTRODUCTION

Archive databases

- are spatial databases developed as stores of information for general use
- provide coverage of some political jurisdiction, e.g. globe, nation, state, province, county
- purposes often not clearly articulated
- system provides limited functionality oriented toward data retrieval

Project databases

- are spatial databases developed to support specific projects
- coverage for study area only
- purposes usually better articulated
- system provides functionality adequate for project

B. NATURE OF ARCHIVES

- map library is traditional archive
- data is partitioned by both theme and geography
    - by theme, e.g.:
        - base maps (usually topographic) include roads, railways, surface hydrology, topography, depending on scale
        - individual thematic map series, e.g. vegetation, soils, transportation, energy
    - by geography, e.g.:
        - USGS 1:24,000 sheets are organized by state, within state by name of sheet
        - index map shows sheets in correct geographical positions
- most atlases and map libraries partition data primarily by geography, secondarily by theme
- digital spatial data archives tend to be organized in an opposite way
    - primary key is thematic
        - topographic data is produced by USGS using DLG (Digital Line Graph) and DEM (Digital Elevation Model) format
        - street network data is produced by Bureau of the Census in TIGER format
        - remotely sensed images are produced by NASA and other space agencies
    - secondary key is geographical
        - topographic data organized by map sheet
        - TIGER organized by county

Landsat images organized by scene

Currency of data

- to be suitable for archiving, data must be stable through time
  - some geographical data never changes - e.g. Census data and satellite images as a representative slice in time never change
  - some changes rarely - e.g. topography and hydrography
  - some changes rapidly - e.g. street networks in rapidly developing cities

- in some cases geographical data needs to be available both in archived and current forms
  - e.g. street network data from TIGER
    - needs to remain in census period format so that census data can be referenced to it
    - needs to be current for use in navigation systems, postal delivery

- advantage of digital archives is that updating, when it does occur, does not require reissuing of hardcopy products
  - updates can be made to central archive
  - updated version can be transmitted by high speed links or distributed as digital tapes

Use of archives

- archives should be established to meet user needs
  - however, users and their needs are often not adequately assessed
  - need to ask:
    - who really needs this type of information? for what purposes?
    - would it get used?
    - does it need to be digital?

- consider the example of a commonly used archive, the phone directory
  - how often does the need to look up phone numbers occur?
  - must be frequently enough to justify the cost of providing phone books with every phone
  - however, even this obvious need for access to phone numbers has not yet justified the development of an on-line database accessible to individual phone customers - access must be through operator

  - (in France the phone directory is available on-line to individuals through an interactive TV system - however, this may be due more to the facts that the TV and telecommunications systems are state-owned and the government is strongly interested in promoting new technology - presumably benefits do not justify the cost of similar systems in North America and elsewhere)

- what does this say about the need for digital geographical archives to replace such things as road maps and area code maps?
  - the phone book archive is very simple compared to the complexity of databases

    - and queries for geographical archives
    - would a digital road map be used frequently enough to justify its cost and updating?

Data for digital archiving

- therefore, data suitable for archiving must
    - be of sufficiently general use to justify cost
    - remain current for sufficient period of time, or be capable of constant update
    - be sufficiently self-explanatory that users do not encounter significant problems of interpretation
        - e.g. user must have access to definition of each object or attribute or attribute value
        - e.g. user should have access to a data quality report

C. EXAMPLES OF SPATIAL DATA ARCHIVE SYSTEMS

- many large raster GIS databases have been built as inventories of natural resources, land use, etc.

CGIS

- started 1962
- Canada Geographic Information System
- designed to allow computer-assisted analysis (measuring area, overlaying different themes) of the data collected by the Canada Land Inventory
- many technical innovations
- one of two claims to original development of the term GIS - the other at Northwestern University (Marble)

MIDAS

- 1964 - U.S. Forest Service
- grid cell overlay and modeling
- first full service GIS for natural resource management

STORET

- 1964 - U.S. Public Health Service
- Division of Water Supply and Pollution Control.
- standardized the data collected by different organizations relating to water quality, flows, treatment processes and location

LUNR

- is a well-documented example of an early inventory project for the State of New York (see Tomlinson et al., 1976)
  - cells were 1 km square which is too large for most planning purposes
  - values could be recorded for percentages within cells, but with no information on where these proportions were located in the cell

NARIS

- Natural Resource Information System
- active in Illinois in 1970s
- raster system based on 40 acre cell
- funded by Ford Foundation, state agencies
- located at Center for Advanced Computation, University of Illinois
- software never adopted for any other purpose, despite the large investment

MLMIS

- is a successful resource inventory for the State of Minnesota - see Unit 9
  - the software developed as part of the analysis and delivery effort for MLMIS, EPPL, has been widely distributed for general purpose use
  - MLMIS cells are 40 acres, but further subdivision is used

D. FATE OF ORIGINAL ARCHIVE SYSTEMS

- most of these original archive systems have become obsolete, expensive experiments that are no longer used
- Why?

Platform

- mainframe systems can not compete with low-cost mini and micro platforms which appeared in the 1980s
- mainframes tend to have unique operating systems which limit portability of software and data
  - many mainframe GIS systems of 1960s and 1970s were never "ported" to newer platforms (e.g. CGIS)
  - entire investment in system on a large mainframe may have to be abandoned because of costs of converting to new platform or operating system

- high cost of keeping large databases on-line (accessible within seconds) on 1970s platforms
  - on-line storage available on typical 1989 workstation would have cost $000s/week on 1970s mainframe
  - these recurring costs could not be offset by revenue generated by database
  - alternative was to store data on tape, incur access delays and maintenance costs

Static data

- systems assumed static data, did not allow for updating

Distribution

- high cost of communicating with mainframes over long distances in 1970s - no high speed, low cost digital communication links

- access effectively restricted to users at central site
  - difficult for remote site, e.g. county planning office, to justify cost of on-line connection

User interface

- command syntax like plain English but rules are still extensive, difficult to learn

- how many users would need the data often enough to justify learning the syntax?
  - e.g. how long does it take a travel agent to forget how to use the airline reservation system?

- command-driven interface requires lengthy, accurate typing

- difficult to make interface user-friendly, forgiving
  - terminals of 1970s did not offer color, operated at frustratingly slow communication speeds

Costs vs benefits

- use failed to materialize at sufficient levels to justify costs
  - inadequate user survey
  - system "over-sold"
  - inadequate anticipation of high recurring costs
  - cost over-runs on development

Development of general purpose systems

- many of the old, special-purpose in-house systems of the 1970s like NARIS have been replaced by general-purpose, vendor-supplied vector systems in the early 1980s
  - these use smaller platforms like the VAX, Prime, Sun costing 1/10 of the large mainframes

special-purpose functionality of archival systems will not support other types of data, other needs
  - high cost of development of system cannot be offset by sales for other uses
  - by contrast, vendor incurs high development costs of general-purpose GIS but can offset costs by multiple sales into diverse market
  - simple low-cost general-purpose raster systems offer comparable functionality without development costs

- general-purpose systems can survive where general-purpose databases cannot
  - ultimately, interest in CGIS and many other archival GISs was in system, not data archive

E. SPATIAL DATA ARCHIVES TODAY

- cheaper platforms, lower recurring cost allows functionality of older systems to be supplied by low-cost workstation

- better awareness of value of digital data archives and GIS-based solutions to resource management has produced a large market for GIS

- important archives today are very general databases like TIGER, USGS digital cartographic data, where user community demand is strong and supporting agency has specific mandate for data collection

REFERENCES

Mounsey, H. and R.F. Tomlinson, 1988. Building Databases for Global Science, Taylor and Francis, New York. A review of the current status of global spatial data archives

Tomlinson, R.F., D.F. Marble and H.W. Calkins, 1976. Computer Handling of Spatial Data. UNESCO Press, Paris. Compares and assesses several archival systems, including CGIS, LUNR, MLMIS.

Wiggins, J.C., R.P. Hartley, M.J. Higgins and R.J. Whittaker, 1987. "Computing aspects of a large geographic information system for the European Community,"

International Journal of Geographical Information Systems 1(1):77-78. A description of the Co-ordinated Information on the European Environment (CORINE) program, a multi-national natural resources GIS project.

EXAM AND DISCUSSION QUESTIONS

1. Review and discuss a selected state natural resource data archive, such as MLMIS, CGIS. How does the system compare with NARIS?

2. The volume Building Databases for Global Science listed in the references contains chapters discussing a number of efforts to construct global databases. What particular problems do they present?

3. Design a study to assess the need for a state natural resource data archive, and to evaluate its potential benefits to the user community.

4. The unit drew a parallel between spatial databases and map libraries. Discuss the validity of this analogy, and the dangers in using it to make design decisions for spatial databases.

Last Updated: August 30, 1997.