PDHonline Course L154G (5 PDH)

Data in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

Slide 1

Lecture 1 Content

- **Geographic Information Systems (GIS)**

Data in GIS – Acquisition and input

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop
1

This lecture is a continuation of the lecture series of Hardware and Software considerations in GIS (course L153) which discusses data in GIS. The topics covered in this series of five lectures are: data acquisition and input; georeference; accuracy; use; sharing; data types; components of a GIS; sources of error in a GIS; data in LIS; scale; types of projections; data distortions; and coordinate systems.

Slide 2

Lecture content:

- **Components of a GIS**
- **Sources of error in a GIS**
- **Data in LIS**
  - **- components of the cadastre system**
  - **- contents of a typical cadastral LIS**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop 2

For this lecture is discussed the essential components of a GIS. Also presented are the sources of error in a GIS. The last component of this lecture is data in a Land Information System (LIS). This term was introduced in an earlier lecture from the lecture series L153 that investigated the various definitions for GIS.

Slide 3

- **Components of a GIS**
  - **A state of the art GIS should be capable of the following:**
    - **Data inputs in one or more formats (e.g. analog maps and overlay information, tabulation, digital image data)**
    - **Storing and maintaining information with necessary spatial relationships**
    - **Manipulate data (search and retrieval, computations, etc.)**
    - **Some level of modeling that takes into account data interrelationships and cause-and-effect responses of appropriate factors**
    - **Present data outputs in a variety of ways (e.g. tabulations, charts, video displays, digital maps, etc)**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop 3

This slide identifies what a state of the art GIS should capable of doing. A state of the art GIS must be able to do the listed functions as presented on this slide.

To facilitate the multiple capabilities of a GIS, the four components of a GIS is collectively utilized. These components are discussed in the following slides.

Slide 4

– **Four components of a GIS:**

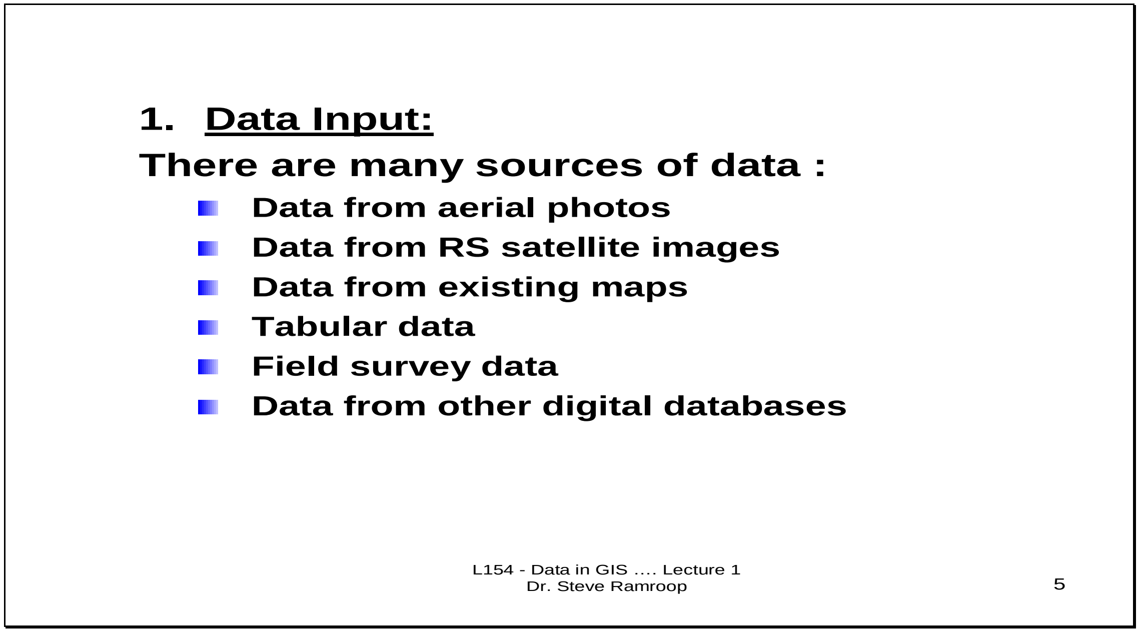
- **Data input**
- **Storage and retrieval (data management)**
- **Data manipulation and analysis**
- **Visualization of products**

The following slides explains the four components

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop 4

This slide shows the four components of a GIS. The components listed are indicative of what is a GIS. It defines GIS in the simplest understanding. Further details into each are discussed in the following slides.

Slide 5



The first component of a GIS is Data Input. Listing of possible data input sources are identified. Details of a set of the various data inputs are variable. The date from aerial photos is using aerial photos to capture larger areas at a faster rate than the standard land surveying techniques. RS (Remote Sensing) are data sets obtained from using sensors on satellites. They are similar to aerial photos but captures global areas at fast rates. Data from existing maps are hard copy maps that are converted (via digitizing or scanning) into digital computer data which can be used in GIS. Tabular data refers to data from existing attribute tabular data that are stored in third party databases such as dBase 1V, access, Oracle, and such like. Lastly, the data from other digital databases are data sets that are available online and also shared by other organizations.

Slide 6

– **Factors influencing data input:**

- **Time**
  – **Approximately 85% of GIS project time on data collection and input**
- **Cost**
  – **(approximately 5-10 times that of GIS hardware and software)**
- **Methods used to capture and store data affects data quality**

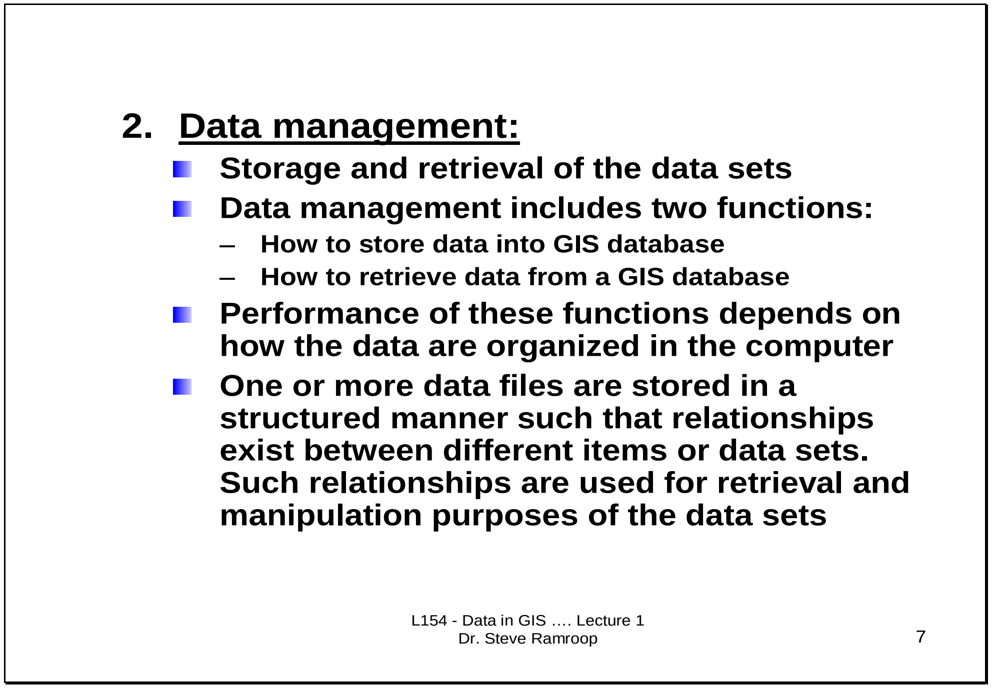
L154 - Data in GIS .... Lecture 1
Dr. Steve Ramroop
6

For the data input component there are three factors that influence data input into a GIS. The first factor is Time. Time is important in that you do not want to spend a lot of time collecting data because it is going to be costly. If you can get access to already available data sets then that is the recommended cheaper alternative, rather than to have the data collected from the field. If data is collected in the filed thane this will mean hiring a crew, and getting instruments, and management. Collectively that will mean a lot of time and a lot of money.

The second factor is Cost. This is related to the time used to collect the data. The cost for GIS software is high but the cost for data is even much higher because data sets are typically collected through a long period of time and requires the hiring of a fieldwork crew, getting the data processed and available for GIS use. This is typically 5-10 times the cost of the hardware and software.

The third factor is a question of data quality. How much of an error is allowed into the data set for it to be useful for the intended GIS application is an indication into its GIS usefulness. The more we process data to minimize errors, the more expensive will be the data set. Various influencing factors can be investigated as it relates to data quality.

Slide 7



The second component of a GIS is the Data Management component which is shown on this slide. The points listed in this slide identify some of the characteristics of what is required under the data management component. In particular this component looks at how the data is stored, and retrieved. Data is stored using various database models which we will explore in a few lectures later in this lecture series. Retrieval of data is through the building of queries (Structured Query Language – SQL). A query is an expression that defines the attributes of the data sets that will satisfy the intended GIS application. It is the relationships between the data sets (graphic and attribute data sets), which are used in the retrieval of suitable data sets.

Slide 8

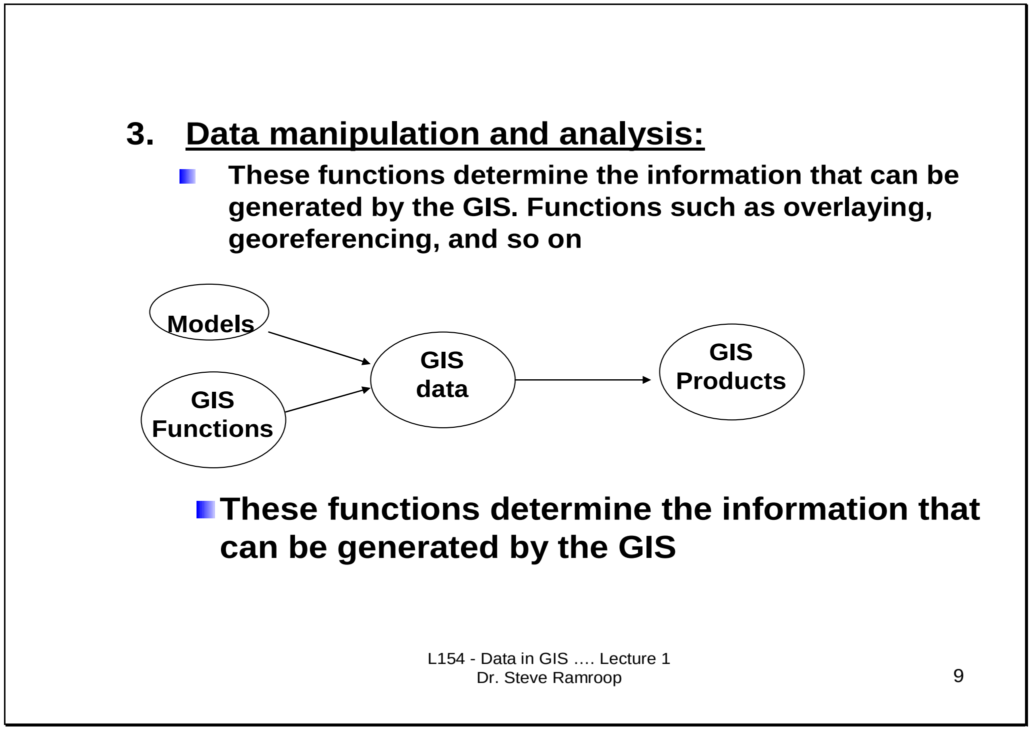
- The organizational aspect of data is often referred to as DATABASE MODEL which can be structured in four ways (discussed later in course L155 in more detail):
  - Hierarchical database model directory structure with redundancies
  - Network database model - directory structure without redundancies
  - Relational database model – tables with defined relationship
  - Object oriented database model – everything is treated as an object
  - Temporal database model – data with time stamp attached with the structure
- Most popular data structure at the moment is “Relational structure” but the trend is towards object oriented
- Examples of relational database are: Oracle, Informix, Ingres, dBaseIV, etc.

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop 8

Data is organized using one of the five database model identified on this slide. The five models are briefly introduced here but will be presented in further detail in a later lecture of the series. The course is L155 which will be with reference to GIS Data Models.

Of the five database models, the relational database model is the most popular. A number or relational database model example exists. A few of the most popular ones are shown on this slide.

Slide 9



This is the third component of a GIS – Data manipulation and analysis. It refers to the GIS processing and analyzing capabilities which are typical of GIS software. Examples of some of the processing capabilities are indicative of the functions of typical functions of GIS software. For example some of the common processing capabilities are: georeferencing, overlaying, buffering, spatial queries, and such like.

The data manipulation brings together the conceptual models of the phenomena, and use of the GIS functions which collectively use the GIS data to get the final GIS products.

Slide 10

- **Users' needs are strongly related in defining models and consequently, analysis functions that GIS needs to perform in order to achieve the required information**
- **Thus user's involvements are necessary during information planning, system design and testing**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop
10

The users' needs are strongly inter-woven into the model's development that defines the conceptual level of modeling. From the model one can make use of the data sets and the GIS functions. Throughout the GIS development, the end user needs to be included because at the end of the day they are the ones who will be using the final product. The success of any GIS application is determined by the acceptance by the end users.

Slide 11

4. Visualization of products

- GIS products or reports can be in the forms of:
  - Maps
  - Tables
  - Charts
- These forms can either be in hard-copy (paper) or soft-copy (digital)
- User involvement is important in specifying the output requirements

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop 11

The fourth component of a GIS is the end product which is typically the visualization of the final end result.

The end result from a GIS can be in the form of hardcopy or softcopy maps, charts, and tables.

The final result must be acceptable to the end user because they will be the ones using it. The end product can also be a customized version of the GIS application, but this will also lead indirectly to the generation of maps. In such software specific output, user specifications must also be met and this can be based upon a standardized set of templates (symbol libraries included in the GIS software) to that of customized systems specifications.

Slide 12

- **Sources of error**
  - **Error is introduced at every step in the process of generating and using geographic information starting from the collection to that of the interpretation**

Data Collection

↓

Interpretation Of results

  - **The objective in dealing with errors in GIS cannot not be to eliminate them but to manage them**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop
12

There are various sources of errors from the time data is being collected. We cannot remove all errors but we must strive to minimize its effect such that it does not affect the intended GIS application. The errors accumulate from the first inception of the GIS phenomena to the final use of the end product. The idea is to choose the data sets that contains the errors that will have the least affect on the GIS application and is in line with the budget of the GIS application. Errors can be minimized using various algorithms (eg. Least Squares Adjustment), but such algorithms may add cost to the budget of the application development. Therefore, a balance must be sought regarding which errors are allowed and which are not.

Slide 13

Stage	Sources of error
Data collection	Errors in field data collection Errors in existing maps used as source data Errors in the analysis of remotely sensed data
Data input	Inaccuracies in digitizing caused by operator and equipment Inaccuracies inherent in the geographic feature (e.g. edges such as forests that do not occur as sharp boundaries)
Data storage	Insufficient numerical precision Insufficient spatial precision
Data manipulation	Inappropriate class intervals Boundary errors Error propagation as multiple overlays are combined Slivers caused by problems in polygon overlay procedures
Data output	Scaling inaccuracies Error caused by inaccuracy of the output device Error caused by instability of the medium
Use of results	The information may be incorrectly understood The information may be inappropriately used

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop
13

This slide gives a brief list of some of the errors throughout the various stages of a GIS application development cycle. This list is exhaustive. You will realize that each component of a GIS has its associated errors. In addition there are errors with data collection procedures and the use of the final GIS results.

Slide 14

- **Data in LIS**
  - **Land Information System (LIS) also termed a Land Related Information System (LRIS)**
  - **both terms used broadly to refer to systems that include land ownership information**
  - **information for LIS is commonly recorded on a large scale map (at scale of 1:1000 to 1:10,000)**
  - **it is administered and maintained by a government unit that is legally responsible for maintaining the land records in the jurisdiction**
  - **the official record of interest in land is termed the <u>cadastre</u>. It is the legally recognized registration of the quantity, value, and ownership of land parcels**

This slide gives you some information about data is used in a Land Information System (LIS). The type of data sets used is typical data sets that are related to the parcel, ownership, and rights to the land. That is, we try to include the land registry with the parcel delineations of property rights.

There is a lot of emphasis on this presently by the GIS and land surveying community. Environmental Systems Research Institute (ESRI), who is the developer of ArcGIS software, has recently introduced the survey analyst software extension that is aimed to bridge the gap between the surveyors and the GIS professionals.

Slide 15

– **The cadastre system used to store and maintain the cadastre consists of three components:**

Records of the cadastral parcels
- **the continuous areas of land (i.e. the land units) within which unique and homogeneous interests are recognized**

The cadastral parcels
- **the graphic (e.g. maps) and text information describing the nature and extent of its land interests**

The parcel index
- **the system for relating parcels and records**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop
15

This slide shows the three important component of a cadastral system and this is reflected by the typical operation of a LIS. The records of the cadastral parcels are the record of property interests, while the cadastral record is the graphic land parcel delineations. The parcel index links the records with the parcels.

Slide 16

– **the database of a LIS will usually include such data layers as:**

- **street network**
- **location and attribute records for land parcels and lots**
- **flood plain hazard**
- **a base map**
- **geodetic control (accurate geographic coordinated points)**
- **utility networks (e.g. water and sewage facilities)**
- **natural resource and environmental data**
- **land use**
- **soil types**
- **etc.**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop
16

This slide shows some additional data sets which can be added to the LIS. When these data sets are added to a LIS, they will all be transformed into the same coordinates used in the LIS. This is important because GIS applications can only be performed using data sets that are using the same coordinate system.

Slide 17

– **Land Information Systems integrate property rights information with information on the uses, values and distribution of natural and cultural resources**

– **Use for a wide variety of such applications as:**
  - **Maintenance of tax and parcel information**
  - **Analysis of changing land use**
  - **Scheduling of road maintenance activities**
  - **Land registry updates**
  - **Maintenance of utility inventories**

– **Potential users:**
  - **public works departments, planners, developers, financial institutions, government officials, general public, etc.**

L154 - Data in GIS …. Lecture 1
Dr. Steve Ramroop

17

This slide shows some of the uses of a LIS and the various professions that will benefit from a LIS.

Slide 18

… The End …