GIS Data Collection

Overview

- Introduction
- Primary data capture
- Secondary data capture
- Data transfer
- Capturing attribute data

Introduction

- Two main types of data capture are
- Primary data sources are those collected in digital format specifically for use in a map.
- Secondary sources are digital and analog datasets that were originally captured for another purpose and need to be converted into a suitable digital format for use in a map.

How is spatial data collected?

- Primary data capture
  - Directly measured: Go out and survey; mostly it measures distance or direction from control points; nowadays GPS is increasingly replacing techniques required for ground surveying
  - Indirect measured: Earth information can be collected without physical contact like using camera mounted on aircraft or electronic recording instrument; such a technique is called remote sensing
- Secondary data capture
  - If some data (directly unusable like it's in analog form) is already available, data can be converted into digital data through scanning or digitizing
- Data transfer
  - If some data (directly usable but in different file format) is already available, only file conversion will be necessary

Data collection workflow

- The workflow commences with planning, followed by preparation, digitizing/transfer (here taken to mean a range of primary and secondary techniques such as table digitizing, survey entry, scanning, and photogrammetry), editing and improvement and, finally, evaluation.
- Planning is obviously important to any project and data collection is no exception. It includes establishing user requirements, garnering resources (staff, hardware, and software), and developing a project plan.
- Preparation is especially important in data collection projects.
- It involves many tasks such as obtaining data, redrafting poor-quality map sources, editing scanned map images, and removing noise (unwanted data such as speckles on a scanned map image)


Planning
Preparation
Digitizing / Transfer
Editing / Improvement
Evaluation

- Planning includes establishing user requirements, garnering resources, and developing a project plan.
- Preparation involves obtaining data, redrafting poor-quality map sources, editing scanned map images, removing noise, setting up appropriate GIS hardware and software systems to accept data.
- Digitizing and transfer are the stages where the majority of the effort will be expended.
- Editing and improvement covers many techniques designed to validate data, as well as correct errors and improve quality.
- Evaluation is the process of identifying project successes and failures.

Data Collection Techniques

	Raster	Vector
Primary	Digital remote sensing images	GPS measurements
	Digital aerial photographs	Survey measurements
Secondary	Scanned maps	Topographic surveys
	DEMs from maps	Toponymy data sets from atlases

Primary spatial data capture

Raster data capture
Vector data capture

Introduction

- Primary geographic capture involves the direct measurement of objects.
- Digital data measurements may be input directly into the GIS database, or can reside in a temporary file prior to input.
- Although the former is preferable as it minimizes the amount of time and the possibility of errors, close coupling of data collection devices and GIS databases is not always possible.
- Both raster and vector GIS primary data capture methods are available.

Primary –Raster data capture:

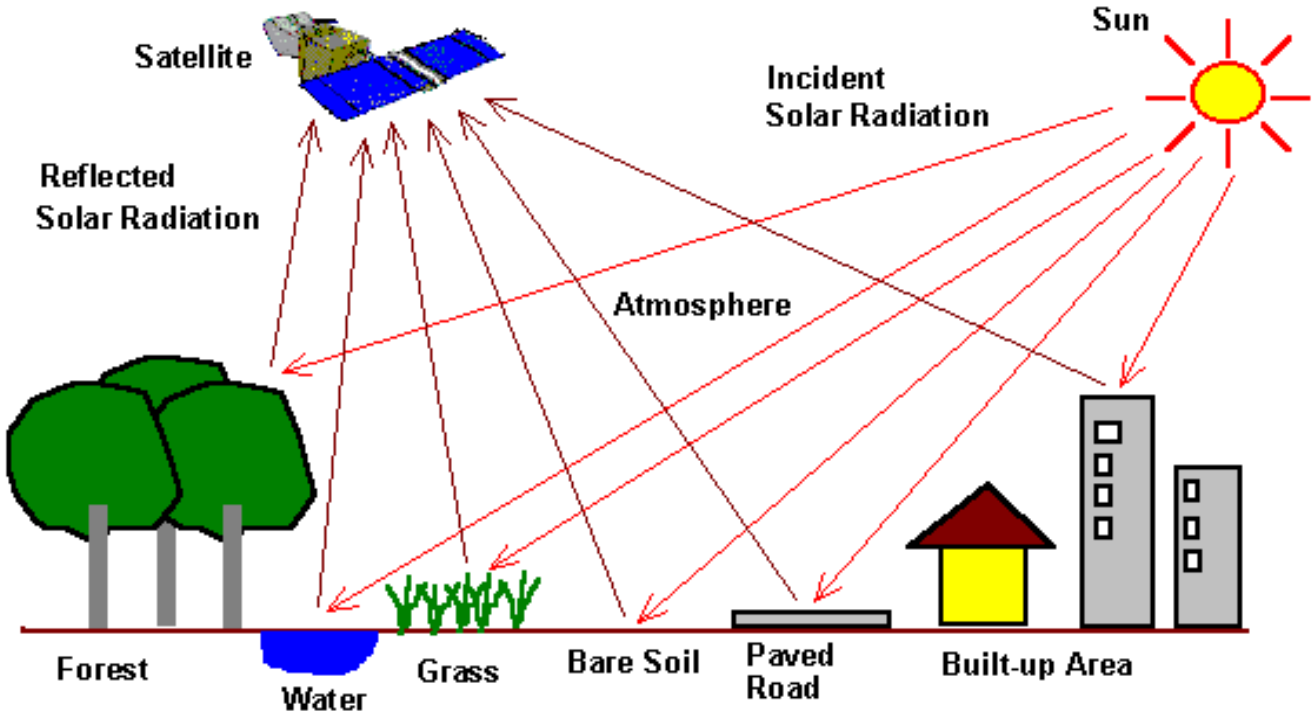
- Remote sensing is a technique used to derive information about the physical, chemical, and biological properties of objects without direct physical contact
  - e.g. SPOT and IKONOS satellites and aerial photography
  - Passive and active sensors
- Resolution is a key physical characteristic of remote sensing systems:
  - Spatial resolution refers to the size of object that can be resolved and the most usual measure is the pixel size.
  - Spectral resolution refers to the parts of the electromagnetic spectrum that are measured.
  - Temporal resolution, or repeat cycle, describes the frequency with which images are collected for the same area.

Primary –Raster data capture:

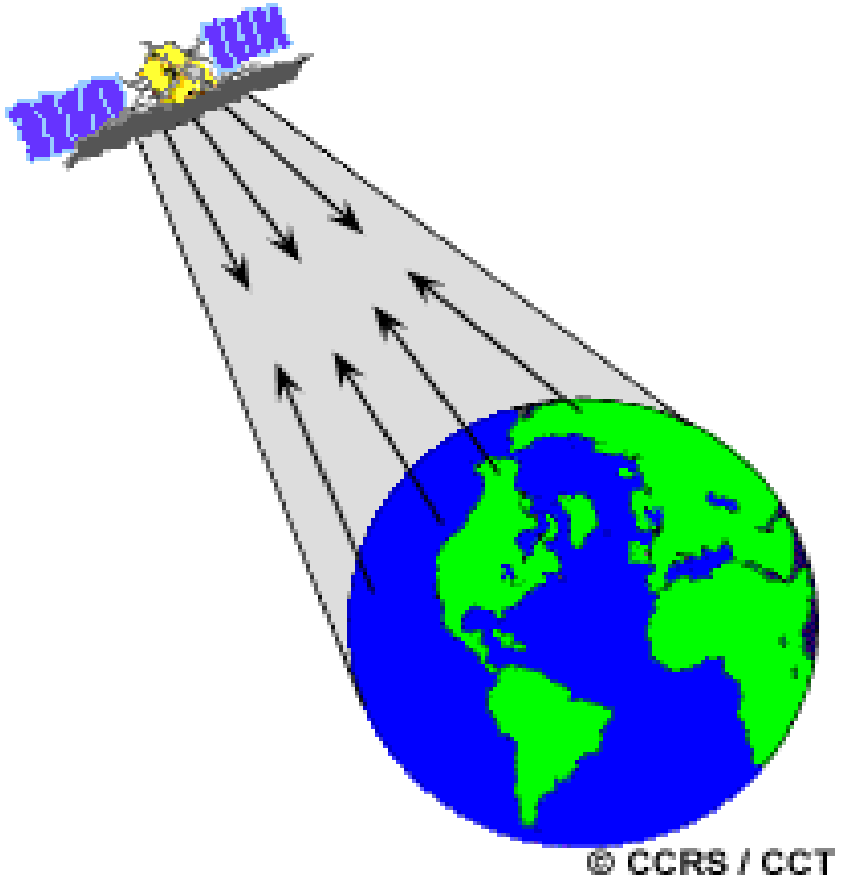
- Advantages are
  - Consistency of the data
  - Availability of systematic global coverage
  - Regular repeat cycles
- Disadvantages are
  - Resolution is often too coarse
  - Many sensors are restricted by cloud cover

1- Passive System:

- It uses an external source of energy, (e.g. SUN).



2- Active System



- In this one the system sends its own energy to the target, receives the reflected energy and senses it. An example is the RADAR.

RADAR = RAdio Detection And Ranging

Primary –Raster data capture:

- Aerial photography is equally important in medium- to large-scale projects

Primary – Vector data capture

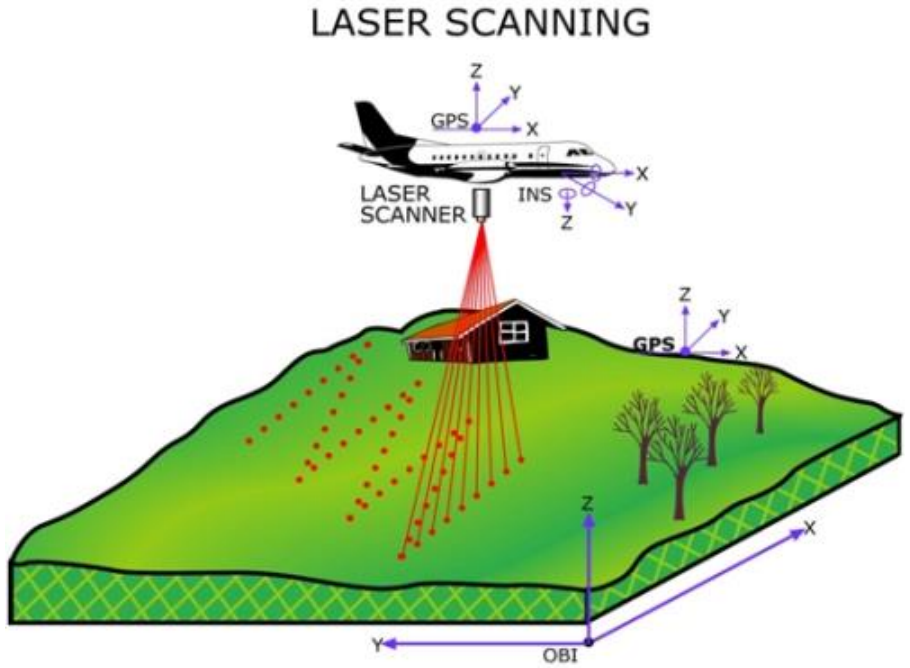
- Primary vector data capture is a major source of
- geographic data.
- The two main branches of vector data capture are:
  - ground surveying
  - LiDAR
  - and GPS

Surveying

- Ground surveying is based on the principle that the 3-D location of any point can be determined by measuring angles and distances from other known points.
- Traditional equipment like transits and theodolites have been replaced by total stations that can measure both angles and distances to an accuracy of 1 mm
- Ground survey is a very time-consuming and expensive activity, but it is still the best way to obtain highly accurate point locations.
- Typically used for capturing buildings, land and property boundaries, manholes, and other objects that need to be located accurately.
- Also employed to obtain reference marks for use in other data capture projects
- Most accurate method for large scale, small areas

LiDAR

- Relatively new technology that employs a scanning laser rangefinder to produce accurate topographic surveys
- Typically carried on a low-altitude aircraft that also has an inertial navigation system and a differential GPS to provide location.



GPS

- GPS
  - Collection of satellites used to fix locations on Earth's surface
  - Differential GPS used to improve accuracy

Secondary spatial data capture

Raster data capture using scanners

Vector data capture

Secondary spatial data capture

- Geographic data capture from secondary sources is the process of creating raster and vector files and databases from maps, photographs, and other hard-copy documents.
- Scanning is used to capture raster data. Table digitizing, heads-up digitizing, stereo photogrammetry.

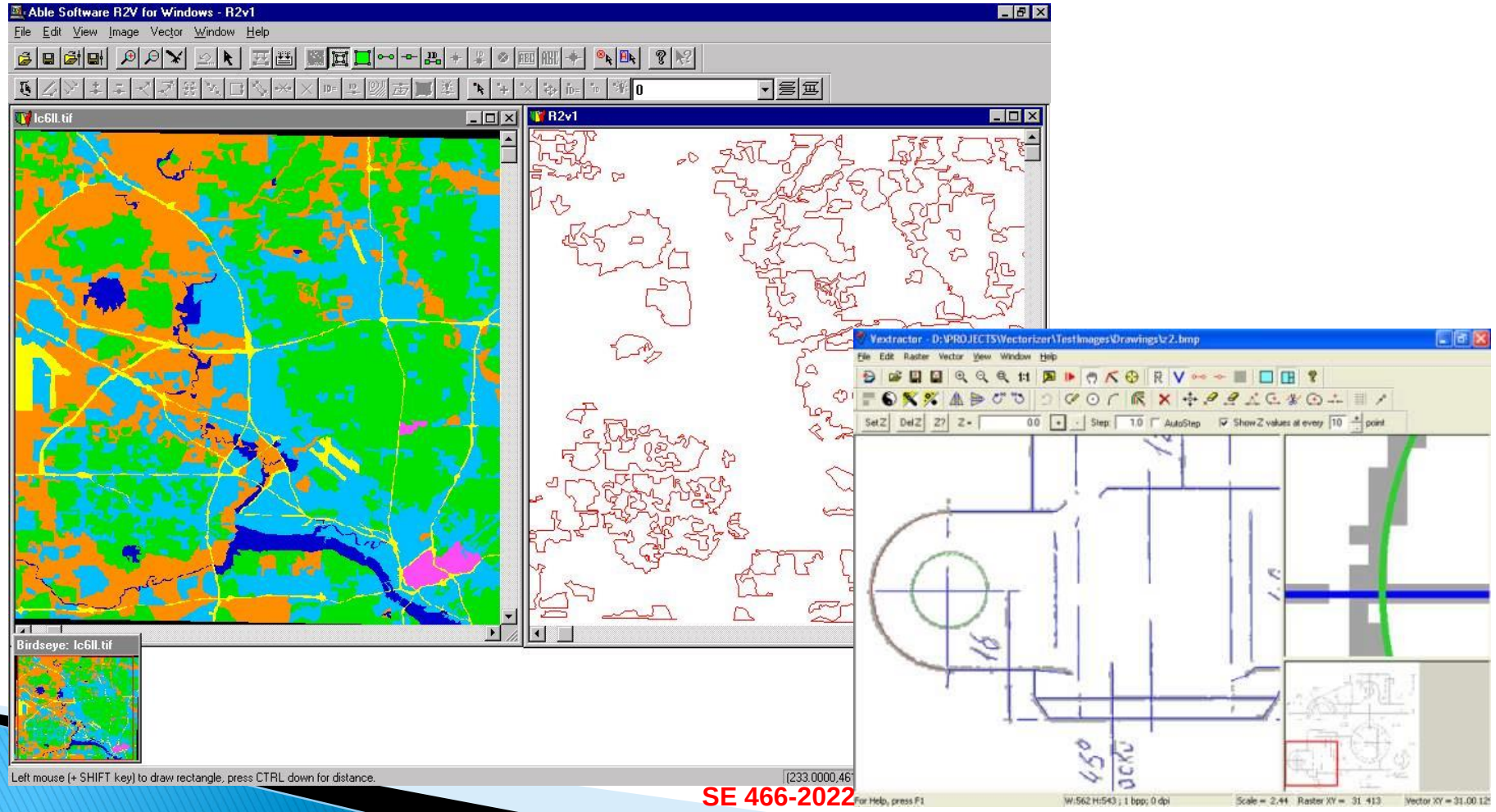
Raster data capture using scanners

- Three main reasons to scan hardcopy media are
  - Documents are scanned to reduce wear and tear, improve access, provide integrated database storage, and to index them geographically
  - Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for other data
  - Maps, aerial photographs and images are scanned prior to vectorization

Scanner

- Scanning is used to capture raster data

Raster to vector conversion



SE 466-2022

Vector data capture

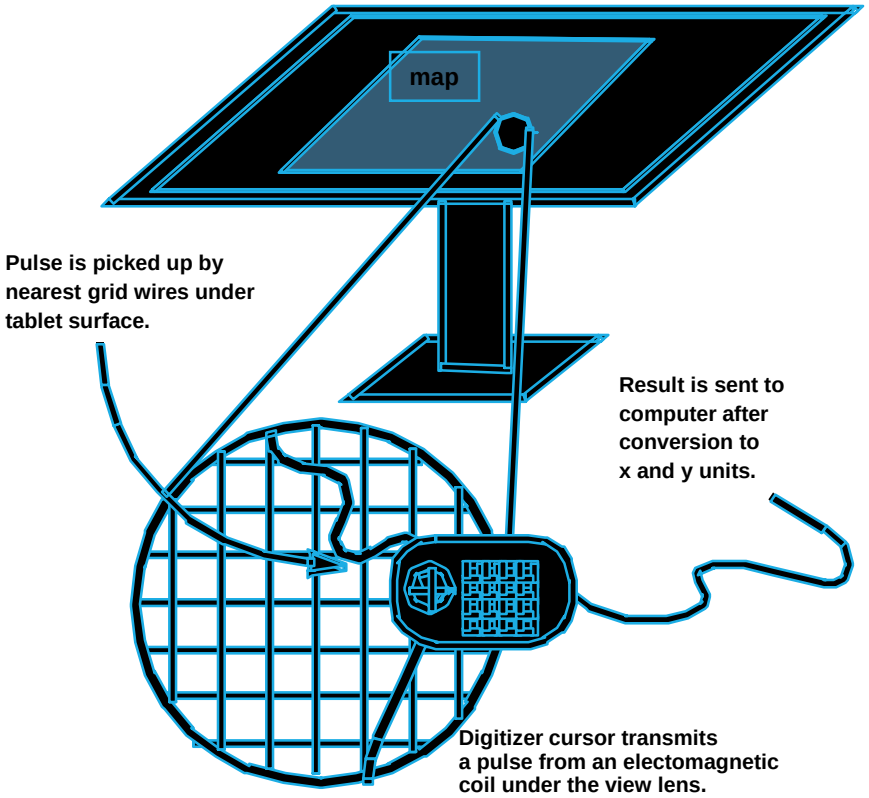
- Secondary vector data capture involves digitizing vector objects from maps and other geographic data sources.
  - Digitizing and vectorization
  - photogrammetry:

Digitizing and vectorization

- Vectorization is the process of converting raster data into vector data.
- The simplest way to create vectors from raster layers is to digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor.
- Describes how automated vectorization is performed

Table digitizing

The Digitizing Tablet



Digitizing

- Stable base map
- Fix to tablet
- Digitize control
- Determine coordinate transformation
- Trace features
- Proof plot
- Edit
- Clean and build

Photogrammetry

- Is the science and technology of making measurements from pictures, aerial photographs, and images.
- Measurements are captured from overlapping pairs of photographs using stereo plotters.
- Orientation and triangulation are fundamental photogrammetry processing tasks.
  - Orientation is the process of creating a stereo model suitable for viewing and extracting 3-D vector coordinates that describe geographic objects.
  - Triangulation (also called _block adjustment‘) is used to assemble a collection of images into a single model so that accurate and consistent information can be obtained from large areas.
- Orthoimages are images corrected for variations in terrain using a DEM.

Photogrammetry

- Stereo photogrammetry is used to capture raster data
- Photogrammetry is the science and technology of making measurements from pictures, aerial photographs, and images.
- Photogrammetry is a very cost-effective data capture technique that is sometimes the only practical method of obtaining detailed topographic data

Vector Over Raster

- An example of raster background data (black and white aerial photography) underneath vector data (land parcels)

Data Transfer

- Buy vs. build is an important question
- Many widely distributed sources of GI
- Includes geocoding
- Key catalogs include
  - Geodata.gov
  - Geography Network
- Access technologies
  - Translation
  - Direct read

Geographic data formats

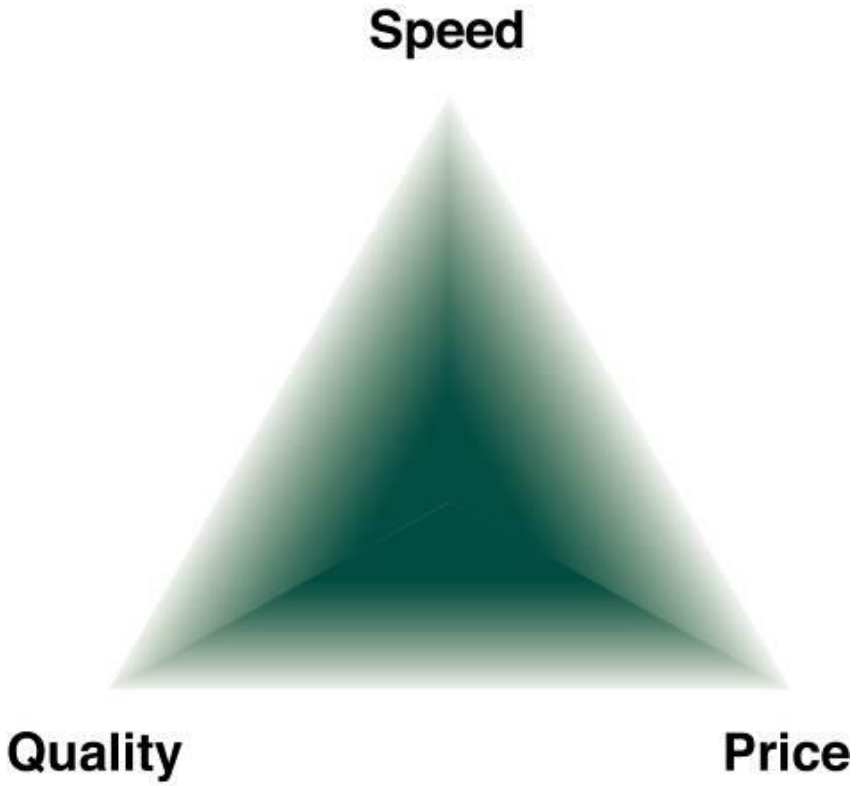
- One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats.
- Many tools have been developed to move data between systems and to reuse data through open application programming interfaces (APIs).
- More than 25 organizations are involved in the standardization of various aspects of geographic data and geoprocessing
- ISO (International Standards Organization) is responsible for coordinating efforts through the work of technical committees TC 211 and 287
- In Europe, CEN (Comité Européen de Normalisation) is engaged in geographic standardization.
- OGC (Open Geospatial Consortium) is a group of vendors, academics, and users interested in the interoperability of geographic systems
- Geographic data translation software must address both syntactic and semantic translation issues.
- Syntactic translation involves converting specific digital symbols (letters and numbers) between systems.
- Semantic translation is concerned with converting the meaning inherent in geographic information.
- While the former is relatively simple to encode and decode, the latter is much more difficult and has seldom met with much success to date.

Capturing attribute data

- Attributes can be entered by direct data loggers, manual keyboard entry, optical character recognition (OCR) or, increasingly, voice recognition.
- An essential requirement for separate data entry is a common identifier (also called a key) that can be used to relate object geometry and attributes together following data capture

Relationship between quality, speed, and price in data capture

- In any data collection project there is a fundamental tradeoff between quality, speed, and price.
- Collecting high-quality data quickly is possible, but it is also very expensive.
- If price is a key consideration then lower-quality data can be collected over a longer period



price in data capture

- If staff costs are excluded from a GIS budget then in cash expenditure terms data collection can be as much as 60–85% of costs.