MODULE 3: GIS data collection methods

OBJECTIVES	✓ Understand the basic concepts of GIS data ✓ Understand how GIS data is collected ✓ Consider the possibilities of working with various GIS data (raster, vector) ✓ Describe the potential use of spatial/EO data ✓ Deal with spatial data interoperability ✓ Consider examples of collecting spatial/EO data, e.g., through an app ✓ Conduct simple spatial data collection on the field
METHODS	Live session, reading material, video's, links to resources, application exercises, quizzes & discussions
DURATION	5.5 hours for participants

SESSION			**DURATION**	**PARTICIPANTS…**
Online	1.0	Introduction to GIS data collection methods	60 min.	✓ Receive a first presentation on the concepts of GIS data collection.
	1.1	Fundamental concepts of spatial data	45 min.	✓ Learn how to define the basic concept of primary and secondary data. ✓ Participants can describe the basic concept of primary and secondary data.
	1.2	Primary GIS data collection methods	30 min.	✓ Learn more about primary data collection.
	1.3	Secondary GIS data collection methods	30 min.	✓ Learn about secondary data capture methods
	1.4	Obtaining data from external sources (data conversion and transfer)	45 min.	✓ Get an understanding of data conversion and transfer.
	1.5	Fieldwork and application exercise on GIS data collection	125 min.	✓ Conduct a simple spatial data collection on the field ✓ Consider examples of collecting spatial/EO data, e.g., through an app ✓ Reflect on content and share experience with peers

Digital Transformation Center Rwanda

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Implemented by
giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Digital Transformation Center Rwanda

FAIR FORWARD
Artificial Intelligence for all.

RSA Rwanda Space Agency

UNIVERSITY of RWANDA

Carnegie Mellon University Africa

GFA CONSULTING GROUP

OPEN MACHINE LEARNING FOR EARTH OBSERVATION (ML4EO)

MODULE 3: GIS data collection methods

FAIR FORWARD
Artificial Intelligence for all.

RSA
Rwanda Space Agency

University of Rwanda

Carnegie Mellon University Africa

GFA CONSULTING GROUP

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Implemented by
giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Digital Transformation Center Rwanda

Digital Transformation Center Rwanda

Before we start

➢ Trainers' team today

Joseph Tuyishimire
Lecturer CGIS

Digital Transformation Center Rwanda

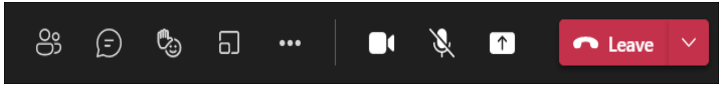
Iris Booyse
ML4EO Teamleader

Rym El Moussaouin
Moodle Expert

Anselme Ndikuryyayo
Education Expert

➢ Communication:
  - Please mute your device if you are not speaking, and switch on the camera/video
  - Please post your questions in the chat and we will answer them on the spot, or during the Q&A session

➢ Material and presentation will be shared



➢ The presentation will be recorded, we assume your consent

german cooperation DEUTSCHE ZUSAMMENARBEIT · Implemented by giz · FAIR FORWARD Artificial Intelligence for all. · RSA Rwanda Space Agency · University of Rwanda · Carnegie Mellon University Africa · GFA CONSULTING GROUP

Structure of today's session

- Course overview
- Review of Module 2
- Introduction to Module 3
- Q&A
- Next steps / assignments
- Closure

Review of Module 2 – FORUM

Experience	Challenges
▪ The overall experience was a challenging one but a learning one as well. ▪ Great experience that required attention to detail; ▪ Very thrilling to finally understand how ESRI images are made ▪ Great navigating through QGIS using different panels and toolbars ▪ QGIS features like the Pansharpening dialog box helped me to go thought, youtube videos were very helpful. ▪ I was impressed by how we provide the training data (images) to the software and how the software uses these training images to show us the places which matches with the ones we provided ▪ Collaborating with colleagues helped overcome some challenges	▪ Tutorial videos used the old version of QGIS=> difficult to follow steps and locating tools ▪ Low accuracy and the challenge to find a better landsat image ▪ The exercise required a lot of time. So did the downloads of complete datasets with the internet connection being slow ▪ Unexpected errors during the image classification process required me to consult other online sources ▪ QGIS maps would disappear abruptly => I had to start over again and failed to provide final graph ▪ I should have saved every step but did not do so and it cost me a lot of time. ▪ Challenge = downloading the correct files due to unclear instructions.

?? QUESTIONS ??

Course overview

The course is composed of **11 Modules**:

- Module 1: Introduction to GIS
- Module 2: Introduction to Remote Sensing
- Module 3: GIS data collection methods
- Module 4: Introduction to ML and Python
- Module 5: Data curation for ML
- Module 6: Visualization of EO data
- Module 7: Predictive modelling using local RS data
- Module 8: Deploying remote sensing-based models
- Module 9: ML workflows and best practice
- Module 10: Business model generation and value proposition design
- Module 11: Project development

Modules 1-4: online + field work
- MS Teams live sessions
- Self-learning on Moodle

Modules 5-10: face-to-face
- Block 1:April 2023
- Block 2: May 2023

Module 11: online & face-to-face
- Project development
- Pitch event

Structure of the Live Session

- Course Overview
- Review of Module 2
- Module 3: GIS data collection methods
  - WHAT are GIS data collection methods?
  - HOW to choose a GIS data collection tool or approach?
  - WHAT are GIS data collection tools?
  - Examples of real-life GIS data collection
- Q&A
- Next steps / assignments
- Closure

MODULE 3:
GIS data collection methods

What do YOU know about GIS data collection methods ?

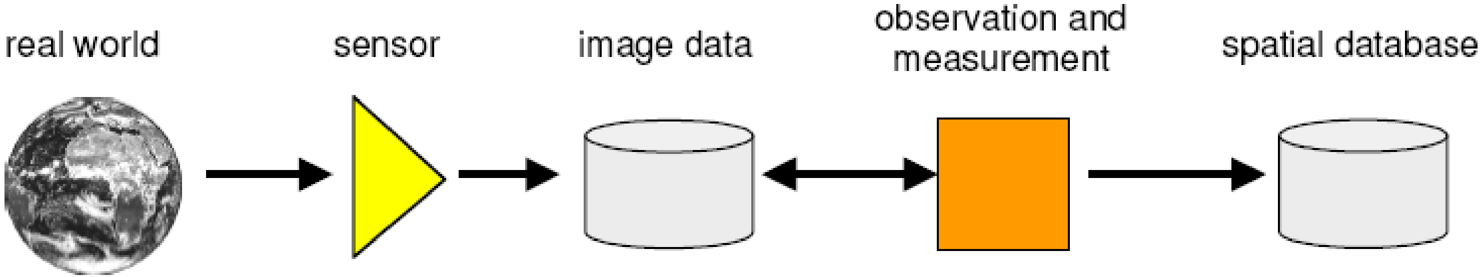
Please raise your hand

or write your questions into the chat box

to respond!

WHAT are GIS data collection methods?

Ground based approach



Remote sensing based approach

WHAT are GIS data collection tools?

Ground based approach

Office based approach

- Preparation
- Time-consuming
- Minimization of errors
- Dust in dust out

Remote sensing based approach

german cooperation
DEUTSCHE ZUSAMMENARBEIT
Implemented by giz
Digital Transformation Center Rwanda
FAIR FORWARD
Artificial Intelligence for all.
RSA
Rwanda Space Agency
Carnegie Mellon University Africa
GFA CONSULTING GROUP

HOW to choose a GIS data collection tool or approach?

1. Problem to be solved
2. Needed details
3. Accuracy
4. Invested resources
5. Size of the area
6. Accessibility of the area
7. Replicability of the methodology
8. Data integration
9. Storage and processing capacity

german cooperation
DEUTSCHE ZUSAMMENARBEIT
Implemented by giz
Digital Transformation Center Rwanda
FAIR FORWARD
Artificial Intelligence for all.
RSA
Rwanda Space Agency

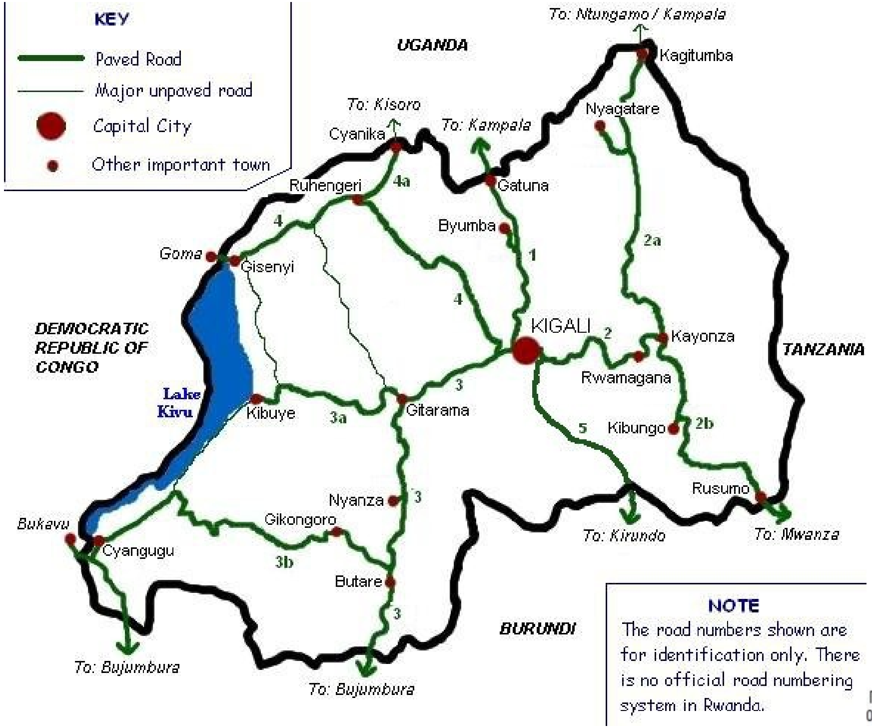
Examples of real-life GIS data collection

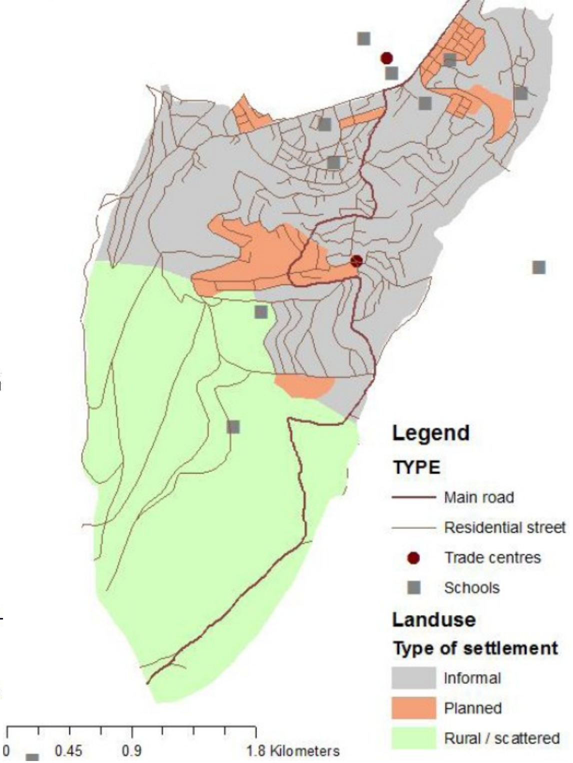
Land registration in Rwanda

german cooperation DEUTSCHE ZUSAMMENARBEIT | Implemented by giz | Digital Transformation Center Rwanda | FAIR FORWARD Artificial Intelligence for all. | RSA | UR | Carnegie Mellon University Africa | GFA CONSULTING GROUP

Examples of real-life GIS data collection

Infrastructure planning-Roads and settlements





- National Scale
- Small scale-urban and rural
- Feeder roads-connecting farmers to markets-access to inputs and sale of production
- National priority-planning, extension and maintenance
- Ongoing in collaboration with different partners

Examples of real-life GIS data collection

Land survey-Demarcation of the Park

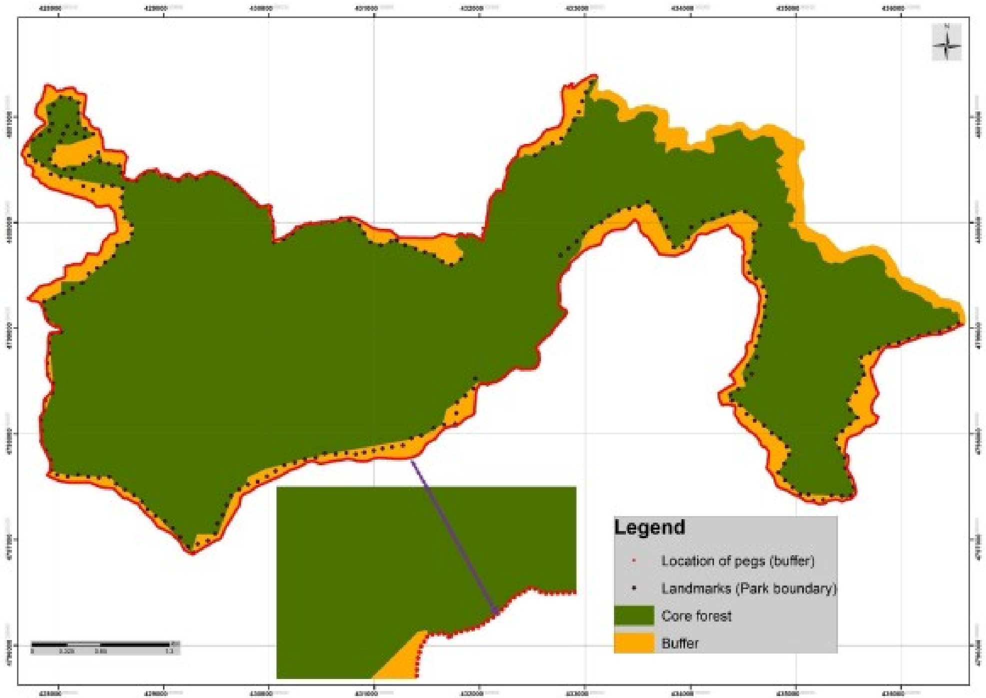


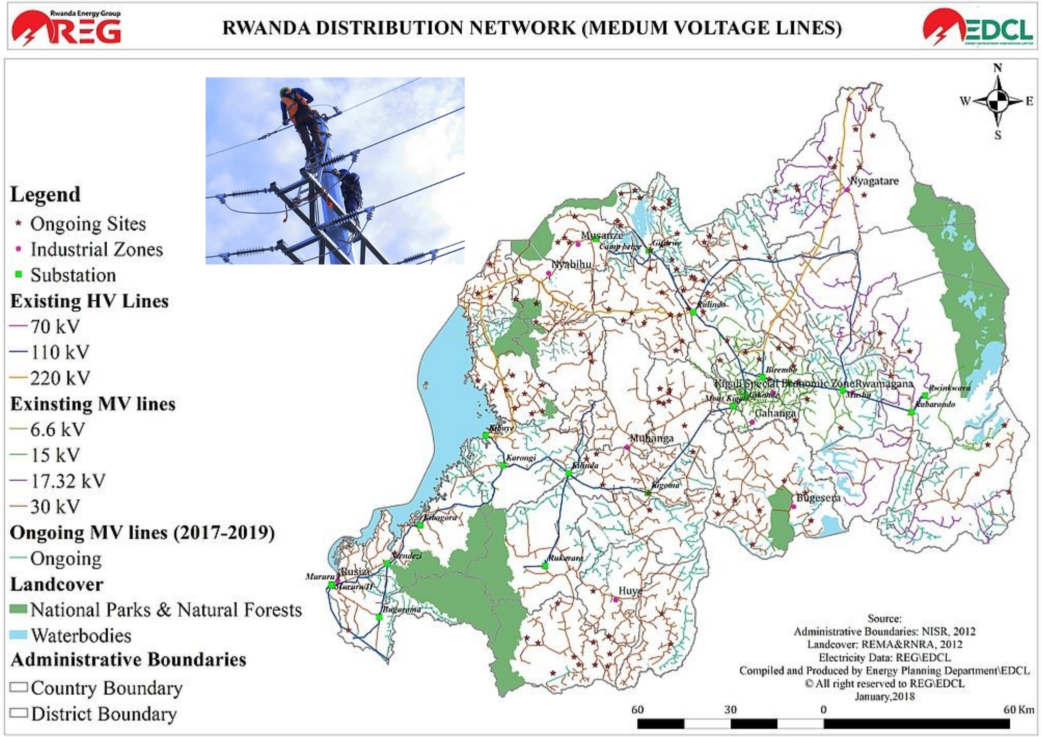
Figure 1. Process for pegs fixation: (a) Painting, (b) digging the hole and (c) Fixed pegs

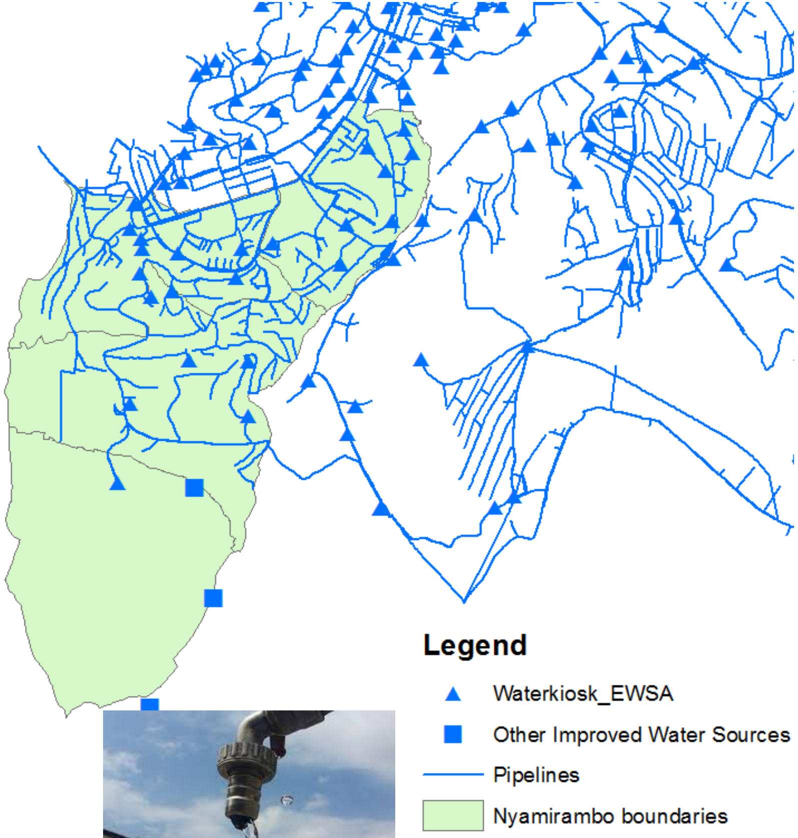
Table 3. Location of pegs around Gishwati Forest Patch

A			B			C			D		
ID	X	Y	ID	X	Y	ID	X	Y	ID	X	Y
1	433117	4801374	364	432045	4799911	727	428382	4800149	1089	428600	4797408
2	433105	4801389	365	432034	4799894	728	428402	4800150	1090	428613	4797394
3	433086	4801390	366	432024	4799876	729	428422	4800153	1091	428631	4797389
4	433066	4801389	367	432033	4799861	730	428442	4800155	1092	428651	4797390
5	433049	4801380	368	432035	4799841	731	428462	4800158	1093	428668	4797380
6	433031	4801370	369	432030	4799832	732	428481	4800161	1094	428685	4797369
7	433013	4801360	370	432012	4799840	733	428493	4800147	1095	428701	4797359
8	432996	4801351	371	431995	4799850	734	428504	4800130	1096	428720	4797351
9	432979	4801340	372	431977	4799859	735	428518	4800115	1097	428738	4797343

Examples of real-life GIS data collection

Infrastructure: Water and electricity supply

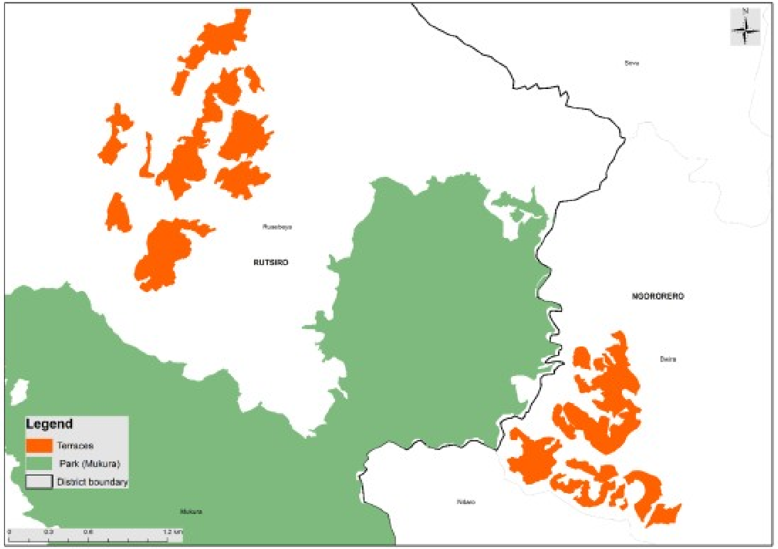




german cooperation DEUTSCHE ZUSAMMENARBEIT | Implemented by giz | Digital Transformation Center Rwanda | FAIR FORWARD Artificial Intelligence for all. | RSA Rwanda Space Agency | UR | Carnegie Mellon University Africa | GFA CONSULTING GROUP

Examples of real-life GIS data collection

Land management



german cooperation DEUTSCHE ZUSAMMENARBEIT · Implemented by giz · Digital Transformation Center Rwanda · FAIR FORWARD Artificial Intelligence for all. · RSA Rwanda Space Agency · UR · Carnegie Mellon University Africa · GFA CONSULTING GROUP

Q & A

DO YOU HAVE ANY QUESTIONS?

Please raise your hand

or write your questions into the chat box!

Structure of today's session

- Course overview
- Review of Module 2
- Introduction to Module 3
- Q&A
- Next steps / assignments
- Closure

Individual tasks

- **Self-study part Moodle – Module 3 -** to be completed by Sunday, **12/03/2023**
  - Read all contents of all sessions
  - Watch all videos
  - answer all quiz questions after each session (3 attempts)
  - Perform the application exercise

- **Please also indicate if your name and email may be disclosed to RSA (see email - asap)**

- **Please confirm if data bundles has been successfully loaded (see email – by tomorrow)**

Group Assignment on Moodle – Module 3

- To be completed by Sunday, **12/03/2023**

Contribution to the Forum:

- Post your results of the application exercise and your answers to the questions in the forum
- Read the contributions of the other group members
- Answer any question you got in response to your own post

Evaluation of the Live Session

- Before we close… please do a short evaluation of our live session
- Either: Click on the link in the chat
- Or: go to www.menti.com + submit the code (given in the chat)
- THEN participate by answering the questions

CLOSING & LAST ANNOUNCEMENT

THANK YOU

for sharing this time with us!

See you
IN TWO WEEKS
CHANGE OF DATE! => **Monday, 20.03, 3pm:**

Module 4: Introduction to Machine Learning

3.2 Fundamental concepts of spatial data

Reading material

Now let us start with the first lesson of the module: Fundamental concepts of spatial data

In this lesson, we are going to define the basic concept of primary and secondary data, and how can we describe them. The data can be displayed as vector or raster. It is used as input for QGIS and many other data analysis tools (refer to Module 1).

Basic concepts:

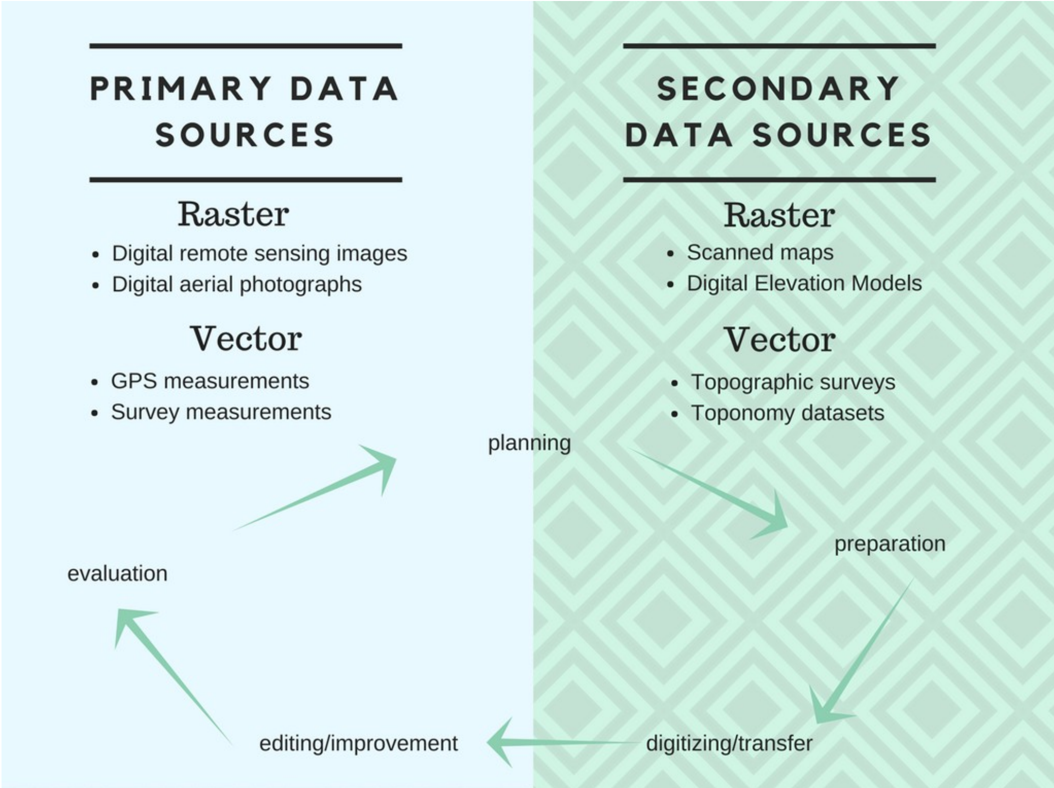
Primary data vs. secondary data

Spatial data is generally obtained from various sources. It can be collected from scratch, using direct spatial-data acquisition techniques, or indirectly, by making use of existing spatial data collected by others. The first source could include field survey data and remotely sensed images. To the second source belongs printed maps and existing digital data sets.

One way to obtain spatial data is by direct observation of relevant geographic phenomena. This can be done through ground-based field surveys or by using remote sensors on satellites or aircraft. Many Earth science disciplines have developed specific survey techniques as ground-based approaches remain the most important source of reliable data in many cases.

Data that are captured directly from the environment are called primary data. With primary data, the core concern in knowing their properties is to know the process by which they were captured, the parameters of any instruments used, and the rigour with which quality requirements were observed.

In contrast to direct methods of data capture, spatial data can also be sourced indirectly. This includes data derived by scanning existing printed maps, data digitized from a satellite image, processed data purchased from data-capture firms or international agencies, and so on. This type of data is known as secondary data. Secondary data are derived from existing sources and have been collected for other purposes, often not connected with the investigation at hand.



Source: https://uizentrum.de/the-process-of-data-collection-in-gis/?lang=en

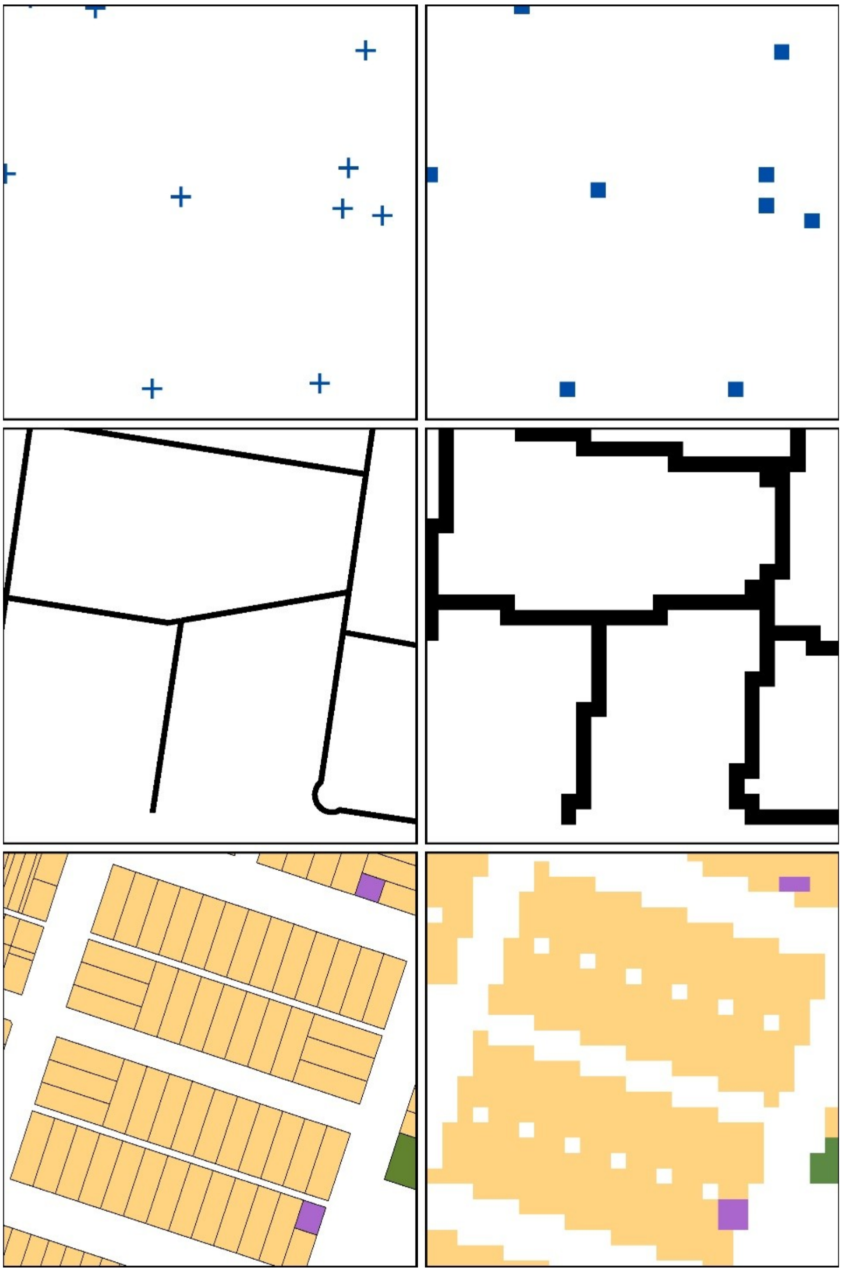
Spatial data, or data related to a specific location on Earth, is collected into two formats or types: raster and vector. Both types of spatial data can be paired with attribute data, which describes any additional information that isn't tied to location. For example, the coordinate's location of a building would be classified as spatial data, while the name of that building would be considered attribute data.

Vector Data

Data in this format consists of points, lines or polygons to simply the view of real-world as spatial data. At its simplest level, vector data comprises of individual points stored as coordinate pairs that indicate a physical location in the world. These points can be joined, in a particular order, to form lines or joined into closed areas to form polygons. Vector data is extremely useful for storing and representing data that has discrete boundaries, such as borders or building footprints, streets and other transport links, and location points. Ubiquitous online mapping portals, such as Google Maps and Open Street Maps, use data in this format.

Raster Data

Raster data provides a representation of the world as a surface divided up into a regular grid array, or cells, where each of these cells has an associated value. In an alternate sense, we can consider a digital photograph as an example of a raster dataset. Here each cell, which in this instance is referred to as a pixel, corresponds to a particular colour value (Digital Number). When transferred into a GIS setting, the cells in a raster grid can potentially represent other data values, such as temperature, rainfall, or elevation. The main point of difference between the digital photograph and the GIS representation is that in the GIS there is accompanying data detailing where the cells can be found on a globe and how big these cells can be.



Source: https://spatialvision.com.au/blog-raster-and-vector-data-in-gis/

Comparing raster and vector data is like comparing apples and oranges. To analyse and draw conclusions from GIS data, it has to be integrated. Scanned maps and some other raster files can be converted into vectors format – a process referred to as vectorization.

Converting pixels to points can take a great deal of time by hand, but luckily, GIS software can help speed up the process. The reverse process is also possible- and is referred to as rasterization.

Attribute data

Attribute data are the information linked to the geographic features (spatial data) that describes features, characteristics of geographic features that are quantitative and /or qualitative in nature, Attribute data is information appended in tabular format to spatial features. Attribute data can include text or numeric descriptors: i.e., nominal, ordinal, or interval/ratio data types Usually, a table is used to display attribute data, each row represents a single feature.

Attribute data are also known as non-geographic information data /non- spatial data. Attribute data helps to obtain the meaningful information of a map. Every feature has characteristics that we can describe.

House Number	Size (sq ft)	Color	Cuteness
1100	5000	Red	Very
1589	5000	White	Average
1760	6000	Yellow	Not
1999	1000	Red	Average

Source: https://storymaps.arcgis.com/stories/7c4d2d73eaf64263a438bcb9c6ece5a4

Attribute data models

A separate data model is used to store and maintain attribute data for GIS software. These data models may exist internally within the GIS software or may be reflected in external Database Management Software (DBMS)- either commercial or open source . A variety of different data models exist for the storage and management of attribute data. The most common are:

- Tabular
- Hierarchical
- Network
- Relational
- Object Oriented

The tabular model is the manner in which most early GIS software packages stored their attribute data. The next three models are those most commonly implemented in database management systems (DBMS). The object oriented is newer but rapidly gaining in popularity for some applications.

The simple tabular model stores attribute data as sequential data files with fixed formats (or comma delimited for ASCII data), for the location of attribute values in a predefined record structure. This type of data model is outdated in the GIS arena. It lacks any method of checking data integrity, as well as being inefficient with respect to data storage, e.g., limited indexing capability for attributes or records, etc.

The relational database model is the most widely accepted for managing the attributes of geographic data.

Relational Model

The relational database organizes data in **tables**. Each table, is identified by a unique table name, and is organized by **rows** and **columns**. Each column within a table also has a unique name. Columns store the values for a specific attribute, e.g., cover group, tree height. Rows represent one record in the table. In a GIS each row is usually linked to a separate spatial feature, e.g., a forestry stand.

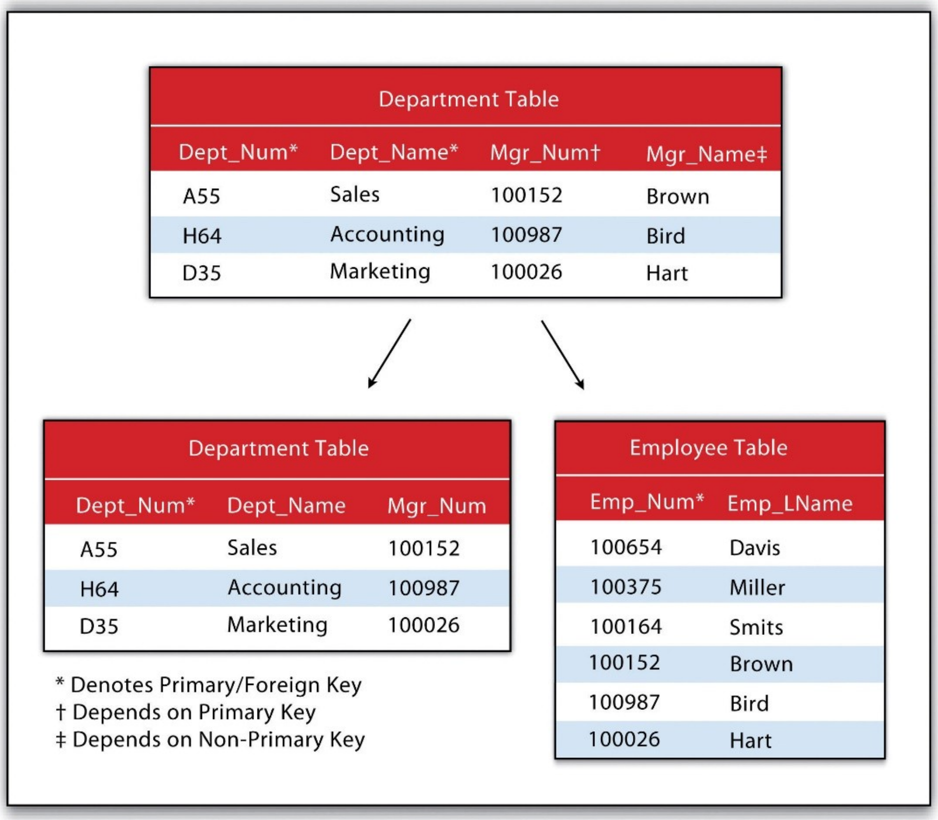
Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature. The following figure presents a sample table for forest inventory features. This table has 4 rows and 5 columns. The forest stand number would be the label for the spatial feature as well as the primary key for the database table. **This serves as the linkage between the spatial definition of the feature and the attribute data for the feature.**

UNIQUE STAND NUMBER	DOMINANT COVER GROUP	AVG. TREE HEIGHT	STAND SITE INDEX	STAND AGE
001	DEC	3	G	100
002	DEC-CON	4	M	80
003	DEC-CON	4	M	60
004	CON	4	G	120

Data is often stored in several tables. Tables can be joined or referenced to each other by common columns (relational fields). Usually, the common column is an identification number for a selected geographic feature, e.g., a forestry stand polygon number. This identification number acts as the primary key for the table. The ability to join tables through use of a common column is the essence of the relational model. Such relational joins are usually ad hoc in nature and form the basis of for querying in a relational GIS product. Unlike the other previously discussed database types, relationships are implicit in the character of the data as opposed to explicit characteristics of the database set up.

The primary key represents the attribute (column) whose value uniquely identifies a particular record (row) in the relation (table). The primary key may not contain missing values as multiple missing values would represent nonunique entities that violate the basic rule of the primary key. The primary key corresponds to an identical attribute in a secondary table (and possibly third, fourth, fifth, etc.) called a **foreign key**.

This results in all the information in the first table being directly related to the information in the second table via the primary and foreign keys, hence the term “relational” DBMS. With these links in place, tables within the database can be kept very simple, resulting in minimal computation time and file complexity. This process can be repeated over many tables as long as each contains a foreign key that corresponds to another table’s primary key.



Department Table			
Dept_Num*	Dept_Name*	Mgr_Num†	Mgr_Name‡
A55	Sales	100152	Brown
H64	Accounting	100987	Bird
D35	Marketing	100026	Hart

Department Table		
Dept_Num*	Dept_Name	Mgr_Num
A55	Sales	100152
H64	Accounting	100987
D35	Marketing	100026

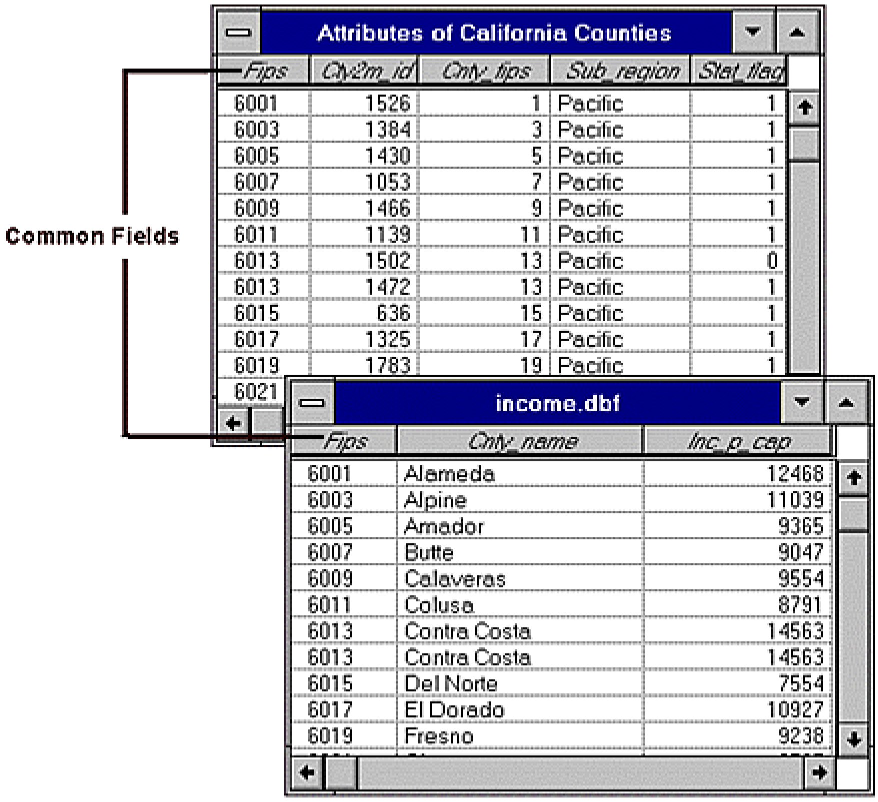
Employee Table	
Emp_Num*	Emp_LName
100654	Davis
100375	Miller
100164	Smits
100152	Brown
100987	Bird
100026	Hart

* Denotes Primary/Foreign Key
† Depends on Primary Key
‡ Depends on Non-Primary Key

Source: https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s09-02-geospatial-database-management.html

There are many different designs of DBMSs, but in GIS the relational design has been the most useful. In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link them together.

This surprisingly simple design has been so widely used primarily because of its flexibility and very wide deployment in applications both within and without GIS.



Fips	Cty2m_id	Cnty_fips	Sub_region	Stat_flag
6001	1526	1	Pacific	1
6003	1384	3	Pacific	1
6005	1430	5	Pacific	1
6007	1053	7	Pacific	1
6009	1466	9	Pacific	1
6011	1139	11	Pacific	1
6013	1502	13	Pacific	0
6013	1472	13	Pacific	1
6015	636	15	Pacific	1
6017	1325	17	Pacific	1
6019	1783	19	Pacific	1
6021				

Fips	Cnty_name	Inc_p_cap
6001	Alameda	12468
6003	Alpine	11039
6005	Amador	9365
6007	Butte	9047
6009	Calaveras	9554
6011	Colusa	8791
6013	Contra Costa	14563
6013	Contra Costa	14563
6015	Del Norte	7554
6017	El Dorado	10927
6019	Fresno	9238

Source: http://wiki.gis.com/wiki/index.php/Database_management_system

In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link them together.

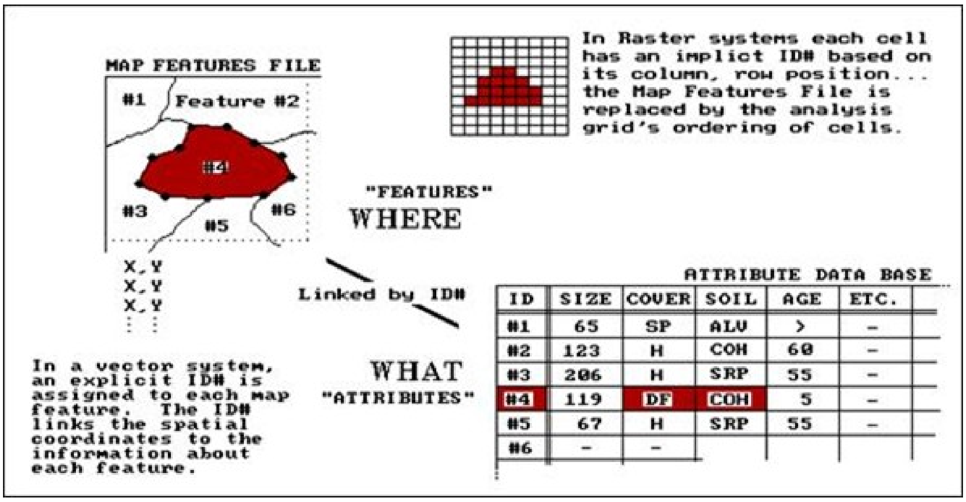
In fact, most GIS software provides an internal relational data model, as well as support from external relational DBMS' such as MySQL, PostGIS, PostgreSQL etc. This approach supports both users with small data sets, where an internal data model is sufficient, and customers with larger data sets who utilize a DBMS for other corporate data storage requirements. With an external DBMS the GIS software can simply connect to the database, and the user can make use of the inherent capabilities of the DBMS. External DBMS' tend to have much more extensive querying and data integrity capabilities than the GIS' internal relational model. The emergence and use of the external DBMS is a trend that has resulted in the proliferation of GIS technology into more traditional data processing environments.

The relational DBMS is attractive because of its:

- simplicity in organization and data modelling.
- flexibility - data can be manipulated in an ad hoc manner by joining tables.
- efficiency of storage - by the proper design of data tables redundant data can be minimized; and
- the non-procedural nature - queries on a relational database do not need to take into account the internal organization of the data.

The relational DBMS has emerged as the dominant data management tool in GIS implementation and application.

The following diagram illustrates the basic linkage between a vector spatial data (topologic model) and attributes maintained in a relational database file.



Source: https://studfile.net/preview/2069449/page:3/

Data dimensions in EO

There are three dimensions of spatial data:

1. **Spatial**: spatial dimension of data includes various characters or symbols that communicate to the user information about the location of the feature being observed
2. **Temporal**: The temporal dimension provides a record of when the data were collected (or the record to which data applies).
3. **Thematic/attribute**: The thematic dimension shows the characteristic of a real-world feature to which the data refer. In GIS, thematic data are often referred as non-spatial, or attribute, data.

Spatial data quality

Data quality is the degree of data excellency that satisfy the given objective. In other words, completeness of attributes in order to achieve the given task can be termed as Data Quality. Data created from different channels with different techniques can have discrepancies in terms of resolution, orientation, and displacements. Data quality is a pillar in any GIS implementation and application as reliable data are indispensable to allow the user obtaining meaningful results.

Spatial Data quality can be categorized into **Data completeness, Data Precision, Data accuracy and Data Consistency**;

- Data Completeness: It is basically the measure of totality of features. A data set with minimal number of missing features can be termed as Complete-Data.
- Data Precision: Precision can be termed as the degree of details that are displayed on a uniform space. More about precision: GIS Data: A Look at Accuracy, Precision, and Types of Errors
- Data Accuracy: This can be termed as the discrepancy between the actual attributes value and coded attribute value.
- Data Consistency: Data consistency can be termed as the absence of conflicts in a

particular database.

But on the other side, Data quality is a relatively abstract construct that is sometimes difficult to interpret. Because the quality depends on the applications or the intended use of data. In short, “Data quality” refers to the **fitness for use of data** for intended applications (Chrisman, 1983).

Criteria for defining a GEOSPATIAL data quality:

- Reliable and accurate
- Current and up to date for the applications
- Relevant and timely for the applications
- Complete and precise
- Format that can be easily maintained, transmitted, distributed, classified, updated etc.
- Adequately protected i.e., control access to data integrity
- Should have an associated metadata

Four generic measures of data quality

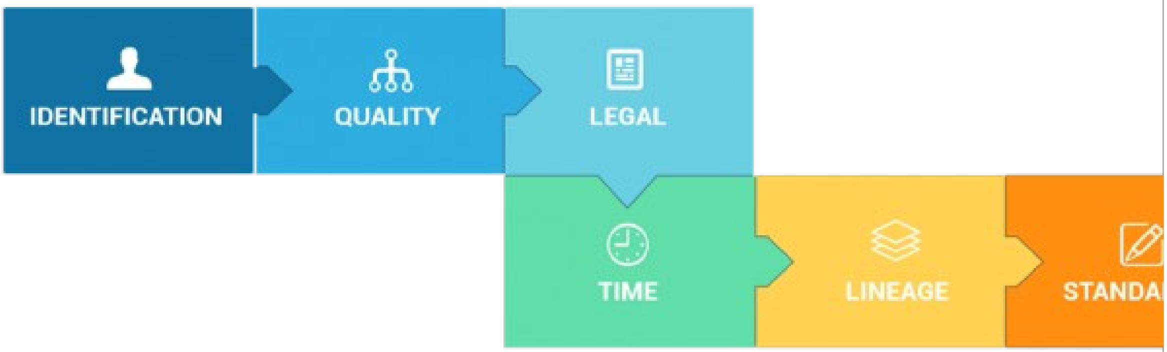
1. **Accuracy**: degree to which data agree with the values or descriptions of real-world features that they represent. It is usually application specific. Accuracy is one of the most important factors governing the cost of data collection.
2. **Precision**: tells one or indicate the measures how exactly data are measured and stored. High precision does not necessarily mean high accuracy while high accura-cy does not necessarily always require high-precision data representation. Precision has different meaning when applied to categorical data.
3. **Error**: measure of error is relative to measure of accuracy in that highly accurate data are supposed to be free of errors. Three major types of errors during measure-ments and observations:
   i) gross error
   ii) systematic errors
   iii) random errors (e.g., Total error = Gross+ Systematic + Random)
4. **Uncertainty**: certain degree of doubt about the applicability or validity of the information derived from the data. Simply it is the lack of confidence in the use of the data due to incomplete knowledge of the data.

For more details about GIS data quality please refer to the following link https://www.e-education.psu.edu/geog160/node/1922

Metadata for EO

Metadata is a summary document providing content, quality, type, creation, and spatial information about a data set. It is like an instruction manual for data because it describes the who, what, when, where, why, and how of data. It has to be detailed, dependable, and well-documented It can be stored in any format such as a text file, Extensible Markup Language (XML), or database record. Because of its small size compared to the data it describes, metadata is more easily shareable. By creating metadata and sharing it with others, information about existing data becomes readily available to anyone seeking it. Metadata makes data discovery easier and reduces data duplication.

Below is the necessary information that have to be incorporated in the metadata file:



Source: https://gisgeography.com/gis-metadata/

GIS metadata identification provides a brief narrative of your data. In other words, it summarizes the purpose of your data in a succinct way. For example, identification assigns the following to your metadata:

- Title – Name of the data set
- Description – The features in the data set and what they represent.
- Keywords – By adding keywords, it helps categorize your data with predefined taxonomy.

GIS metadata contact includes details on who developed and makes the data available. For example, it includes the following three entities: 1.) the ***Originator, the entity that*** developed the data set, 2.) ***the Publisher, the entity that*** assists in producing, editing, and finalizing the end product. and 3.) the ***Distributor, whose*** main focus is to make the data available.

GIS metadata quality explains the accuracy and standards the data set followed. For example, it includes **horizontal and vertical positional accuracy**. Quality also evaluates tests of quality including **completeness, integrity, and inspections** of the data.

Spatial Reference information assigns a geographic extent and coordinate system Projection information includes a projection, datum, and units. Geographic extent comes in the form of a bounding box, place keyword, or thumbnail.

Entity and Attribute: entities refer to the map data type such as points, lines, polygons, or grids. The purpose of this metadata item is to describe how to represent the spatial information in the data. For the entity attributes, it includes a description with a list of valid values and domains.

Lineage describes in detail the creation of the data. For example, it lists the processing steps and responsible parties. Each processing step has a date when it took place so users can track changes. It's like a changelog listing the evolution of the data from start to finish.

Legal section outlines the constraints for accessing and distributing the data. It describes the liability to assure the protection of privacy and intellectual property. Metadata includes a security classification that handles the restriction over security concerns. For example, confidential, restricted, sensitive, unrestricted, and unclassified are examples of security classification in metadata.

- **Temporal**: Temporal information focuses on when the data were collected or updated and how long it's valid for. It also states the progress such as when there will be future updates.

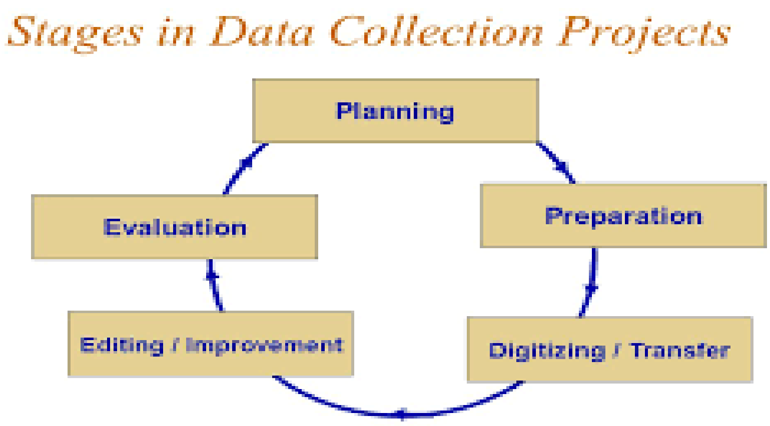
The frequency of updates can be anywhere from daily, weekly, monthly, or annually.

- **Metadata Reference:** the metadata reference section is specific to the metadata. It gives a point of contact when there are uncertainties such as how to cite information when used. The metadata reference has a temporal component for when the user created and will revisit it next.
- **Metadata Standard:** For GIS metadata standards, geographic data providers follow guidelines from the Federal Geographic Data Committee (FGDC), ISO 19115, EPA, Esri, Inspire, and MEDIN. Each schema was developed to best suit their particular requirements and needs. More on this later.

Spatial data collection process

There are mainly five stages of data collection process:

1. **Planning:** this stage includes establishment user requirements, gatherings resources and developing project plan.
2. **Preparation:** this involves obtaining data, redrafting poor quality map sources, editing scanned map images, removing noise, setting up appropriate GIS hardware and soft-ware to accept data
3. **Digitizing and transfer** are the stages where most of the effort will be expended.
4. **Editing and improvement** covers many techniques designed to validate data, as well as correct errors and improve quality.
5. **Evaluation is** the process of identifying project successes and failures.



Source: Muvunyi Germain, INES Ruhengeri (lecture notes)

Data collection and production are the most cumbersome and time-consuming steps of any kind of GIS and remote sensing techniques. They cover 75% of the total implementation cost for most of GIS project.

The processes of data collection are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing.

Exercise materials and tasks

Quiz questions

Please answer the following questions to test your understanding so far:

1. Please complete by filling in the blank space with the correct term.

Data that is captured directly from the environment is called _______ data. (**primary**/secondary)

2. Spatial data, or data related to a specific location on earth, is collected in which format?

   a. Raster
   b. Vector
   c. Metadata

What are three dimensions of spatial data?

   a. Accuracy
   b. Spatial
   c. Precision
   d. Metadata
   e. Temporal
   f. Thematic/attribute
   g. Error

What criteria is defining geospatial data quality?

   a. Current and up to date
   b. Reliable and accurate
   c. No need to have metadata
   d. Complete and precise
   e. Does not have to be protected

Sort in order stages of data collection:

preparation, editing and improvement, evaluation, planning, digitizing and transfer

Step 1	
Step 2	

Step 3	
Step 4	
Step 5	

Answer:

Step 1	**planning**
Step 2	**preparation**
Step 3	**digitizing and transfer**
Step 4	**editing and improvement**
Step 5	**evaluation**

3.3 Primary GIS Data Collection methods

Reading material

Primary GIS Data Collection methods

In this session you will learn more about primary data collection.

Primary data or raw data is a type of information that is obtained directly from the first-hand source through experiments, surveys, or observations. A typical example of primary data are household surveys. In this form of data collection, researchers can personally ensure that primary data meets the standards of quality, availability, statistical information and sampling required for a particular research question.

You will now go through five data collection methods:

1. Remote Sensing
2. Surveying
3. LiDAR – vector data capture
4. Global Navigation Satellite System (GNSS)
5. Mobile Sensor system

1. Remote Sensing

Remote sensing is used to determine the chemical, physical, and biological properties of an area without physical contact – making it one of the most popular methods for gathering raster data. This type of GIS data collection is primarily carried out by satellites and aircraft sensors, which can assess the characteristics of a location by measuring the electromagnetic radiation objects on the surface emit.

Remote sensing is advantageous to GIS professionals for several reasons. To start, remote sensing can cover massive areas, including spots that would be difficult to get to physically (such as the middle of an ocean). In addition, this GIS data collection technique allows for continuous information gathering, which can be helpful when monitoring local temperatures, assessing water levels, measuring air quality, etc.

One form of remote sensing used for GIS data collection, typically conducted by high-quality analogue optical cameras, is photography. GIS photography can come in many different forms, including satellite images, aerial photographs, thermal images, and digital elevation models (DEMs).

While extremely useful, digital elevation models are more challenging to create than traditional photos are. To create one, you'll have to use stereo imagery, which means you'll need two satellites to take a picture of the same location from different angles.

In contrast, aerial photography is one of the earliest, simplest remote GIS data collection methods – and today, it's one of the cheapest. Aerial photography tends to capture small areas from a much lower height than satellite imagery. While the two can find similar types of information, aerial photography is typically reserved for smaller-scale applications such as agricultural management, land use, and marketing properties.

Aerial surveys (a) and satellite remote sensing (b) are employed to map relatively large areas at comparably large scales, as illustrated in the images below:

Shuttle Radar Topography Mission, U.S. Geological Survey Department of the Interior/USGS and NASA, JPL. https://ltb.itc.utwente.nl/509/learningoutcome/show/76976

2. Surveying

Surveying is the science of accurate measurement of natural and human made features on the earth. Data collected by surveyors are then used to create highly precise maps. Surveyors calculate the precise position of points, distances and angles through geometry. This includes most of the time the use of instruments such as total stations, EDM etc, ...

Though not the most efficient or cost-effective methodology, ground surveys are the most accurate form of vector-based GIS data collection. For this reason, surveys tend to be the method of choice when it comes to measuring buildings, property lines, and other features where precision is a top priority. The basic principle in play here is that location can be determined by assessing the direction and distance from other known locations. Accordingly, surveys are an excellent way to determine a reference point for other forms of GIS data collection.

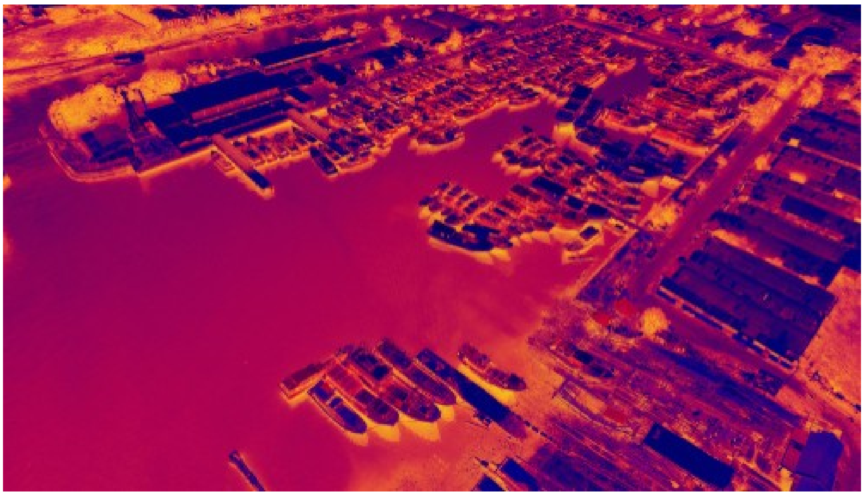
Source:https://ltb.itc.utwente.nl/509/learningoutcome/show/76976

More on how land surveying works, you can find in the next video.

https://www.youtube.com/watch?v=SPCewaAfqPA (time duration 6:25)

3. LiDAR – vector data capture

LiDAR which stands for Light Detection and Ranging, is a relatively new technology used for GIS data collection. If topographic surveys weren't exciting enough already, these surveys are created with lasers. LiDAR is rather complex, but the main components are low flying aircraft and GPS technology. The instruments gather in-depth information on the Earth's shape by scanning an area with flashing lights, resulting in an accurate, almost instant topographic survey.



Source: https://www.aidash.com/remote-sensing-the-art-behind-geospatial-data-collection-for-non-experts/

Find out how LiDAR data is collected in this video:

https://www.youtube.com/watch?v=H2-Yp30TGk4 (time duration 4:09)

4. Global Navigation Satellite System (GNSS)

Global Navigation Satellite System (GNSS) refers to a constellation of satellites providing signals from space that transmit positioning and timing data to GNSS receivers. The receivers then use this data to determine location. Earth-orbiting satellites broadcast their locations in space and time, of networks of ground control stations, and of receivers that calculate ground positions by trilateration.

Source : https://www.gps.gov/

The performance of GNSS is assessed using four criteria:

1. Accuracy: the difference between a receiver's measured and real position, speed or time;
2. Integrity: a system's capacity to provide a threshold of confidence and, in the event of an anomaly in the positioning data, an alarm;
3. Continuity: a system's ability to function without interruption.
4. Availability: the percentage of time a signal fulfils the above accuracy, integrity and continuity criteria.

At present GNSS include two fully operational global systems: the United States' Global Positioning System (GPS) and the Russian Federation's GLObal NAvigation Satellite System (GLONASS), as well as the developing global and regional systems, namely Europe's European Satellite Navigation System (GALILEO) and China's COMPASS/Bei-Dou, India's Regional Navigation Satellite System (IRNSS) and Japan's Quasi-Zenith Satellite System (QZSS). Once all these global and regional systems become fully operational, the user will have access to positioning, navigation, and timing signals from more than 100 satellites.

5. Mobile sensor system

With a mobile Sensors system and the support of a satellite receiver, we can take GIS into the field with us on powerful, compact mobile computers and view, capture and update information, and then synchronize changes between the field and office. **Mobile GIS provides the integration of mapping, GIS and positioning to field users via hand-held and mobile devices.**

Source: https://ltb.itc.utwente.nl/509/learningoutcome/show/76976

Exercise materials and tasks

Quiz questions

Please perform the following sorting exercise to check if you have understood the different methods and their features:

Please link the features on the right with their corresponding method:

Town expansion, car navigation, tree height, water pipeline, crop observation, plot border, boat location, traffic sign, digital elevation model, garbage can, flood, civil engineering, forest

Remote sensing	
Surveying	

LiDAR	
GNSS	
Mobile GIS/GPS	

Answer:

Remote Sensing - town expansion, flood, forest, crop observation

Surveying - plot border, civil engineering, water pipeline

LiDAR - tree height, digital elevation model

GNSS - boat location, car navigation

Mobile GIS/GPS - traffic sign, garbage can

3.4 Secondary GIS data methods

Reading material

In this lesson, you are going to learn about secondary data capturing methods.

Secondary data collection refers to gathering information that is already available. The data was previously collected, has undergone necessary statistical analysis and is not owned by the researcher. This data is usually one that was collected from primary sources and later made available for everyone to access.

Scanning & Georeferencing of Analogue Maps

Scanning is one of the simplest ways to convert physical documents into digital formats. When you scan a document, the scanner uploads a digital image of the page. Like all images, scanned documents will appear in raster format.

Maps are often scanned in order to:

- Use digital image data as a background for other (vector) map data.
- Convert scanned data to vector data for use in a vector GIS.

Scanning requires that the map scanned be of high cartographic quality, with clearly defined lines, text and symbols; be clean and have lines of 0.1mm width or wider.

Scanning comprises two operations:

- scanning, which produces a regular grid of pixels with grey-scale levels (usually in the range 0-255)
- binary encoding – to separate the lines from the background using automated feature recognition techniques.

Editing of scanned data can include pattern recognition of shapes and symbol candidates; line thinning and vectorization; error correction; supplementing missing data and forming topology.

- Once maps are scanned, they will then be subjected to ***Georeferencing process*** which is a process of associating features on the scanned image with real world x and y coordinates. It is therefore the process through which geographical information systems use the positions of some reference points to fix the geographic location of each topographic feature so that it can be viewed, queried, and analyzed with other geographic data. Georeferencing may involve shifting, rotating, scaling, skewing, and in some cases warping, rubber sheeting, or orthorectifying the data.

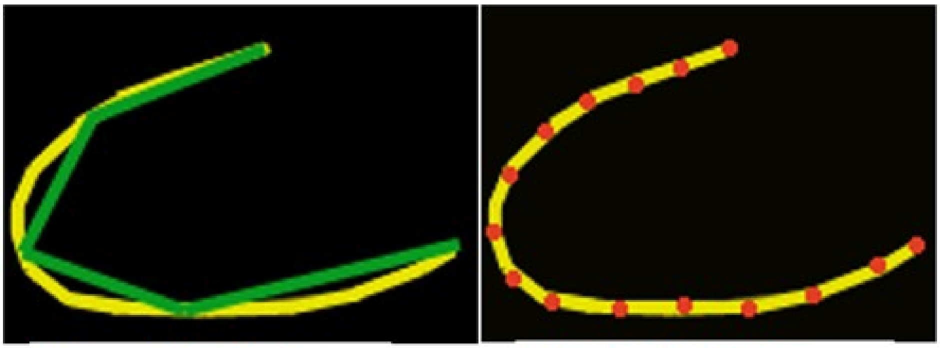
Digitizing and vectorization – vector data capture

Digitizing is the transformation of information from analogue format, such as a paper map, to digital format, so that it can be stored and displayed with a computer. Digitizing can be manual, semi-automated (automatically recorded while manually following a line), or fully automated (line following).

Manual digitizing involves an operator using a digitizing table (or tablet) (known as heads-down digitizing), or with the operator using a computer screen (heads-up digitizing).

The two most commonly used digitizing methods are :

- Point-Mode digitizing : A method of digitizing in which the digitizer selects particular points, or vertices, to encode.
- Stream mode digitizing : a method of digitizing in which, as the cursor is moved, points are recorded automatically at preset intervals of either distance or time. With stream digitizing you set a tolerance, and points will be added only when you move the specified number of pixels in either the x or y direction. A large tolerance leads to fewer points, while a small tolerance leads to smoother curves.



Point-mode digitising   Stream-mode digitising

Source : Muvunyi Germain, INES Ruhengeri (lecture notes)

Topology expresses the spatial relationships between connecting or adjacent vector features (points, polylines and polygons) in a GIS. Topological or topology-based data are useful for detecting and correcting digitising errors (e.g. two lines in a roads vector layer that do not meet perfectly at an intersection). Topology is necessary for carrying out some types of spatial analysis, such as network analysis.

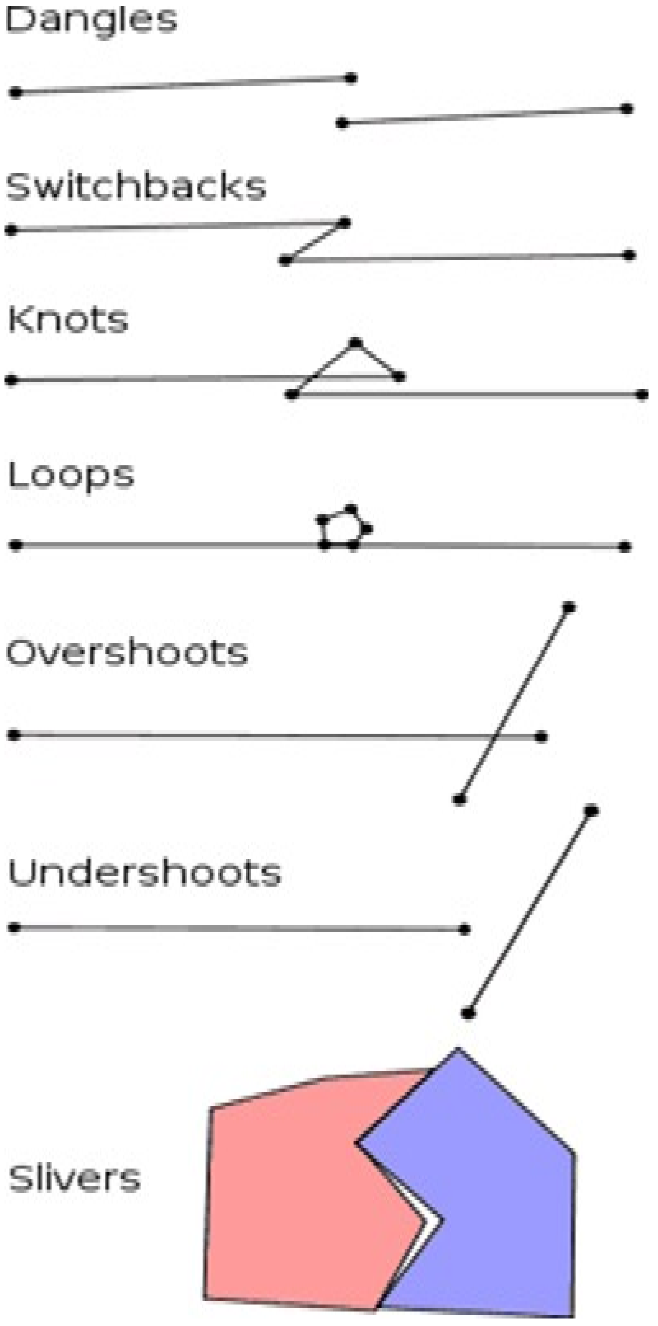
The process of Digitization involves ***topology rules and errors.***

Topological errors:

- **Dangles or dangling nodes** are lines that are not connected but should be connected. With dangling nodes, gaps occur in the linework where the two lines should be connected. Dangling nodes also occur when a digitized polygon doesn't connect back to itself, leaving a gap where the two end nodes should have connected, creating what is called an open polygon.
- **Switchbacks, Knots, and Loops:** These types of errors are introduced when the digitizer has an unsteady hand and moves the cursor or puck in such a way that the line being digitized ends up with extra vertices and/or nodes. In the case of switchbacks, extra vertices are introduced and the line ends up with a bend in it. With knots and loops, the line folds back onto itself, creating small polygon like geometry known as weird polygons.
- **Overshoots and Undershoots:** Similar to dangles, overshoots and undershoots happen when the line digitized doesn't connect properly with the neighbouring line it should intersect with. During digitization a snap tolerance is set by the digitizer. The snap

tolerance or snap distance is the measurement of the diameter extending from the point of the cursor. Any nodes of neighbouring lines that fall within the circle of the snap tolerance will result in the end points of the line being digitized automatically snapping to the nearest node. Undershoots and overshoots occur when the snap distance is either not set or is set too low for the scale being digitized. Conversely, if the snap distance is set too high and the line endpoint snaps to the wrong node. In a few cases, undershoots and overshoots are not actually errors. One instance would be the presence of cul-de-sacs (i.e., dead ends) within a road GIS database.

- **Slivers** are gaps in a digitized polygon layer where the adjoining polygons have gaps between them. Again, setting the proper parameters for snap tolerance is critical for ensuring that the edges of adjoining polygons snap together to eliminate those gaps. Where the two adjacent polygons overlap in error, the area where the two polygons overlap is called a sliver.



Source : https://www.gislounge.com/what-are-weird-polygons/

Topology rules

Fortunately, many common errors that can occur when digitising vector features can be prevented by topology rules that are implemented in many GIS applications. Except for some special GIS data formats, topology is usually not enforced by default. Many common GIS, like QGIS, define topology as relationship rules and let the user choose the rules, if any, to be implemented in a vector layer.

The following list shows some examples of where topology rules can be defined for real world features in a vector map:

- Area edges of a municipality map must not overlap.
- Area edges of a municipality map must not have gaps (slivers).
- Polygons showing property boundaries must be closed. Undershoots or overshoots of the border lines are not allowed.
- Contour lines in a vector line layer must not intersect (cross each other).

GIS data collection
Source : https://www.agiratech.com/introduction-to-topology-gis

Photogrammetry – vector data capture

Photogrammetry is the process of determining measurements from overlapping images, essentially creating 3D models by analysing multiple photos of a single area. This complex science is surprisingly affordable and can result in very detailed topographic visualizations.

One situation where vectorization may be necessary is when using an aerial photograph to create a property map. In this case, the process could look like the following:

Step 1. Scan the aerial photograph

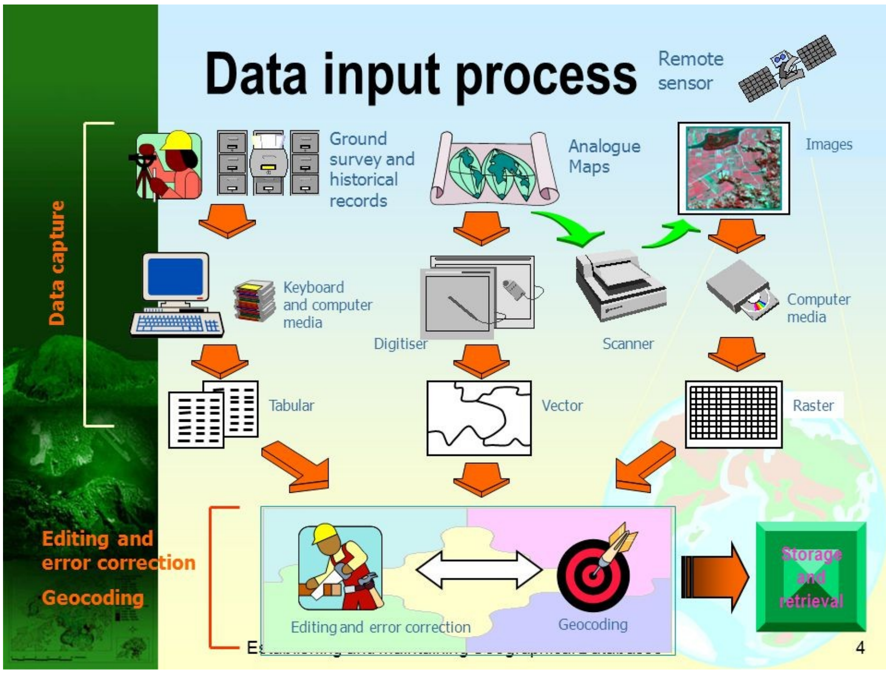
Step 2. Upload the scanned photo to GIS software

Step 3. Vectorize the photograph (converting it from pixels to polygons)

Step 4. Create a map layer with the vector data

Volunteered Geographic Information (VGI)

VGI refers to a new application through which, every single person with a smartphone, can share spatial data and upgrade the information we have for any place. This happens through social media sites. For example, when the user takes a picture, there is an automated question which asks the user, “where is that photo taken?” The user then can answer this question and improve the geodata. In this way, there is a bidirectional help because we can share, as well as collect data.



Source: Qiming Zhou, Geographical Information Systems Establishing and Maintaining Geographical Databases.

Choice of data acquisition methods

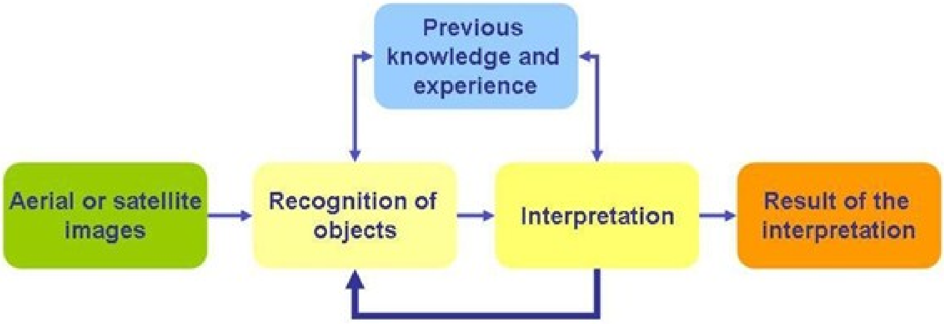
Data Sources	Method	Equipments	Accuracy	Cost
Analog Map	Manual Digitizing	Digitizer	±0.1 mm (on a map)	cheap
	Semi-automatic Scanning	Scanner	±0.1 mm (on a map)	high
Aerial Photos	Analytical Photogrammetry	Analog Stereo Plotter	±10 cm	high
	Digital Photogrammetry	Digital Phot. Workstation	±10 cm	very high
Satellite Images	Visual Interpretation	Image zoom scope	±30~50m	cheap
	Digital Image Processing	Image Processing System	±10~30m	high
Ground Survey	Field Measurement	Total station, GPS	±1 cm	very high
Reports	Keyboard Entry	Keyboard, PC		cheap

Source: Guy Picton Phillipps, Brent PCT, https://www.csp.org.uk/system/files/documents/2019-05/gis-basics-surfaces.pdf

Capturing attribute data

Attribute can be obtained from various sources including:

1. **Remote sensing image interpretation and processing**



Source: Remote Sensing Image Classification: A Comprehensive Review and Applications by Maryam Mehmood et al (March,20220), https://doi.org/10.1155/2022/5880959

2. Field investigations & Site Visits: field investigations are sometimes conducted using field questionnaires or other documents. In this case, thematic/attribute data can be collected by sampling method.

Source: https://www.shutterstock.com/image-photo/businessman-lost-field-using-map-23639869

3. 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.

Exercise materials and tasks

Quiz questions

You finished this session! Well done!

Now, please answer the following questions to test your knowledge:

1. Please complete the blank space with the correct term:

_________ (**Scanning**/Printing) is one of the simplest ways to convert physical documents into digital formats. When you scan a document, the scanner uploads a digital image of the page. Like all images, scanned documents will appear in _______ (vector/**raster**) format.

In what ways can we capture attribute data?

- **a. Field investigation**
- **b. direct data loggers**
- **c. Site visits**
- **d. Remote sensing image processing**

Photogrammetry is the process of taking reliable measurements from photographs.

- **a. True**
- b. False

3.5 Obtaining data from external sources

Reading material

Now that you know more about data capture, we will dive into data conversion and transfer.

For many projects, it would be nearly impossible to gather all the necessary data on your own. That is where external data sources come in. Regardless of where the data comes from, GIS software can overlay all the information into a single, layered map.

Now let us start with looking at Spatial data conversion and Spatial data transfer Standards (SDTS) as data management processes for obtaining external data.

I. Spatial data conversion

One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats. Data conversion is the process of translating data from one format to another. It is a critical process in the migration of information from existing information databases to new ones that often require changes in data formats. This step enables the data to be read, altered, and executed in an application or database other than that in which it was created, when the systems undergo replacement or updates. It is also of great importance in the insurance sector. Companies can make use of different strategies for converting data to ensure that the data is compatible with their systems.

Source : https://www.igismap.com/gis-data-conversion/

There are many ways to convert GIS data and the conversion may use special conversion programs. Simply, it may involve complex data exporting or importing procedures. Following are the most used spatial data format.

- The commonly used formats for raster data model are**: GeoTIFF (TIFF), ERDAS Imagine, Grid, ECW**
- The commonly used formats for the vector data model are**: .shp ,(shapefile), *.gpkg(Geopackage)** , SpatiaLite, **GeoJSON (a lightweight format based on JSON, used by many open source GIS packages), KML (Keyhole Markup Language a XML**

based), etc. Vector data can also be stored in database formats as opposed to file-based formats.

Right data conversion should ensure the following:

- Data is converted into an appropriate format and is transferred correctly.
- Data works in the new destination database.
- Data retains its quality and data consistency is maintained at all times across all the systems.

As said above one of the biggest challenges with spatial data obtained from external sources is that they can be encoded in many different formats. For that, many tools have been developed to move data between systems and to reuse data through open application programming interfaces (APIs).

GIS Consortium has the capability to convert hard copy into a wide range of electronic formats. GISC's team of experienced and multi – skilled specialists have a vast knowledge of converting various types of geospatial data. This helps to deal with the present-day challenges such as complexity of data, project timelines and effect on the quality and accessibility of the data, ensuring a smooth and successful data conversion.

What data conversion is NOT

Data conversion is often confused with processes known as ***data migration***, ***data transformation***, and ***data cleansing.*** Let's take a look to clarify these processes.

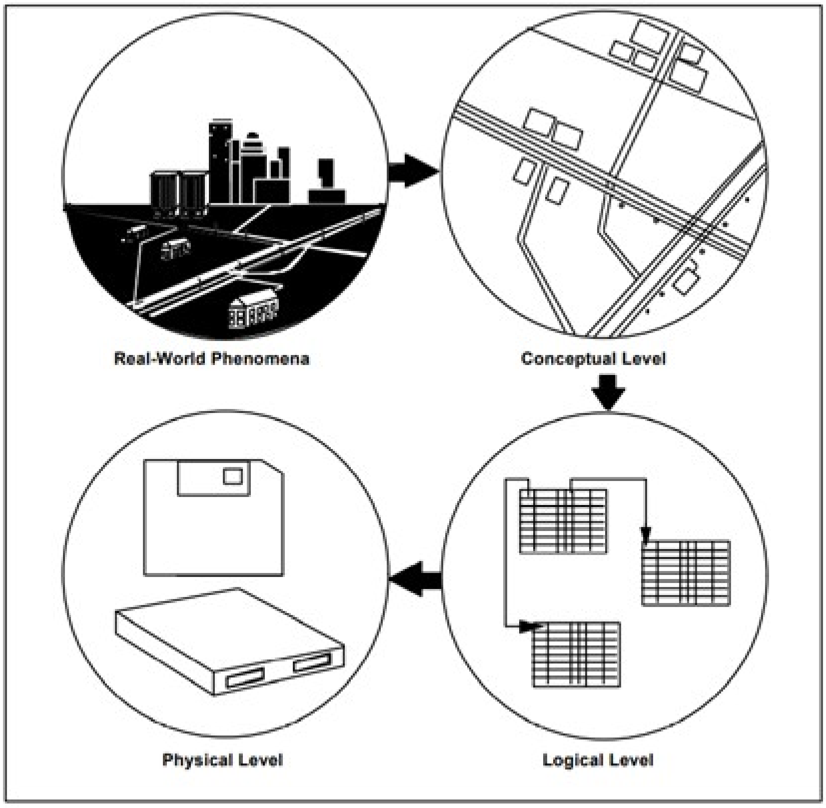
- **Data migration:** Where data conversion translates individual computer objects and data types from one format to another, data migration transfers entire databases or programs from one location to another. Data migration often entails data conversion, data transformation, and/or data cleansing.
- ***Data transformation:*** Data conversion translates one format to another. An example would be converting an ***.shp file*** to ***raster***. Data transformation changes the data presentation. A common data transformation process is to condense the data. Note: the format itself does not change.
- ***Data cleansing:*** Data cleansing finds and corrects inaccurate, repeated, and incomplete data. This procedure often occurs after a data conversion, data transformation, or data migration process.

II. Spatial data transfer Standards (SDTS)

Data transfer is the exchange of data and is essential to foster sharing of geographic data. The ***Spatial Data Transfer Standard (SDTS)*** is a robust way of transferring earth-referenced spatial data between different computer systems with the potential of no information loss. It is a general mechanism for the transfer of geographically referenced spatial data and its supporting metadata (i.e., attributes, data quality reports, coordinate reference systems, security information).

Data transfer standards are operating rules that accomplish the transfer of data between incompatible systems and provide a common format for exchanging data from various software systems. Data transfer standards provide a road map for interpreting various geographic data models. Since each GIS contains its own unique, conceptual, logical, and physical data model, standards can provide a means for exchanging data between two GIS systems. Resolving characteristics of various spatial data models used by the different systems is a fundamental requirement for data transfer standards.

The full SDTS specification creates a framework for spatial data transfer by defining different "levels," from the real world to the physical encoding of the data. The **conceptual level** describes a way to represent real world entities, including their geometric and topological characteristics and relationships. The **logical level** presents a data model for identifying and encoding information for an SDTS transfer. SDTS also defines the **physical level** with rules and specific formats for encoding data on a medium of choice (e.g., magnetic tape).



Source: "Design of a Spatial Data Transfer Processor." Cartography and Geographic Information Systems, Vol. 19, by Altheide, Phyllis

As usual, at an abstract level, every GIS packages supports at least one of the two spatial data models - ***the raster model and the vector model.*** In the raster model, the information is provided in the form of images (or pixels). In the vector model, the information is provided in the form of topology comprising of points, lines, and polygons. A data format should adhere to any one of the two data models.

Purpose of SDTS

The purpose of the SDTS is to promote and facilitate the transfer of digital spatial data between dissimilar computer systems, while preserving information meaning and minimizing the need for information external to the transfer. Implementation of SDTS is of significant interest to users and producers of digital spatial data because of the potential for increased access to and sharing of spatial data, the reduction of information loss in data exchange, the elimination of the duplication of data acquisition, and the increase in the quality and integrity of spatial data. SDTS is neutral, modular, growth-oriented, extensible, and flexible--all characteristics of an "open systems" standard.

Interoperability and Spatial Data Standards

The International Organization for Standardization defines interoperability as “the capacity to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units” (ISO 1993). In the context of geographic information, the functional units are Geographic information (GI) systems or Web services. The communication between these units comprises transfer of spatial data as well as querying and execution of remote services.

In this respect, two types of interoperability can be distinguished: ***syntactic interoperability*** refers to format of the transferred data that is the compliance with spatial data standards; ***semantic interoperability*** builds on syntactic interoperability and refers to the accurate preservation and interpretation of the meaning of the transferred information.

Interoperability is a key requirement for the seamless exchange of spatial data between users and organizations employing different Geographic information systems and Web services. On a broader scale, the standards enabling interoperability are the cornerstones of Spatial Data Infrastructure. Without agreements on the formats of the transferred spatial data and the interfaces for accessing the corresponding Web services, mutual data exchange between different geographic information systems would at least require manual transformation of the data or even be impossible.

Source: Encyclopedia of Geography by Wright, John Kirtland (1891–1969), http://dx.doi.org/10.4135/9781412939591.n1264 , ©2010 SAGE Publications

Geospatial information, a variant of spatial information, is generally collected on thematic basis, where individual organizations are involved on any particular theme. Geospatial thematic data is being collected from decades and huge amount of data is available in different organizations. Information communities find it difficult to locate and retrieve required geospatial information from other geospatial sources in reliable and acceptable form.

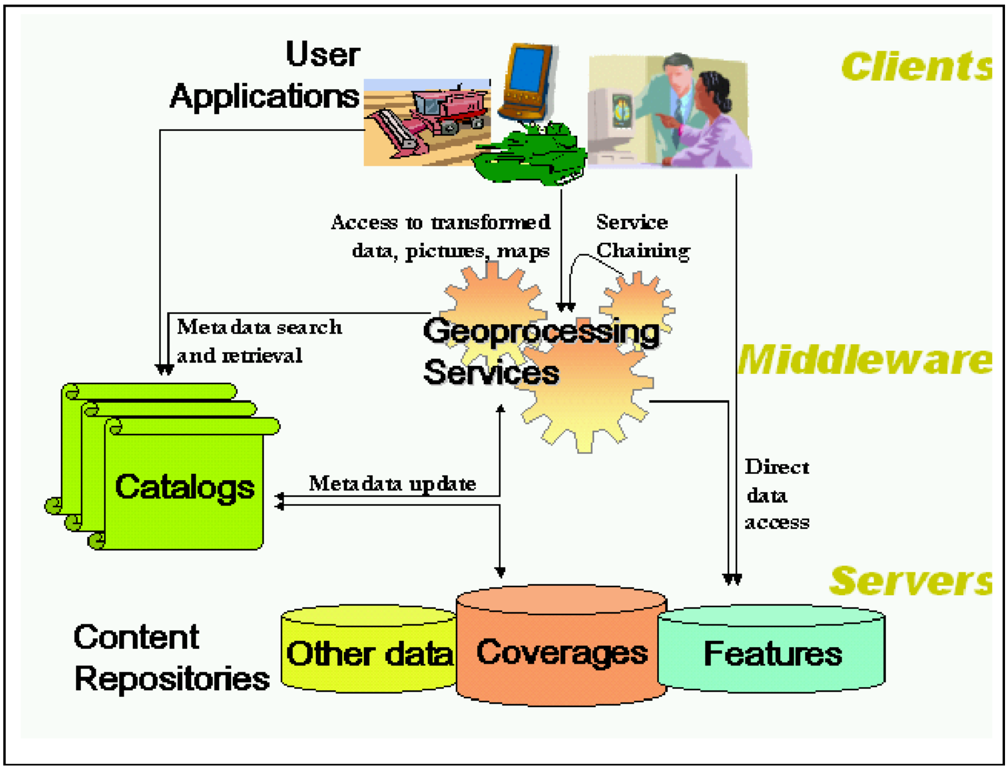
The problem that has been incurred is the lack of standards in geospatial data formats and storage/access mechanism. Heterogeneity in geospatial data formats and access methods poses a major challenge for geospatial information sharing among a larger user community.

With the growing need of geospatial information and widespread use of Internet has fostered the requirement of geospatial information sharing over the Web. The Geo-Web is one of the reliable distributed networks of interconnected geographic information sources and processing services that are:

- Globally accessible, that is, they live on the internet and are accessed through standard Open Geospatial Consortium (OGC) and W3C interfaces

- Globally integrated data sources that make use of standard data representation for sharing and transporting geospatial data.

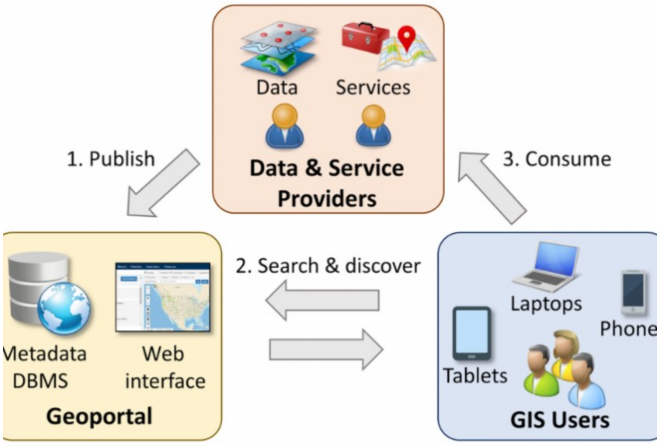
Without successful interoperability approaches, the realization of Geo-Web is not possible.



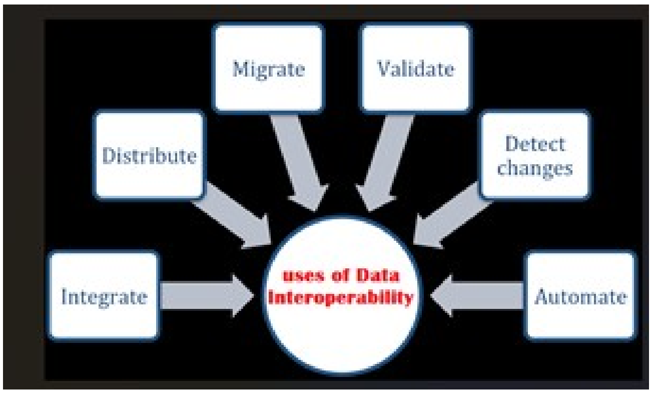
Source: Geospatial Standards, Interoperability, Metadata Semantics and Spatial Data Infrastructure by R. Longhorn (2005)

Traits of successful interoperability:

- Data producers must ensure that their data are readily accessible and understandable to potential data consumers.
- Users must be able to identify and locate relevant information and know whether a given set of data is germane to their work.
- Queries to dispersed sites must be formulated and processed in a manner meaningful to both data server and application client.
- Geodata from one source must be capable of being integrated with data from another, in terms of both structure and semantics.
- Display and analytical functions must be associated with particular data models and made available to the requester.



Source: University of consortium of Geographic Information science



Source: Muvunyi Germain, INES Ruhengeri, Rwanda (Lecture note)

Exercise materials and tasks

Quiz questions

As a recap of this session's contents, please answer the following questions to test your understanding:

1. The biggest problems with data obtained from external sources is that they are in many different formats. Data transformation is the process of translating data from one format to another.
   a. True
   b. False

2. Link these two types of interoperability with their definition by matching the following terms with the correct definition:
   syntactic interoperability, semantic interoperability

refers to the accurate preservation and interpretation of the meaning of the transferred information	
refers to format of the transferred data that is the compliance with spatial data standards	

Answer:

refers to the accurate preservation and interpretation of the meaning of the transferred information	**semantic interoperability**
refers to format of the transferred data that is the compliance with spatial data standards	**syntactic interoperability**

3. Pick the right traits of successful interoperability
   a. Data producer must ensure that the data is accessible
   b. Geodata cannot be integrated with other's data
   c. Users must be able to locate relevant information
   d. Data does not have to be understandable

3.6 Fieldwork and application

Exercise materials and tasks

Application Exercise

Congratulations on completing the theory part of this module!

Now let's continue with the practical part of the module. In this practical application of GIS data collection, you will be asked to go to the field and learn how to collect simple data using a mobile app. The exercise will allow you to explore field data collection and become familiar with data acquisition using accessible GPS techniques.

Field work and Application exercises

This exercise consists of three parts:

1) GIS data collection in the field work
2) Data analysis on your computer.
3) Application report summarizing your findings.

Please familiarize yourself with the report requirements prior to starting your field work.

Part 1: Fieldwork – GIS data collection

To collect data in the field we suggest using the SW Maps mobile app. Before any further instructions, please download SW Maps on your smart device (SW Maps - GIS & Data Collector – Aplikacije v Googlu Play (google.com)).

SW Maps is a free GIS and mobile mapping app for collecting, presenting and sharing geographic information. Whether you are conducting a full scale GNSS survey with high precision instruments, need to collect large amount of location-based data using nothing but your phone, or just upload shared shapefile on your phone in the app, view shapefiles with labels over a background map on the go, track a new point, polyline or draw a polygon. The surveyed output can be shared in different GIS formats that can read in EO application to monitor landscape change overtime.

As a help with the use of the application and additional ambiguities, the video shows how the application is used and how you will use it to collect data - https://www.youtube.com/watch?v=bNRmrhdmujU. You are now ready to start the fieldwork:

1. Please select an area near your home. You are going to collect data using the SW Mapp application on your smart device.
2. Using the application, please generate:
    - 5 reference points
    - 3 reference lines (at least 100m long)
    - 2 polygons (bigger house and agriculture field) close to your home.

Part 2: Data analysis

After finishing data collection, go back to your computer and download the data. To find out how to download data, use the link of the video from above in the fieldwork part of the exercise.

Steps:

1. Open QGIS and start a new project
2. Under data Source Manager > Vector add exported layers (*.shp) from downloaded data
3. Try to use Google map or OpenStreetMap as background to check on your data (Hint: XYZ Tiles or WMS)

Part 3: Application report

Done? Now you just need to draft your report!

Forum instructions

Generate report of the application exercise and share your results in the forum

Welcome to the forum of Module 3!

As a last step, please generate a report about your findings and share your results in the dedicated forum.

The report you shall post in the forum should contain the following information:

1. Make a screenshot of data in QGIS
2. Include a brief description of the selected area – what reference data you digitized
3. Discuss obtained results of accuracy. In order to do so, here are some questions that will help you:
    - Do your lines match with the lines on maps?
    - Why is it that the corners of the house do not match with yours?
    - How accurate are you (estimation)?

Once you have finalised, please read the contributions of the other group members. Post at least one comment or question to another participant's contribution with the idea of exchanging experiences.

Please do not forget to answer any question you got in response to your post in the country forum.