Advanced GIS

Course designers:

Nguyen Thi Nhat Thanh: *University of Engineering and Technology – Vietnam National University Hanoi, Vietnam*
Dominque Laffly: *University of Toulouse, France*

1. INTRODUCTION

The introductory course on GIS and remote sensing clarified the already ancient historical context. GIS were an information system or a database management system for vector data attributes with the Canadian Geographical System (CGIS). The phrase, "geographic information system", was coined by Roger TOMLINSON in 1963, when he published the scientific paper, "A Geographic Information System for Regional Planning". Ultimately, TOMLINSON created the framework for a database that could store and analyze huge amounts of data and leaded to the Canadian government being able to implement its National Land-Use Management Program (*Wikipedia*).

The principle of GIS is based on the topological consistency of layer geometry and their georeferencing precision, which is why GIS develops all their intrinsic qualities. The first thing to remember is geographical datum level. It may be fine but guarantees geometric precision as shown in Figure 1. Using an image with high spatial resolution, we can observe at ground level the facades of buildings whereas the cartographic projection only visualizes the detour of the footprint.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

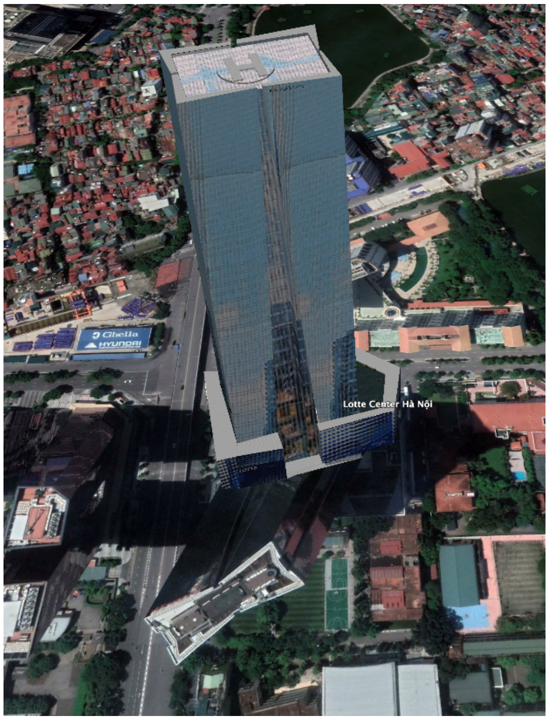

Figure 1. On the left in vertical view, the 3D vector representation is in good projection. On the right, same scene in cavalier view (Source: Google Earth)

Even for symbolic data such as vector data - a vector is a more or less precise drawing providing thematic information - the notion of precision remains unclear. Figure 2 shows on Google Maps another example of the difference between the route of a river in relation to the relief and the border between Vietnam and China: either the river is in wrong place, or it is the problem of the relief and border? Figure 3 shows the border route differences from Plans (Apple), Google Maps and GADM (https://gadm.org) at the same area.

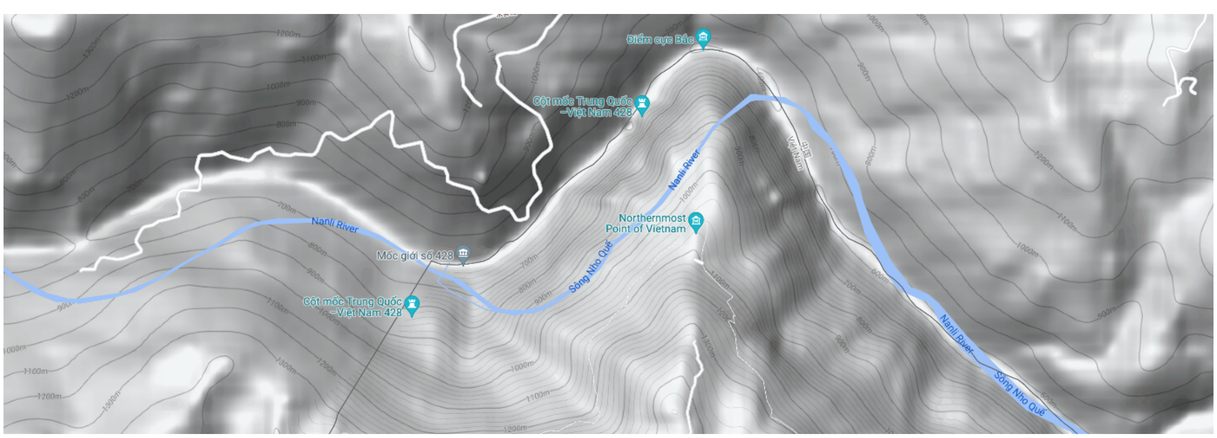

Figure 2. Northernmost point of Vietnam: 23.383771166438247 ° North, 105.32253902840087 ° East.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

Note the absurdity of the geographic coordinates given to within 10^{-15}, or on the order of a tenth of a nanometer! While the river is several hundred meters from its bed in the middle of the slope.

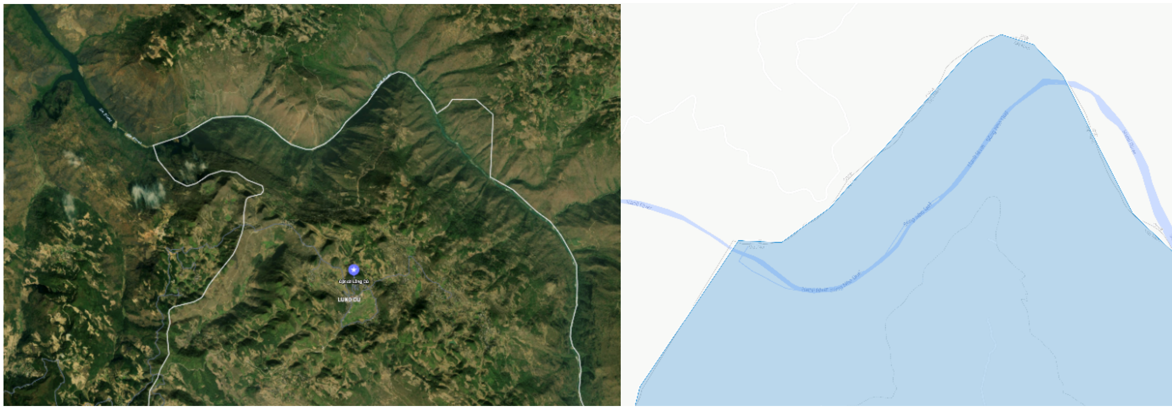

Figure 3. On the left Maps which give additional indentations at the border compared to Google Maps (gray) and GADM (blue) for which the limits are almost identical (but a river outside its course).

The GIS data layer has been collected and processed in different ways. Each data layer has different errors/uncertainties depending on its own characteristics. Data could be collected directly in the field, which is referred to as in situ or in-place data collection, or from distance using innovation technology, called to as remote sensing. Scientists and/or experts will go to different sites, do surveys, collect samples, make measurements of objects and phenomena of their interest, such as people, sex, income, living conditions, atmospheric conditions, soil moisture, etc., using survey form, transducers or other in situ equipment. However, in situ data could have errors as results of human intervention to objects/phenomena (referred to as intrusive errors), biased procedures or measurement device miscalibration. In situ data is often considered as the ground truth but it still has uncertainties and errors.

Remote sensing data is collected in another way. (RN Colwell, 1997) stated the definition of remote sensing as follow "Photogrammetry and remote sensing are the art, science, and technology of obtaining reliable information about physical objects and the environment, through the process of recording, measuring and interpreting imagery and digital representations of energy patterns derived from non-contact sensor systems". Sensors on-board satellite and aerial vehicles are used to observe objects from space. The data collected for Earth resource applications is the result of sensors that record electromagnetic energy. Therefore, methods to obtain meaningful information on the environment like estimate area, types, characteristics, and health of vegetation, water, atmosphere, urban, … need various computation techniques as well as extensive knowledge in physics, biology, and society. Figure 4 present relation of remote sensing, GIS, surveying, and cartography in research and application.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

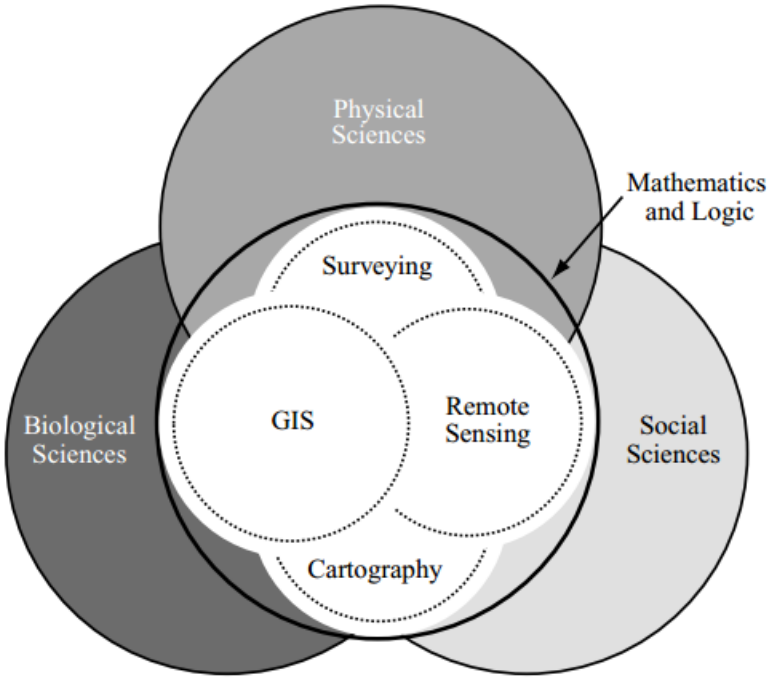

Figure 4. The relationship of the remote sensing, geographic information systems, remote sensing, cartography, and surveying with other sciences (J. R. Jensen, 2014)

Remote sensing data provide thematic information on a very large geographic areas rather than just single-point observations. Moreover, a lot of significant information can be extracted from climate modelling (precipitation, air pollution, temperature, humidity, etc.), from people culture/activities (e.g., land cover, land use change, urbanization, …) or digital elevation. However, remote sensing data has one limitation in that it only provides certain information based on interaction between object/event and satellite signals. Remote sensing also contains measurement errors that can be fixed with calibrations. The active remote sensing can retrieve more information but can also have certain effects on the observed objects/phenomena and is costly in terms of data measurement and processing. Figure 5 present popular applications using satellite images in constraints of observation frequency and spatial range.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

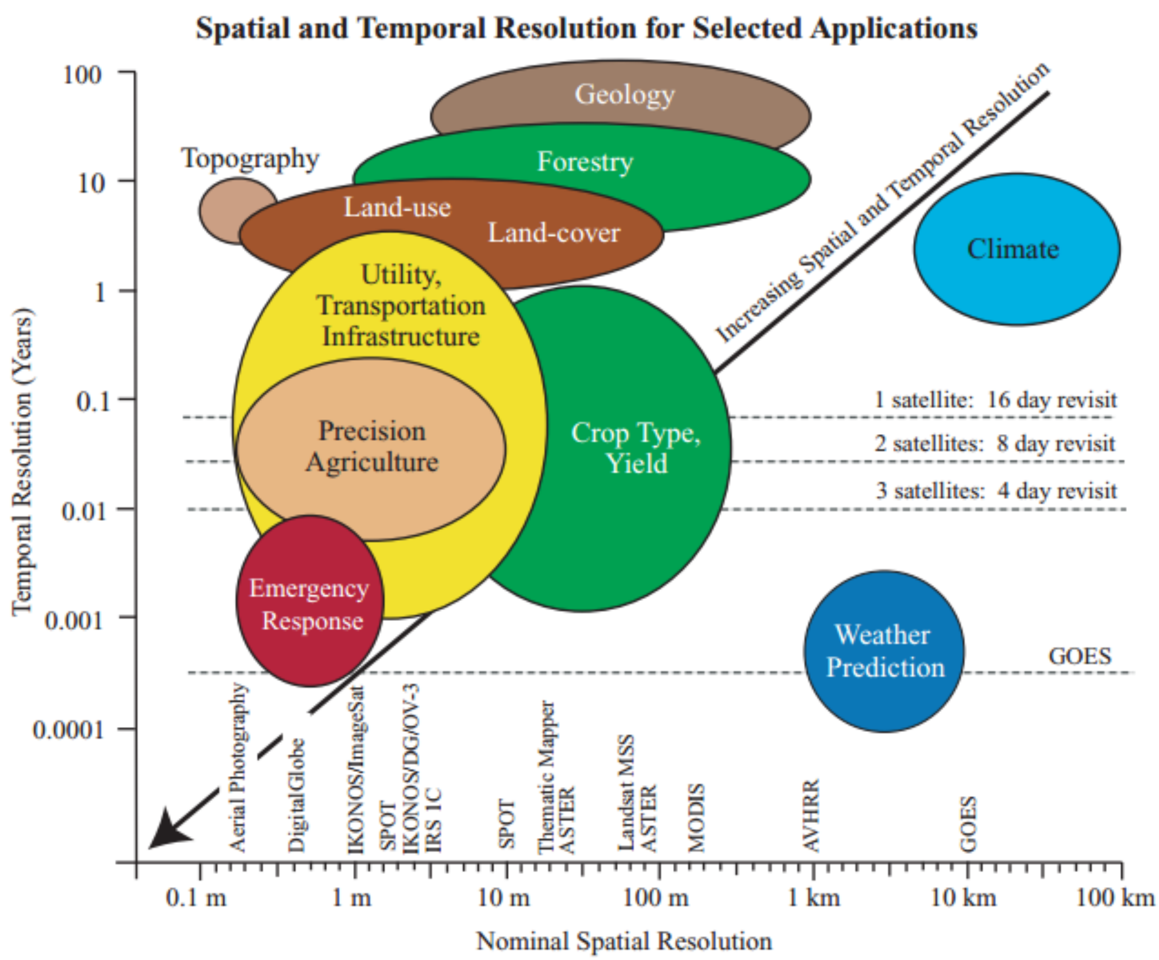

Figure 5. The nominal spatial and temporal resolution for selected application (J. R. Jensen, 2014)

What if the "advanced level" of Remote Sensing and GIS in this course is already being data-critical, doubting data by default, always checking data? This apparent skepticism is a fundamental value of science and of research methodology. It is aimed at drastically reducing uncertainties to minimize risks and allow actions. The COVID-19 pandemic is a perfect illustration of decision making based on data. Decision makers in countries around the world have had to act on the reductions in uncertainty that science is trying to create such as wear a mask or not; full or targeted containment; vaccination by age group and/or geographic area; health pass and access regulations to public places…

What if the "advanced level" of Remote Sensing and GIS that we introduce in this course is processing methodologies for remote sensing data to obtain significant information of vegetation, water, atmosphere, and urban landscape?

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

What if the "advanced level" of Remote Sensing and GIS that we are discussing in this course were the promotion of remote sensing and geographical information processing flows for scientific research and then their operationalities for decision-makers? Charles Picquet and John Snow[1] did just that, in 1832 and 1854 respectively, to understand cholera epidemics, their sources and methods of distribution as well as risk factors by comparing data on population density, the location of patients and water supply points.

The "advanced level" of remote sensing and GIS addressing in this course is then the ability to think of geographic space - not to master software although this is necessary - as a complex whole. A holistic thinking of the geographical space where economic, social, societal, historical, cultural, political forces as well as geological, climatic, biological forces interact ... It is obviously extremely complicated to approach the complexity. The students have to master the tools, the data and the problem all at the same time. Let's be honest, with very few exceptions - who else? - it will be a transdisciplinary teamwork, that is to say a team which does not content itself with collecting the disciplines around a project - which is then said to be multidisciplinary, as one would collect the layers in a GIS, but which releases an innovative added value from the curiosities, respects and skills of the disciplines involved.

2. ADVANCED REMOTE SENSING

a. Optical remote sensing for Vegetation

Vegetation is an important component of the ecosystem. The knowledge of vegetation is vast in many aspects, from species and community, distribution patterns, growth cycles, to the modification in the plant physiology and morphology. One of the branches in remote sensing is to research and develop sensors and algorithms to calculate vegetation indices from the sensors' data. Popular applications of this include agricultural, forest, mangrove soil, urban vegetation surveillance, etc.

This lecture introduces the basic concepts of vegetation characteristics, their connection to remote sensing data, processing techniques to obtain information on a vegetation area such as the photosynthesis process, spectral characteristics, natural morphological cycles. Next, vegetation indices, which are used to extract biological information on the vegetation layer from sensors' data including leaf-area-index (LAI), percentage green cover, chlorophyll content, etc., will be introduced. Different metrics like height, size, etc., are also mentioned as useful information about the vegetation layer to be extracted. Landscape surveillance can be done by observing the variation of indices such as land-cover composition and pattern, riparian extent and distribution, ground water, greenness pattern, degree of biophysical constraints, erosion potential in both space and

[1] See the introductory course on GIS and remote sensing.

time (Jones et al., 1998). Estimations of these variations are necessary to determine the stability of the current landscape, based on the different models and its past condition (M. E. Jensen & Everett, 1993). The application of the principles of landscape ecology requires an understanding of the natural variability of landscape patterns and processes in both space and time. Estimates of this variability are essential to determine whether the present condition of the landscape is sustainable, based on its historical patterns and changes (M. E. Jensen & Everett, 1993). Furthermore, these estimates are extremely useful in both broad-based assessments of risk to resources, as well as smaller-scale assessments.

Several applications provided in this lecture include the application of remote sensing in rice cultivation area surveillance, land cover classification, observation of the land use changes in Vietnam.

b. Optical remote sensing for water and atmosphere

Water accounts for 74% of the Earth's surface, in which 97% are found in the vast salty oceans and only 0,02% are found in the rivers, lakes and freshwater reservoirs. The rest either resides inside deep pockets in the Earth's crust or in the Earth's atmosphere and surface as vapor or ice caps. Water exists in many forms on Earth, including fresh water, salt water, vapor, rain, snow and ice. Measuring, monitoring and predicting the spatial distribution, volume and motion off water throughout the hydrologic cycle is of great importance for the use and adaption of this precious resource.

Various indices related to water such as precipitation, water depth, temperature, salinity, velocity, volume, etc., can be measured using on-ground measuring equipment (in situ measurement) at a given location. This type of observation is quite essential but still inadequate in reflecting the current state of an area. Even with interpolation techniques to create distribution maps from on-ground station data, the number of observation station is insufficient for the results to be reliable. Moreover, it is rather hard to obtain the spatial distribution of some important hydrologic indices such as water-surface area (streams, rivers, ponds, lakes, reservoirs, and seas), water constituents (organic and inorganic), water depth (bathymetry), water-surface temperature, cloud cover, precipitation, water vapor, etc. Therefore, remote sensing images and calculation methods are researched and developed in order to obtain quantitative, spatial measurements of these important hydrologic variables.

This lecture will introduce basic principles of remote sensing for water surfaces and its various components: clouds, water vapor, precipitation and snow, as well as methods of analyzing remote sensing images for inventory and surveillance of the spatial distribution of inorganic/organic substances, depth and temperature. Definition of water surfaces, subsurface water volume and bottom radiation will be introduced. Next, the spectral response of water as a wave function will be mentioned to determine the purity of water. For water containing inorganic or organic substances, the spectral response of water as a function of inorganic/organic composition will be used to observe suspended minerals in the water (turbidity), chlorophyll and soluble organic substances. In addition, the temperature of inland water bodies and the sea-surface temperature (SST) in the daytime and nighttime can be calculated using infrared remote sensing techniques. Regular water regions usually have a high consistency (thermal inertia), meaning there can only a few degrees difference between daytime and nighttime temperature.

Precipitation is one of the essential atmospheric indices. This lecture will introduce different methods of initial estimation based on the brightness of reflected sunlight hypothesis in the visible and infrared spectrum from clouds to indicate its thickness and the possibility of rain. Next, information on the cloud top temperature is used to estimate rainfall with the hypothesis that colder cloud tops equal higher chances of rain. One algorithm based on this principle has been developed and applied to the polar-orbiting infrared sensors onboard AVHRR to estimate monthly rainfall and determine GOES Precipitation Index (GPI) based on an analysis of thermal infrared data. The GPI was the standard climatological rainfall product that provided a climate-scale precipitation record throughout the tropics and subtropics.

One crucial innovation in the area of precipitation estimation is the development of the microwave remote sensing instruments, which can respond directly in the presence of water/ice in clouds. The first ever implementation was the Special Sensor Microwave Imager (SSM/1), launched in 1987, capable of calculating the precipitation on land with a spatial resolution of 15x15km on a global scale. The Tropical Rainfall Measuring Mission (TRMM), launched on November 27, 1997, by NASA and the National Space Development Agency (NASDA) of Japan, is another important milestone in the evolution of precipitation estimation, especially for tropical and subtropical regions.

Aerosol particles may be solid or liquid and range in size from 0.01 μm to several tens of micrometers. Aerosols tend to cool the surface below them because most aerosols reflect Sunlight back to space, reducing the amount of solar radiation that can be absorbed by the surface below. Aerosols have a tight connection to air pollution, specifically to Particulate Matter concentration on ground. Algorithms to obtain information on amount and distribution of aerosol (called optical depth) and particle properties based on the Dark Target or Deep Blue approach are used to construct global aerosol products from the MODIS instrument on board the Terra and Aqua are currently widely used.

Clouds are one of the main drivers in the regulation of Earth's climate due to its effect on the energy flow in the Earth's atmosphere. Clouds can heat or cold the Earth depending upon its thickness and height above the surface. There are a variety of cloud types, with differences in height, humidity and location and it is extremely difficult to record their effects on global climate. One of the few instruments to take on the challenge is the first ever meteorological satellite GOES, which provides data in both the visible and thermal infrared portion of the spectrum. During the daytime, images from the sensor can give information on clouds from above, which is similar to naked eye observations. In the nighttime, infrared sensors can record cloud patterns and other important meteorological events at night.

Water vapor is an essential contributor to the formation of rain and an important greenhouse gas, playing a significant role in the understanding and predicting of most weather process from convection to the formation of clouds and rainfall and huge storm. Vapor maps can be constructed from remote sensing data using a collection of wavelengths; the most prominent being 6.7 µm. This rule of producing vapor maps was issued in 1978 by the Europe Space Agency and GOES and used by the METEOSAT 1. Information on water vapor can also be extracted from spectral data at the near-infrared portion. Specifically, in the wavelength range of 890 to 990 nm, there are 3 sub-ranges that mainly correlate to the absorption of water vapor: the strongest wavelength that centers near 942 nm and weaker wavelengths around 906 and 977nm. The MODIS instrument have a number of bands of wavelengths that are sensitive to vapor in the atmosphere in these portions of the spectrum, namely band 17 (890 - 920 nm), 18 (931 - 941 nm) và 19 (915 - 965 nm).

c. Optical remote sensing for urban landscape

The urban landscape comprises multiple diverted components like plants, crops, houses, roads, factories, industrial compounds, etc., which provide the goods and services to accommodate human life. In modern days, the process of urbanization is happening at a staggering rate, without proper management and planning, which gives rise to a plethora of negative impacts on the environment and quality of life. Information on urban areas is collected and extracted from a variety of sources like non-image-based sources, sensors, surveys, etc. Most ideally, this information needs to be stored in a national spatial data infrastructure (NSDI), which can be accessed freely by users of the public and private sectors. The NSDI consists of three framework foundation spatial databases (Geodetic control, Digital terrain (elevation and bathymetry), Digital ortho-rectified imagery), four framework thematic databases (cadastral, boundaries or political units, hydrology, transportation) and several other the thematic databases, such as urban/suburban land-use and land cover.

This lecture will consider the process of using remote sensing data to gather information on the specific characteristics of urban/suburban areas with regards to suitable resolutions of time, optical spectrum and space. Land cover and land use represents the biophysical materials that are present on land and their usefulness towards humans. This information is necessary for lots of applications including choosing residential/industrial/commercial locations, population estimation, tax evaluation, construction of zonal regulations, etc.

Information on land use and land cover need to be organized according to a standardized land-use or land-cover classification scheme. The most comprehensive hierarchical classification system for urban/suburban land-use is the Land-Based Classification Standard (LBCS) developed by the American Planning Association (2006). The U.S. Geological Survey Land-Use/Land-Cover Classification System was specifically designed to be resource-oriented (land cover). There is a relationship between USGS Land-use/Land-cover Classification System and the spatial resolution of the remote sensing system (often referred to as ground-resolved distance in meters), where Land Cover classification levels of I, II, III, IV are corresponding to spatial resolution from 20 – 100 m, 5 – 20 m, 1 – 5 m, and < 1m.

In the field of home management, inventory of location, type, condition and quantity are important tasks, which are performed using remote sensing data. The distinction between land use purposes of grade II for residential, commercial, industrial and traffic areas require a sensor system with a spatial resolution of 5 - 20 m. The identification of land use types at level III requires sensors with a spatial resolution of 1 - 5m. In addition, socioeconomic attributes can also be extracted directly or indirectly from the image, such as population and quality of life indicators. Trade and service areas can also be identified from remote sensing images when experts have a good understanding of the cultural characteristics of the country of interest and its pattern. These regions are defined as the specific set of materials and structures that characterize certain activities in a particular culture.

Information on industrial land can also be extracted from high-resolution remote sensing images. Rules are discussed to allow interpreters to define what the industry is. The term industry includes those enterprises engaged in the extraction of raw materials, the processing of raw materials, and the manufacture of intermediate and finished products. Industries often have unique combinations of raw materials, equipment, products, waste and characteristic buildings. These components can be seen directly or inferred from captured images. It can be classified into three major industries: Mining, processing, and manufacturing. When classifying an industry using remote sensing data, it is advisable to 1) Analyze first to determine whether it is a mining, processing, or manufacturing industry; 2) If it's a process industry, decide if it's chemical, thermal, or mechanical, in that order; and 3) If it's a manufacturing industry, decide if it's light or heavy manufacturing. The industries are so numerous that only a few representative examples of each major industry can be provided.

Transit planners often use remote sensor data to update transit network maps, assess railway networks and rail conditions, and study urban traffic patterns at points of congestion such as roads, tunnels, bridges, shopping malls and airports, and parking lots (Hauck et al., 1997). The most comprehensive work to date on remote sensing applied to transportation problems has been carried out by the NASA-funded National Consortium for Remote Sensing in Transportation and the Department of Transportation (NCRST, 2006). (Thirumalai, 2003) evaluated the NCRST RESEARCH conducted by consortium members on: a) Environmental Impacts for Multimodal Corridor Impact Planning, b) Risk and Disaster Preparedness and Safety plasma security, c) Infrastructure and security asset management, and d) Multimodal transport flow management.

Urban/suburban environments are huge consumers of electricity, natural gas, phone service, and mobile water and generate large amounts of refuse, sewage, and wastewater. Automated mapping/facilities management (AM/FM) and geographic information systems (GIS) have been developed to manage extensive right-of-way corridors for various utilities, especially pipelines. Other categories for cities include urban digital elevation modeling, meteorological data, urban hydrology, disaster emergency response, observations.

3. ADVANCED GIS

a. Modeling concept and modeling process

A model captures a view of a physical system. It is an abstraction of the physical system, with a certain purpose. This purpose determines what is to be included in the model and what is irrelevant. Thus, the model completely describes those aspects of the physical system that are relevant to the purpose of the model, at the relevant level of detail.

The Modeling Process has four steps. The first step is to define the goals of a model. The second step is to break down model into elements and to define properties of each element and interactions between the elements. A flowchart is a useful tool for linking the elements. The third step is the implementation and calibration of the model. Finally, the fourth step is to validate the model before it can be generally accepted.

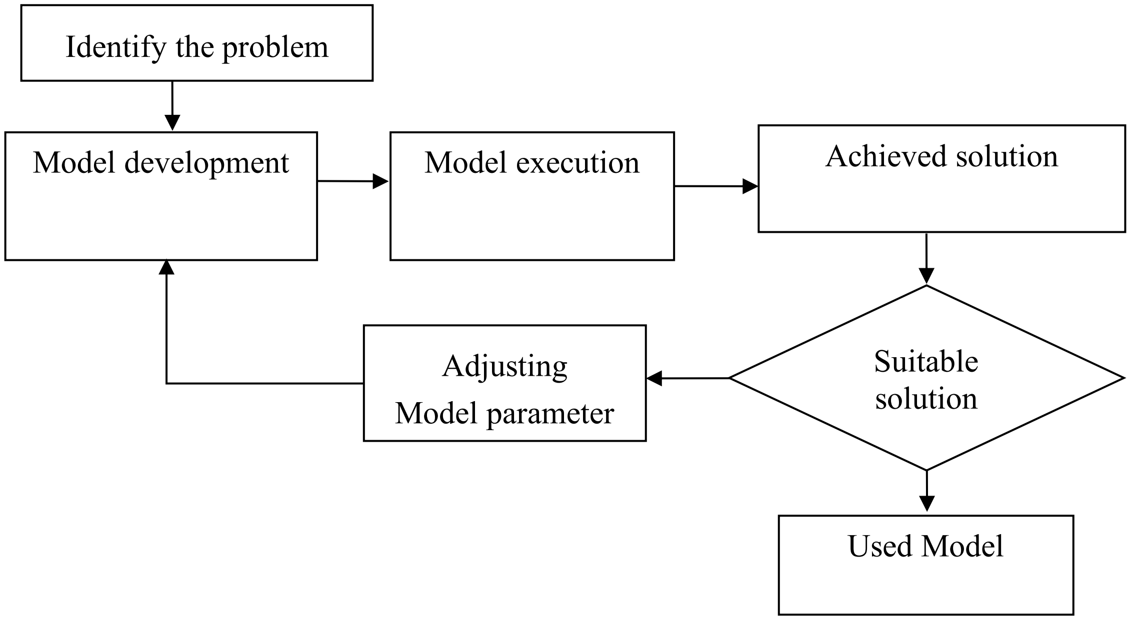

Figure 7. The modelling processes

Depending on the degree of complexity, GIS modeling may take place in a GIS or require the linking of a GIS to other computer programs. In general, binary models and index models belong to the former, and regression models and process models are the latter. A statistical analysis package can better accommodate regression analysis, which is the core of regression models, than a GIS. In the case of process models, a GIS is often a tool for database management, data visualization, and spatial analysis while simulation models embedded in other computer programs work with complex and dynamic analysis. A GIS can be linked to other computer programs using the following three strategies. A loose coupling involves data transfer from one to another, a tight coupling provides a common user interface to the GIS and other computer programs, and an embedded system bundles the GIS and other computer programs with shared memory and a common interface. Most GIS applications described in the literature are based on the first two strategies.

A cartographic model is an integrated sequence of data processing tasks that organize, combine, analyze and display information to answer a question. Cartographic modeling is effective in GIS environments because they rely heavily upon visualization, making it easy to show input and output layers in map form (Le, 2017, 2019). In many GIS platforms, the sequence of tasks can be created and modified graphically as well.

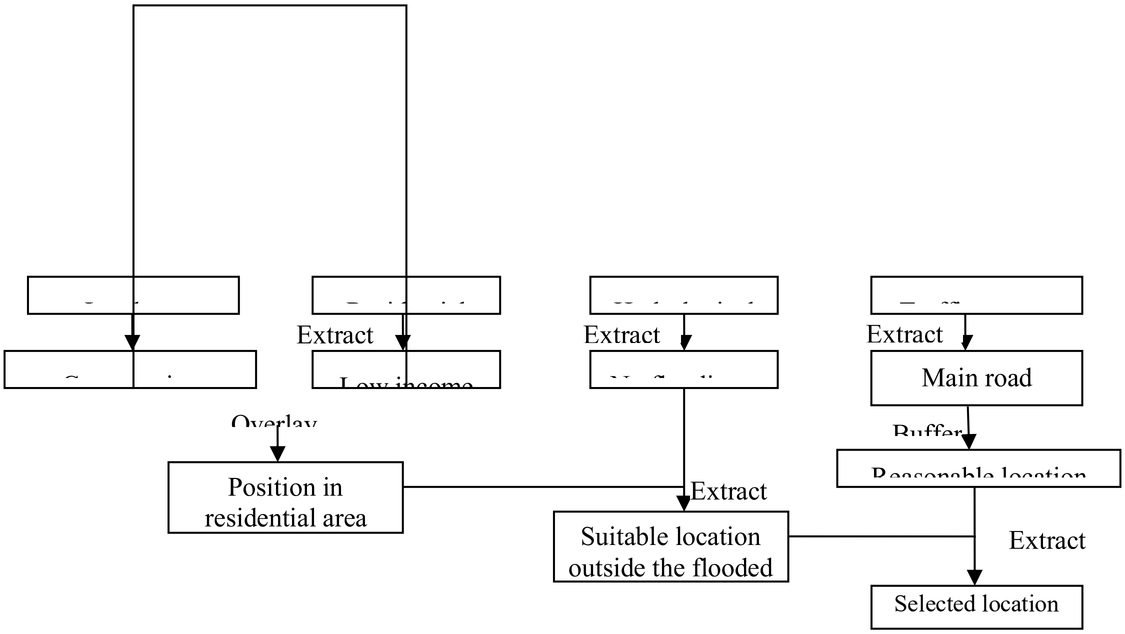

Figure 8. Select location for building school

A binary model uses logical expressions to select map features from a composite map. The output of a binary model is in binary format: 1 (True) for map features that satisfy the logical expressions and 0 (False) for map features that do not. A common application of binary models is siting analysis, with each logical expression corresponding to a siting criterion. Suppose a county government wants to select potential industrial sites based on the following criteria.

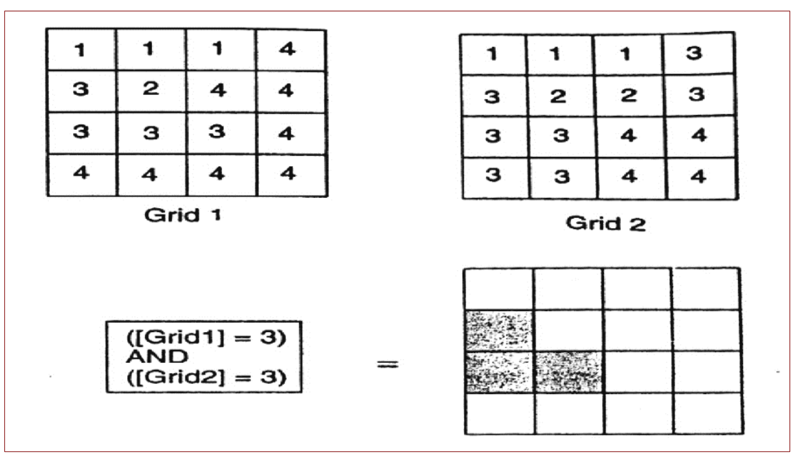

Figure 9. Grids

An index model calculates the index value for each unit area and produces a ranked map based on the index values. An index model is similar to a binary model in which both involve multicriteria evaluation and depend on overlay operations for data processing. But an index model produces for each unit area an index value rather than a simple yes or no.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

Regression model relates a dependent variable to a number of independent (explanatory) variables in an equation, which can then be used for prediction or estimation. A regression model can use overlay operations in a GIS to combine variables needed for the analysis. There are two types of regression model: linear regression and logistic regression.

b. Spatial Database and Data Warehouse

Spatial data is associated with geographic locations such as cities, towns etc. A spatial database is optimized to store and query data representing objects. These are the objects which are defined in a geometric space.

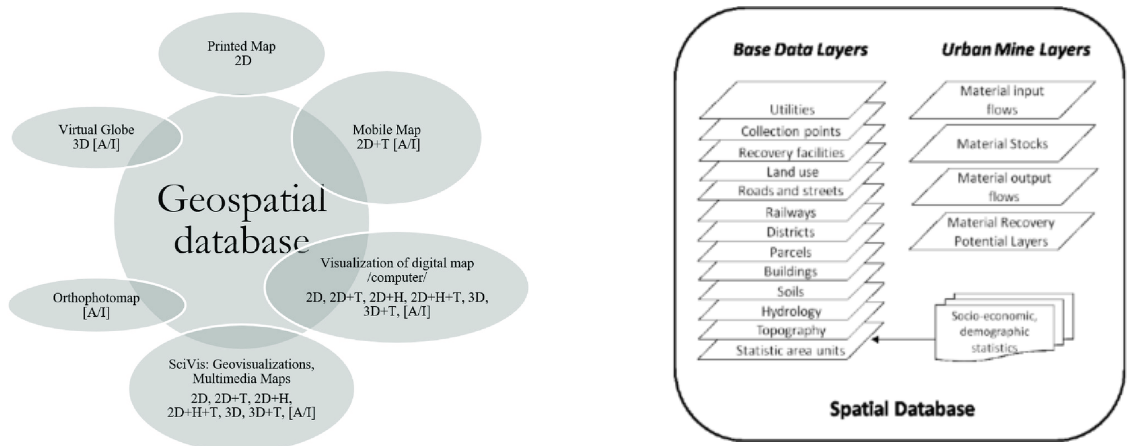

Figure 10. Spatial database

We cannot calculate how much data nowadays is electronically available. Neither can we imagine how much it has cost to gather this data and to feed the systems. Almost all data is captured and used in special purpose systems, so called operational systems. Data in these systems is being modelled for these systems in a specific format. This is done just on behalf of the specific transactional system. Now this data is here, we want to use it for not only the operational tasks but more for decision support. This is becoming more and more important. If you want to keep ahead of your competitor, you must get a better understanding of their needs, the trends in the market, the correlation between events, etc. Not only in competition, but good information also makes the whole company perform better. Good information means information that is there at the right moment, at the right place and in the right format.

For people who are not used to work with GIS and even those who do, it is hard to imagine that over eighty percent of the business data has some spatial context. This means that if you want to use data for your decision support systems, you must consider the use of this spatial factor.

To be able to use spatial data, and to take full advantage of the spatial dimension, the locational element data must be integrated in the Datawarehouse. The following GIS-concepts is

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

being used and, with GIS-technology, being implemented in the organization. There are four main items to distinguish:

- Geo-reference. Making the objects without geo-references, spatial enable. Objects are connected to a model which represents the real world. This is done by giving the x- and y-co-ordinates in a co-ordinate system. For the object tree: a point object with one co-ordinate, for the object building: a polygon object with several co-ordinates.
- Geo-coding. Objects like statistics, are objects without a geo-reference. To make it possible to visualize them in the model of the real world, to display them on a map, they must be connected to co-ordinates.
- Topology. Separated objects which in the "real world" have relations with each other must be in the model of the real world. An example is a road which is connected to another road.
- Spatial aggregation. Depending on the demands on the Data warehouse, a certain aggregation of the data will be necessary. Not all the detail data has to be stored in the Data warehouse. For instance, for distribution planning analyses, it is not necessary to have all the addresses (streets and house numbers) only the names and location of the streets are needed.

The advantages of using a Data warehouse lie in the better understanding of the business, the possibility of the customer being served better, understanding of the business risks, improvement of the business processes, being able to make more tailored made products and services. The most important attraction of spatial enabling your Data warehouse is being able to make dynamically geographic queries on your data, to aggregate your data to geographic areas, to analyses data, spatial (re) organization of your data and, finally, presentation and visualization.

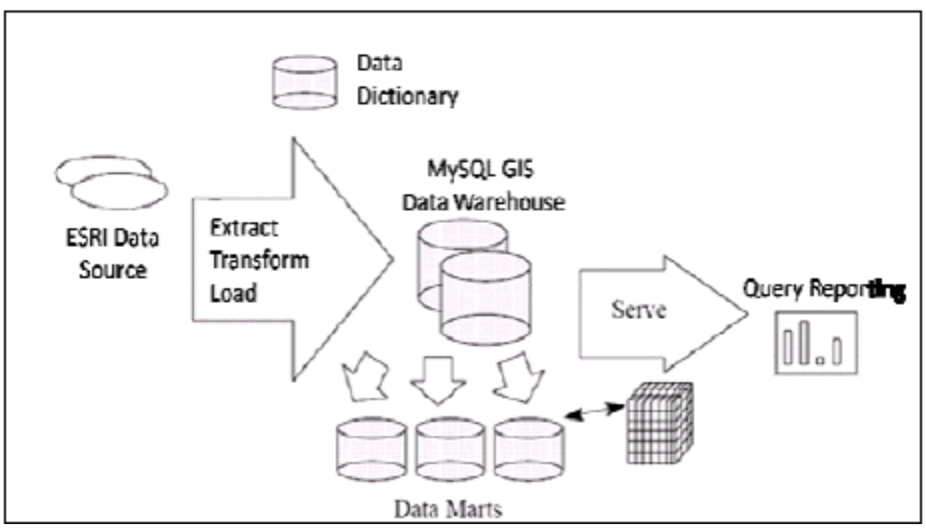

Figure 11. Data warehouse architecture for GIS applications

c. Modeling and Model Builder in GIS: batch and framework

Model Builder is an application used to create, edit, and manage models. Models are workflows that string together sequences of geoprocessing tools and providing the output of one tool to another tool as input. Model Builder can also be thought of as a visual programming language for building workflows.

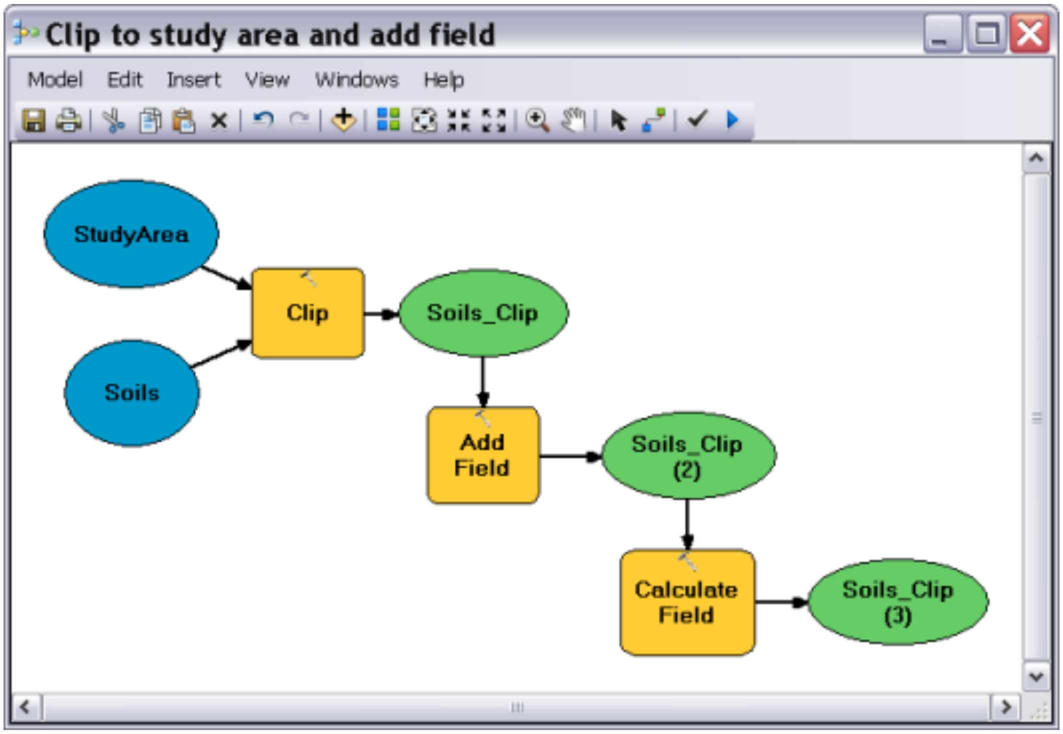

Figure 12. Example of GIS ModelBuilder

This model clips Soils for the study area, adds a new field, and calculates a value for the new field. While Model Builder is useful for constructing and executing simple workflows, it also provides advanced methods for extending ArcGIS functionality by allowing you to create and share your models as tools. Model Builder can even be used to integrate ArcGIS with other applications.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

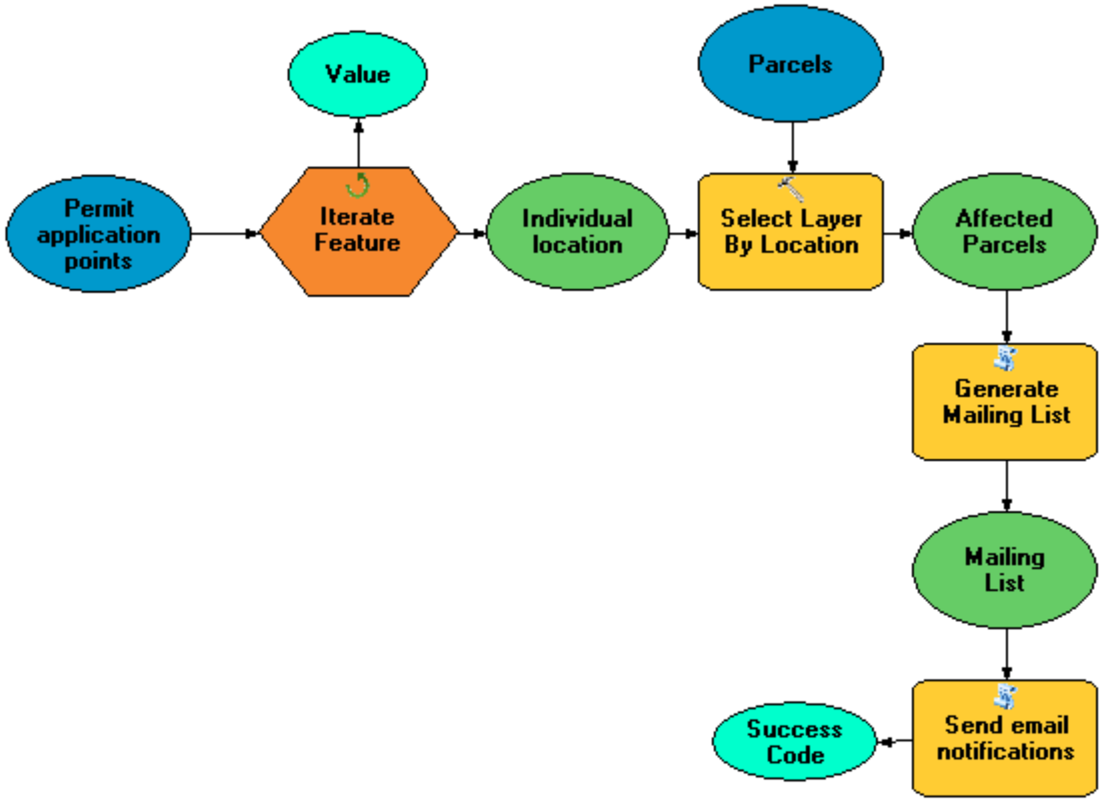

Figure 13. Example of GIS ModelBuilder with ArcGIS

An example is shown in Figure 13 to send email notifications The above model is used by a municipality to send email notifications to all addresses within 1 mile of an address for which a building permit application is filed. The model starts with a feature class of multiple permit application point locations. This feature class is fed into an iterator that loops over each individual point and feeds the point into the Select Layer By Location tool, where all addresses (parcels) within 1 mile of the point are selected. These addresses are then passed to a custom script tool (one that you or your colleague created), Generate Mailing List, that executes Python code to output a mailing list in HTML format. Finally, the mailing list is fed to another custom script tool, Send Email Notifications, which runs a custom executable that sends email notifications and produces a success code. The benefits of Model Builder are summarized as follows:

- Model Builder is an easy-to-use application for creating and running workflows containing a sequence of tools.
- You can create your own tools with Model Builder. Tools you create with Model Builder can be used in Python scripting and other models.
- Model Builder, along with scripting, is a way for you to integrate ArcGIS with other applications.

d. WebGIS and Mobile Application

Web-service architecture

Nowadays, people use both web applications and mobile applications. In the server, people build a web service that responds to clients' requests.

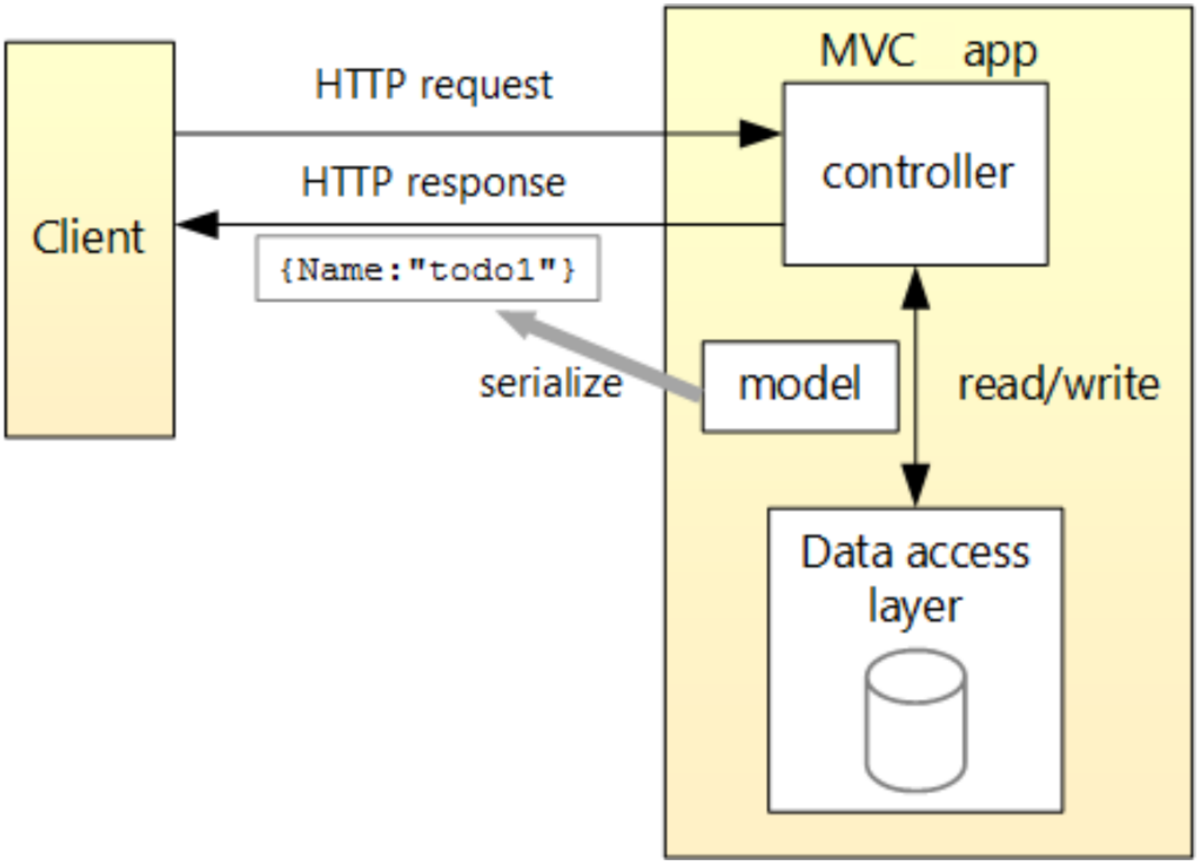

Figure 14. RestFul API for web service [Vainikka, J., 2018]

Client side can be a web application and mobile application. The data exchanged between client and service in this case is Json string. For a WebGIS application, GeoJson string is used for Geographic components such as: geometry, coordinator, configuration. Just like any other client-server system, the client in this case makes a request to the server. Content of request is wrapped inside a Json string. Server receives, processed, returns the results via another Json string. Client receives results, then visualizes them into a web browser and mobile application.

The main advantage of this architecture is that we can build only one server component that serves both mobile application and web application

GeoDjango for web service

GeoDjango is a built-in extension to the popular Django web application framework. Using GeoDjango, you can develop sophisticated web-based applications that let the user view and edit geospatial data. Before we can work with GeoDjango, however, we need to understand Django

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

(Rubio, 2017) itself. In this case, GeoDjango is used to build a web service. It processes the request and returns results via the GeoJson string.

Leaflet.js library for client visualization

On the client side, most of the cases, Javascript is chosen to be the programming language to visualize and manipulate the GUI. For web applications, Javascript and HTML are used to build webpages. In order to build a client application, we normally use many Javascript libraries together. In this case, Leaflet.js (or Leaflet) is just a Javascript library to visualize and manipulate GIS data in Web GUI. Leaflet can be stand-alone in a website. It also can be integrated into another component inside a web page or inside a Javascript framework for mobile application (such as React Native).

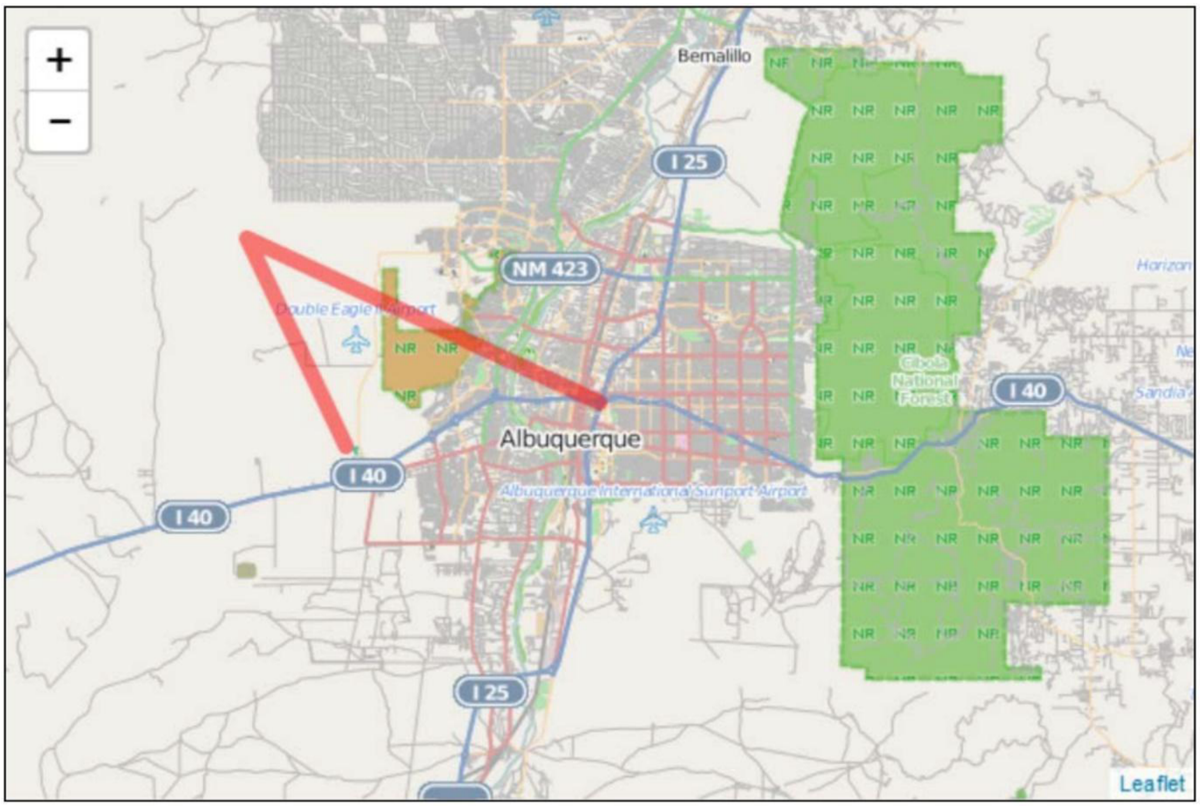

Figure 15. Example view of leaflet.js [Crickard III, P., 2014]

e. Spatial Data Infrastructure (SDI)

A spatial data infrastructure (SDI) is a data infrastructure implementing a framework of geographic data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way. Another definition is "the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data".

The emergence of spatial data infrastructures (SDIs) is closely associated with the efforts of collecting and producing geospatial data, as well as the advancement of surveying and computer

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

technologies. In the past decades, a large amount of geospatial data, such as remote sensing images and GPS locations, have been collected by government agencies such as the U.S. Geological Survey (USGS) and the National Oceanic and Atmospheric Administration (NOAA). Meanwhile, the fast development of geographic information systems facilitates the derivation of various data products from the collected data, such as topographic maps, land cover data, transportation networks, and hydrographic features. As location-based services are becoming increasingly popular, vast amounts of volunteered geographic information (VGI) has also been contributed by the general public through smart mobile devices and social media platforms. In addition, the componentization of GIS brings geospatial services that provide data processing and spatial analysis functions in the general Web environment. The large number of geospatial data, services, maps, and others, however, do not ease the use of these geospatial resources. On one hand, it is challenging to find and access these digital resources which are widely distributed at different government agencies and websites. On the other hand, a lot of data redundancies exist, and money and human resources were wasted in duplicated data collection and maintenance efforts.

Spatial data infrastructure presents a solution to the problems of resource discovery and data redundancy. It provides a unified platform where people can go and search geospatial data, maps, services, and other digital resources. As multiple government agencies are sharing their data on one platform, SDI reduces data redundancy and the extra efforts in collecting duplicated geospatial data. From a cost/benefit perspective, SDI allows geospatial data to be collected once and reused multiple times in different applications. More generally, SDI can be considered as an important element in the e-government and open- government movement to increase the transparency of governmental activities and to enhance public participation. Better access to geospatial data also stimulates the growth of new businesses which may not be possible otherwise.

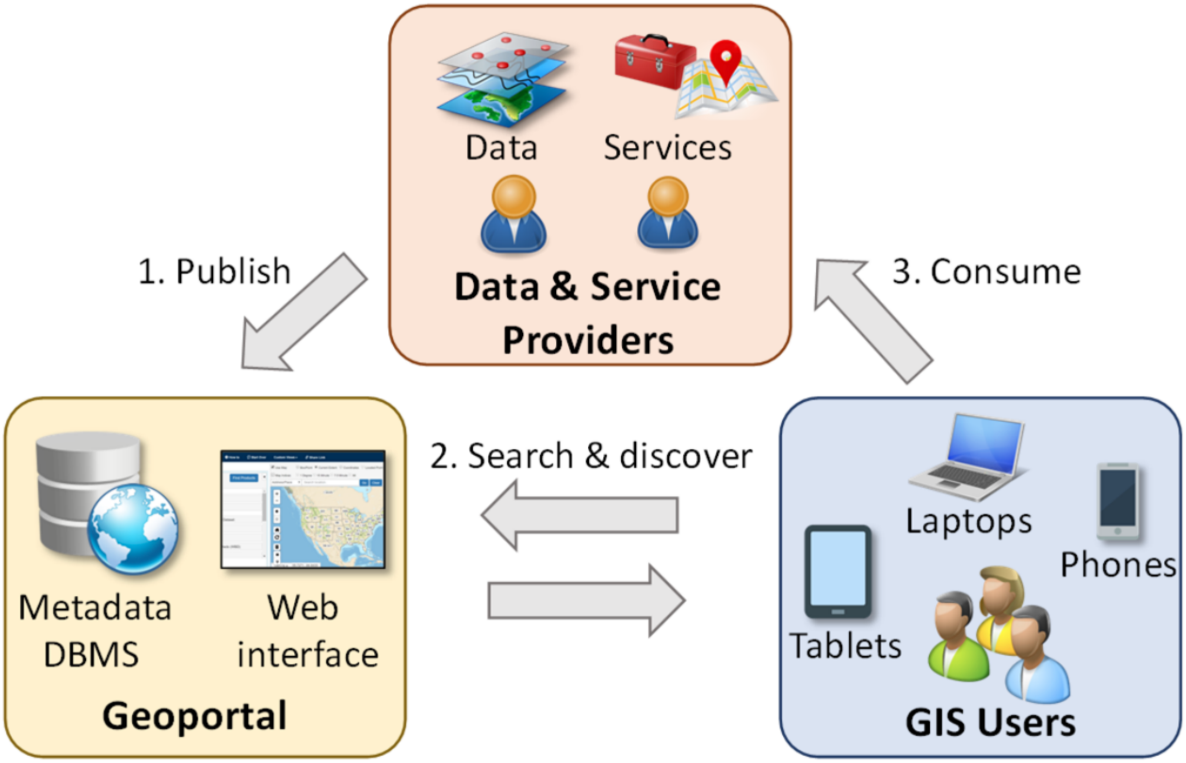

Figure 16. Example about 3 components of SDI

Geoportals are typically developed using Web-based technologies and off-the-shelf GIS software packages. A database management system (DBMS) is used to store and manage the metadata of the geospatial resources contained in the SDI. A Web interface, which often contains a map, enables end users to interact with the system and to conduct searches (Figure 15). When a search is performed, a HTTP (Hypertext Transmission Protocol) request will be sent to the Web server which hosts the geoportal. After querying the metadata stored in the database, the geoportal will then send back the result to the client through a HTTP response. Geoportals are typically designed to be used by both GIS professionals and the general public. One important function of geoportals is helping users discover the existing geospatial resources. This resource discovery process often follows the publish-find-bind pattern, in which: 1) providers publish the metadata of their data and services to a geoportal; 2) users perform a search on the geoportal and potentially find the data; 3) users consume the data and services from the providers. Figure 2 illustrates these three steps.

4. USE CASES

Several case studies will be introduced and illustrated during lectures. They are method development and analysis techniques for remote sensing data in applications of land cover classification (Man et al., 2018), paddy rice mapping and impact of flooding on rice area (Phan et al., 2019), air pollution modeling from satellite images (Nguyen et al., 2015, 2020; Phan et al., 2019). The programming environment is on Google Earth Engine or based on Python programming language and GDAL image processing library (Crickard, 2018; Rubio, 2017).

Software such as Google Earth, Arcmap will be used to obtain remote sensing image information and collect samples (Vainikka, 2018).

The GIS case study will be illustrated by selecting land to build an industrial zone in Cam Pha city, Quang Ninh Province, Viet Nam using a cartographic model. Using a database of elevation models of the study area and the required standards for the land to build the industrial zone according to the government's regulations. Besides, examples of using GeoJSON to develop WebGIS will be introduced.

After this course, students will gain knowledge and skill in following topics:

- Fundamentals of optical remote sensing for vegetation, water, atmosphere and urban landscape;
- Techniques for processing and integrating of multi-source satellite datasets, ground measurements, and field survey data;
- Development of machine learning models such as Random Forest, Neural Network, SVM, XGBoost, … for remote sensing applications;
- Analysis and evaluation of developed models;
- Familiarization with programming tools (Google Earth Engine, Python, GDAL library) and supported software (Google Earth, Arcmap).
- Know how to build the Spatial database and data warehouse (CAD, ArcGIS, Mapinfor)
- Know the use of GIS in the process of building the use of GIS in the process of building models with spatial data models with spatial data (ArcGIS, ArcTool box).
- Fundamentals of GIS and webGIS;
- Techniques for processing and integrating of GeoJSON;
- Development of Javascript (library leaflet) and Python (library GDAL, Flask) to build a web page of webGIS;

REFERENCES

Crickard, P. (2018). Leaflet.js Essentials. In *ACS Applied Materials & Interfaces* (Vol. 1, Issue 19). https://pubs.acs.org/doi/10.1021/acsami.9b03822

Jensen, J. R. (2014). Remote sensing of the environment: an earth resource perspective second edition. In *Pearson Education Limited,Harlow, England* (Vol. 1).

Jensen, M. E., & Everett, R. (1993). An Overview of Ecosystem Management Principles. In *Vol. II, Ecosystem Management: Principles and Applications* (pp. 7–16).

Jones, K. B., Ritters, K. H., Wickham, J. D., Tankersley, R. D., O'Neill, R. V., Chaloud, D. J., Smith, E. R., & Neale, A. C. (1998). *An EcologicalAssessment of the United States Mid - Atlantic Region*.

Le, T. M. Phuong, (2019), Mapping and GIS. Construction Publish

Le, T. M. Phuong. (2017). GIS in Urban management. Construction Publish ,

Le, T. M. Phuong, Vu Le Anh, Nguyen Hong Quang, Research on Urban Land Use Change in Ha Noi, Viet Nam Using Remote Sensing and GIS for Planning Oriented Work, International Journal of Sustainable Construction Engineering and Technology, 2020.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.

Man, C. D., Nguyen, T. T., Bui, H. Q., Lasko, K., & Nguyen, T. N. T. (2018). Improvement of land-cover classification over frequently cloud-covered areas using landsat 8 time-series composites and an ensemble of supervised classifiers. *International Journal of Remote Sensing*, *39*(4), 1243–1255. https://doi.org/10.1080/01431161.2017.1399477

Nguyen, T. N. T., Bui, Q. H., Pham, V. H., Luu, V. H., Man, D. C., Pham, H. N., Le, T. H., & Nguyen, T. T. (2015). Particulate matter concentration mapping from MODIS satellite data: A Vietnamese case study. *Environmental Research Letters*, *10*(9). https://doi.org/10.1088/1748-9326/10/9/095016

Nguyen, T. N. T., Luu, V. H., Pham, V. H., Bui, Q. H., & Nguyen, T. K. O. (2020). Particulate matter concentration mapping from satellite imagery. In *TORUS 3 - Toward an Open Resource Using Services: Cloud Computing for Environmental Data* (pp. 103–130). https://doi.org/10.1002/9781119720522.ch5

Phan, A., Ha, D. N., Man, D. C., Nguyen, T. T., Bui, Q. H., & Nguyen, T. N. T. (2019). Rapid assessment of flood inundation and damaged rice area in Red River Delta from Sentinel 1A imagery. *Remote Sensing*, *11*(17), 1–24. https://doi.org/10.3390/rs11172034

RN Colwell. (1997). History and place of photographic interpretation. In *Manual of photographic interpretation* (Vol. 2, pp. 33–48).

Rubio, D. (2017). Beginning Django. In *Beginning Django*. https://doi.org/10.1007/978-1-4842-2787-9

Thirumalai, T. K. (2003). *Remote Sensing and Geospatial Information Technologies Applicable to Multimodel Transportation*.

Vainikka, J. (2018). *Full-stack web development using Django RESTframework and React*. *May*.

This document was produced within the MONTUS project financed by the European Union in the framework of Erasmus + Capacity Building 598264-EPP-1-2018-1-FR-EPPKA2-CBHE-JP.