GEOGRAPHICAL INFORMATION SYSTEMS

Frans J.M. van der Wel
KNMI, Royal Netherlands Meteorological Institute, the Netherlands

Keywords: Geographical Information Systems, metadata, data quality components: lineage, attribute and positional accuracy, GIS operators, Internet, OpenGIS

Contents

Summary

Geographical information systems (GIS) emerged as far back as the 1960s to meet the requirements of more complex spatial analyses. Nowadays, GIS is increasingly part of mainstream Information and Communication Technology and almost invisibly present in our daily life.

The unique approach of GIS to spatial data has been embraced by numerous disciplines that benefit from a standardized toolbox to process their data flows. The amount of data has increased dramatically over the past 30 years, mainly as a result of image data from space-borne satellites. GIS contributes strongly to the processing, storage and visualization of these data, such that the distinction between GIS and Image Processing Systems is slowly becoming artificial.

With GIS, mankind has a tool to manage its surroundings and anticipate timely and natural as well as human processes such as floods and migration respectively. Advances in Internet technology help to bring spatial data to the individual by intelligent mapping applications based on GIS. In this way, geographical knowledge is shared and, it is to be hoped, awareness of our precious environment is intensified.

1. Introduction

Driven by an ongoing willingness to understand our environment, mankind has put effort into mapping the world from the early beginning of human life. Knowing your "territory" is crucial for simply surviving and carrying on in a society that sometimes provides only limited conditions for a sustainable life style. The need for information about what is surrounding us is evident; it helps us to answer questions related to such issues as weather, crop growth, seismic activity and urban planning. All this data can be referred to as geographical data.Geographical data relate to meaningful data about earthbound objects, and therefore include a reference to position and topological connections in addition to attributes. In other words, they describe the place of an object in the geographical space by means of some coordinate system, its qualitative and quantitative appearances and the way in which it is linked with other objects in terms of adjacency, proximity and connectivity. Grimshaw expresses himself in concrete terms by defining the questions *"where is it?"*, *"what is it?"* and *"what is its relationship to other spatial features?"* respectively. Geographical data constitute the core of many decision-making strategies. Grimshaw quotes a study by Moloney *et al.* who states that 90% of business data is geographical, or, more general, spatial. The availability of geo-information is undoubtedly improving increasingly, and the dimensions of the accompanying data flow are so impressive that the notion of an information revolution is by no means exaggerated. Advancements in computer development have created the preconditions for the introduction of Geographical Information Systems (GIS), which have made both new and existing data sets accessible in a systematic way. Also, remote sensing techniques have really caused a flood of data that is likely to swell even more in the near future (see chapter *Remote Sensing*).

2. Geographical Information Systems – a Definition

Defining GIS is not an easy task to do as is apparent from the large number of divergent views spouted in literature. Chrisman refers to it as a complicated type of software covering the whole life cycle of geographical data, from data collection to interpretation and on. A better and more widely accepted definition of GIS is given by Burrough & McDonnell who consider GIS a complex of computer hardware and software embedded in a proper organizational context. The latter refers to such issues as training of staff and appropriate implementation of the system in the present workflow.

As far as the software is concerned, Burrough & McDonnell distinguish the following five technical tasks (see figure 1):

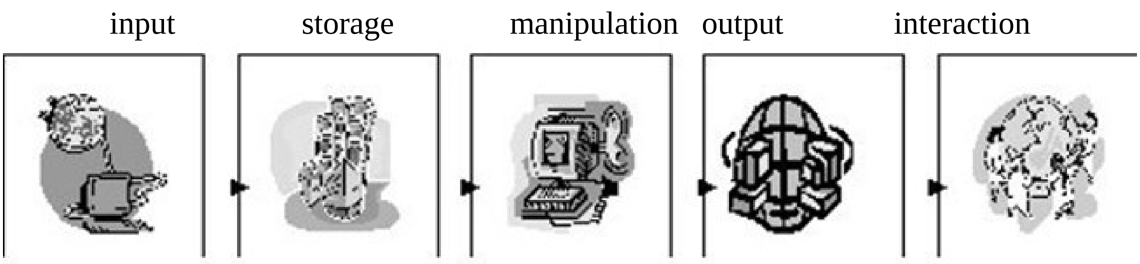

Figure 1: Five main tasks of GIS software, schematically represented

- *Data input and data verification:* The conversion of collected data into a suitable, digital format, for example by means of digitizers, scanners or keyboard. Moreover, it involves some kind of pre-processing, as data can be subjected to generalization or simple classification procedures.
- *Data storage and database management*: Once passed the input stage, data are stored in a database according to a particular data structure and database structure.
- *Data manipulation*: This involves all transformations being applied to the data. Berry uses the term “operations” whereas Abel prefers “transactions”. In general, a distinction can be made between analyses (spatial or not) and more trivial processing tasks like updating and simple error removal.
- *Data output and presentation:* The data, processed or not, can be presented in a graphic or alphanumeric way, as hardcopy (e.g. a paper map) or softcopy (e.g. so-called ephemeral output on a computer screen).
- *Interaction with a user:* A user is able to communicate with the information system (“query input”) in order to extract information from the stored data.

Marble & Peuquet give a well-chosen description of the functionality of a GIS, that summarizes the above-mentioned tasks: *“... a GIS is designed to accept large volumes of spatial data, derived from a variety of sources, including remote sensors, and to efficiently store, retrieve, manipulate, analyze and display these data according to user-defined specifications...”*

3. History of GIS

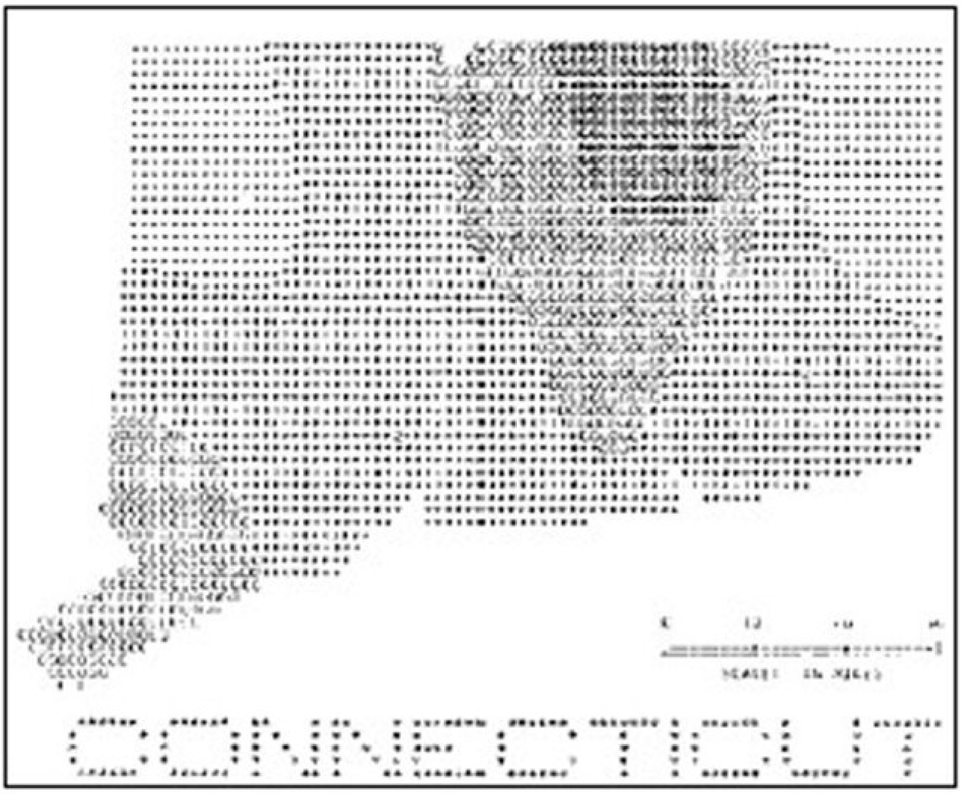

Figure 2: Example of a line printer map produced by the SYMAP system

Already as far back as the 1960s geographical information systems were being developed as a means to facilitate the traditional manual cartographic procedures of overlaying and visualization. One famous example is SYMAP from the United States which was rooted within the academic world. The maps that were produced were produced by a simple line printer and cartographically considered unacceptable, but the system was one of the first attempts to make the mapmaking process easier and faster (see figure 2).

In fact, one can state that GIS has its roots in North America, in which academic institutes and governmental agencies contributed to the concepts that underlie the current generation of GIS systems. The real commercial milestone, however, was only in the 1980s when computer technology had evolved sufficiently to become affordable for a growing group of users. Environmental Systems Research Institute (ESRI) launched its Arc/Info software in 1981 and in 2001 it released the next generation of GIS systems known as ArcGIS. The history of this market leader is exemplary for the speed with which the field has evolved since the early 1980s. ESRI held its first user conference in 1981 hosting 18 people; nowadays it attracts well over 9,000 participants from all over the world!

The technological shift in the 1980s and 1990s brought GIS closer to the user. Personal computers and a focus on Microsoft's Windows operating system were responsible for a high level of acceptance of these information systems. A generation of “desktop GIS” characterizes the last decennium: ESRI’s Arcview, Geomedia from Intergraph Corporation and MapInfo Professional are among the most widely used packages.

GIS is primarily concerned with data acquisition and data processing techniques, and its advent has enabled the integrated use of data from different sources. Imagine how in the early days a land suitability map was produced: different maps revealing soil information, height and slope data, hydrology, etc., would have been projected on transparencies and overlaid on a light table. Areas that met certain requirements were outlined manually, a cumbersome and time-consuming piece of work! By automating this process, multiple layers of data can be combined and queried according to some user-defined statement. The awareness that GIS could help to work efficiently through large geographical data sets in order to explore possible spatial relationships opened doors to an enormous group of people using GIS and developing applications for it. Recently, network technology – especially the Internet – has added another argument for using a GIS: sharing spatial data has become easy and within the reach of numerous non-experienced users that benefit from user-friendly, standardized tools like geographical information systems. According to Burrough & McDonnell, an estimated 93 000 sites worldwide had a GIS installation in 1995.

4. The Relevance of Metadata

In his paper *The Digital Earth: Understanding our planet in the Twenty-First Century*, former vice-president Al Gore of the United States referred in 1998 to the huge amount of georeferenced data that is available to our community (Al Gore's "the digital Earth" speech). He also stressed the fact that only a small part of this data is actually used. Now that we have the disposal of technologies to collect (*remote sensing*) and process

(*GIS, image processing*) these data almost instantaneously, the data flow is apparently too unstructured to be of immediate use.

At this point metadata, or data about data, enter the stage because they describe the considered data not only for reasons of storage but also for retrieval. If we know the specifications and the fitness for use of the data, we are provided with handles that help us to decide whether or not data are useful. If additional data are made available for the description of data sets, according to a number of *quality components*, the resulting increase of *data consciousness* is expected to contribute to better decisions. In the literature, these quality components are generally considered a subset of metadata and placed in one of the following categories according to Moellering's classification for the National Committee on Digital Cartographic Data Standards (NCDCDS):

- *lineage* – describes the "pedigree" of a data set, including information about source, age and level of processing;
- *attribute accuracy* – refers to the correctness of a non-spatial characteristic that has been assigned to an object;
- *positional accuracy* – indicates the extent to which the location and height of an object have been correctly represented;
- *completeness* - describes the relationship between objects represented in a data set and "*…the abstract universe of all objects…*", which is dependent on a specific context and real world model;
- *logical consistency* – deals with the data set itself as it focuses on the validity of object representations, both from a geometric and thematic point of view, and the relationships between these objects.

These quality components have been adopted by the International Cartographic Association (ICA) as the ingredients of quality reports which are in turn part of *de jure* spatial data standards. The European Committee for Standardisation (CEN) has published its guidelines for the description of spatial data while the US Content Standard for Digital Geospatial Metadata of the Federal Geographical Data Committee (FGDC) legally prescribes the standardized documentation of spatial data.

Within the context of GIS, the work of the OpenGIS Consortium is worth mentioning (http://www.opengis.org). Industry, academics and governmental organizations co-operate in order to define open interface specifications to advance the unbounded exchange of geographical data, at least as far as technology is concerned. Their ideas on metadata are inspired by the work of ISO's Technical Committee 211. As with all standards, it might take some time before a crystallized standard appears!

-
-
-

Bibliography

Abel, D.J. (1989) A model for data set management in large spatial information systems. *International Journal of Geographical Information Systems*, vol.3, no.4. pp. 291-301. [example of a contribution to the establishment of a geo-information science]

Berry, J.K. (1987) Computer-assisted map analysis: potential and pitfalls. *Photogrammetric Engineering and Remote Sensing*, vol.53. pp. 1405-1410. [With the advent of automation, a new perspective on the reliability of data processing was required, as this paper shows.]

Brassel, K., F. Bucher, E.-M. Stephan & A. Vckovski (1995) Completeness. In S.C. Guptill & J.L. Morrison (eds.): *Elements of spatial data quality*. pp. 81-108. ICA/Pergamon, Oxford. [chapter from one of the first books on spatial data quality, a must for everyone interested professionally in spatial data quality in the broader context of metadata]

Burrough. P.A. & R.A. McDonnell (1998): *Principles of geographical information systems*. Oxford University Press, Oxford. [a new edition of what is widely considered to be THE textbook on GIS]

CEN (1996) *Geographic information - Data description - Quality*. Draft European Standard, prEN 287008 (July 1996). European Committee for Standardisation (CEN), Brussels.

Chrisman, N.R. (1984) The role of quality information in the long-term functioning of a GIS. *Cartographica*, vol.21, pp. 79-87. [Nicolas Chrisman's papers deal for a large part with the issue of errors and uncertainties in spatial data: with this particular paper he was far ahead of his time.]

Clinton, W.J. (1994) Coordinating geographic data acquisition and access: the National Spatial Data Infrastructure. Executive Order 12906, April 11, 1994.

Davis, F.W. (1991) Processing of GIS and remotely sensed data for environmental analysis. In J.L. Star (ed.): *Proceedings - the integration of remote sensing and Geographic Information Systems*. pp.137-150. ASPRS, Bethesda. [paper on the integrated use of GIS and remotely sensed data, typical for a time when several levels of integration were discussed, both on a system and data level]

Grimshaw, D.J. (1994) *Bringing geographical information systems into business*. Longman Scientific & Technical, Harlow. 273 pp. [example of a contribution to the establishment of a geo-information science]

Hellyer, G.M., D.L. Civco & M. Berbrick (1990) Integration of Landsat Thematic Mapper derived land use/cover information into a Geographic Information System (Arc/Info) to assist in the identification, characterization, and protection of a riparian wetland dominated landscape in southern New England. *Proceedings ISPRS Commission II/VII International Workshop on advances in spatial information extraction and analysis for remote sensing*. pp. 25-40. ASPRS, Bethesda. [example of the extra value of remotely sensed data in environmental studies when used together with GIS]

Janssen, L.L.F. (1993) *Methodology for updating terrain object data from remote sensing data*. PhD-thesis. Landbouwuniversiteit Wageningen, Wageningen. 173 pp. [interesting study focusing on the use of GIS data to improve land cover classification of remotely sensed data]

Kainz, W. (1995) Logical consistency. In S.C. Guptill & J.L. Morrison (eds.): *Elements of spatial data quality*. pp. 109-137. ICA/Pergamon, Oxford [chapter from one of the first books on spatial data quality, a must for everyone interested professionally in spatial data quality in the broader context of metadata]

Marble, D.F. & D.J. Peuquet (1983) Geographic information systems and remote sensing. In D.S. Simonett & F.T. Ulaby (eds.): *Manual of remote sensing*. (second edition) Volume 1 - theory, instruments and techniques. pp. 923-958. American Society of Photogrammetry, Falls Church. [authoritative paper on GIS and remote sensing]

Moellering, H. (ed.) (1987) *A draft proposed standard for digital cartographic data*. National Committee for Digital Cartographic Data Standards, Columbus (Ohio).

Moleney, T., A.C. Lea & C. Kowalchuk (1993) Manufacturing and packaged goods. In *Profiting from a Geographical Information System*. pp.105-129. GIS World Books, Fort Collins. [examples of the extra value of GIS in a number of applications]

Morrison, J.L. (1988) The proposed standard for digital cartographic data: report of the digital cartographic data standards task force. *The American Cartographer* 15, no.1. pp. 129-135.

Rhind, D. (1988) Geografische informatiesystemen en kartografie (in Dutch). *Kartografisch Tijdschrift XIV*, no.2. pp. 25-27. [Translated in Dutch, this paper reflects the discussion on the position of GIS, cartography and remote sensing in the 1980s in which the advent of the personal computer required a new vision on geoinformatics.]

Strahler, A.H. (1980) The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment* **10**. pp. 135-163. [clear and comprehensive paper on the use of Bayes' Maximum Likelihood Classification applied to remotely-sensed data]

Trotter, C.M. (1991) Remotely-sensed data as an information source for geographical information systems in natural resource management: a review. *International Journal of Geographical Information Systems*, vol.5, no.2. pp. 225-239. [remotely-sensed data as an input for GIS in environmental studies]

Wilkinson, G.G. (1996) A review of current issues in the integration of GIS and remote sensing data. *International Journal of Geographical Information Systems*, vol.10, no.1. pp. 85-101. [example of a contribution to the establishment of a geo-information science]

Biographical Sketch

Frans van der Wel (Utrecht, 1965) was educated in Physical Geography at Utrecht University. During his studies he focused on ways to process and present spatial data and their metadata: Hence his interest in cartography, image processing and geographical information systems (GIS). In 1990 he graduated as a cartographer and started his research activities at the Cartography Section of Utrecht University. In this period, integration of remotely sensed data with GIS data gained much attention within the scientific community. The role of GIS ("mapping") data related to the reduction of uncertainties in remote sensing land cover classifications has been the subject of his PhD study. In addition to assessment and reduction of uncertainty, progress was made in the field of cartographic visualization of spatial metadata, resulting in the thesis *Assessment and Visualization of Uncertainty in Remote Sensing Land Cover Classifications* (2000). Currently, Dr van der Wel is working at the Royal Netherlands Meteorological Institute (KNMI) where he is responsible for the use of GIS in climatological and meteorological applications. Moreover, he is involved in data infrastructure issues, concentrating on the accessibility of data through the Internet via web applications.