STANDARDIZE GEODATA MANAGEMENT IN VIETNAM – AN URGENT NEED

Dinh Quang Duong[1], Harro Stolpe[1], Christian Jolk[1], Sandra Greassidis[1], Nils Führer[1], Björn Zindler[1], Manfred Fink[2], and Duc Phuoc Vo[3]

[1] Environmental Engineering and Ecology (EE+E), Ruhr University of Bochum, Germany
[2] Department of Geography (DEO), Friedrich-Schiller-Universität Jena, Germany
[3]Da Nang University of Sciences and Technology (DUT)

ABSTRACT

In Vietnam, many legal documents indicated that the purpose of acquiring geographical information is to print maps. Therefore, the CAD (Computer Aided Design) data model is commonly used. This fact brings a huge challenge for the users if they want to use the data for analyzing purposes (statistical analysis, spatial analysis, hydrological modelling, etc.). In this context, the paper aims to summarize and highlight the critical shortcomings of geodata from Vietnamese authorities, explain why there is an urgent need of geodata management standardization in Vietnam and to share ideas on possible solutions as well as discussing a "golden key" like the INSPIRE (Infrastructure for Spatial Information in the European Community) Directive.

Keywords: geodata model, geodata management

1. INTRODUCTION

The University of Bochum, Environmental Engineering and Ecology (EE+E) research group and the Jena University, Department of Geography have been conducting various research projects in several provinces of Vietnam since 1999 and gained vast experience in handling Vietnamese geodata. The handled geodata focus on water and land management, mining, urban and environmental planning. The geodata are being processed in order to compile e.g. planning and risk analysis maps at provincial and regional scales. Misunderstandings and technical problems repeatedly occur during data collection and processing. How could such problems be avoided and what are the solutions?

2. GEODATA MODELLING AND MANAGEMENT

2.1 Development history

According to Zeiler (2010), geodata reflects the description of our physical world with objects (buildings, water bodies, land, etc.), the behavior of each, and the interactions among them. With a long history of development, geodata appeared from simple maps on a paper to a complex geodatabase powered by computer systems. From the 1960s, the first geodata model was introduced namely Computer Aided Design (CAD) with four fundamental elements: point, line, polygon (features) and text (annotation). The most important purpose of a **CAD model** was to produce printed maps with certain predefined scales. Vectors and raster data are not supported by a CAD model, hence, the **coverage model** came out in 1980s which solved these shortcomings and added structure tables to sort descriptive information of features. The coverage model has a capability to store exact location of every single object. However, each feature type (point, line and polygon) has a generic behavior, meaning there is no difference between a line of road and a line of river. In 2008, a **geodatabase model** was introduced with an incredible capacity because it was a combination of geodata and a relational database (Zeiler, 2010) The following points highlight the capacities of a geodatabase (after Zeiler, 2010; ESRI, 2013):

- Fully object-oriented data model, adding logical objects such as owners. Each object could be defined to have their unique behavior
- Integrity dataset with vectors, rasters, tables, and locators (geographic locations)
- Objects are not only interacting with each other by single relationship but also with complex interaction sequences
- Having full power of a relational database (searchable, multiuser, versioning, remote deployable - WebGIS, authenticating, etc.)

The obvious disadvantage of a geodatabase is the extensive effort of building it (designing, structuring, and unambiguity) versus the simple structure of CAD files.

Metadata, as a curriculum vitae of the data, contains important information about the origin of the data, usage restrictions, keywords (tags) and so on. Depending on the data model, metadata are stored in different ways: they are texts with the CAD model but associated tables with the other models.

2.2 Status quo in Vietnam

To date, there is no concept of spatial data management in Vietnam. All geographical data were surveyed and managed based on thematic and singular purposes: cadastral, land use, soil, geology, surface water, groundwater, topography, etc. The government issues a number of decrees, circulars and standards of how data has to be obtained and processed. All legal framework are based on the claim that the final product will be a printed map (e.g. Ministry of Natural Resources and Environment, 2014a, 2014b). The legal documents concentrate only on clearly defining layouts of maps including specific symbol and annotation styles. Based on that, the data providers apply simplest-possible techniques to produce *a print friendly version of data, not a database*. Table 1 shows certain techniques and data models currently used in Vietnam in different thematic applications.

Table 1: Data models of different thematic applications in Vietnam

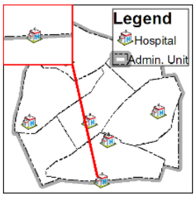
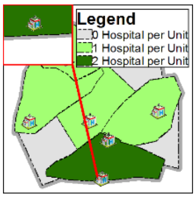
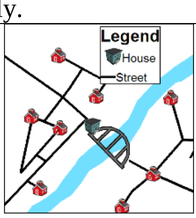
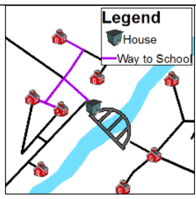
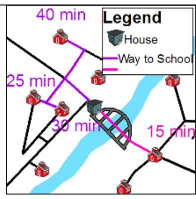
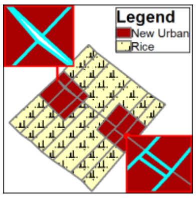
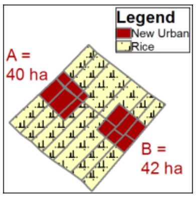
Thematic applications	Data model
Cadastral	CAD
Land use, land use planning, forestation	CAD, Coverage
Soil, geology, mineral resource, water management	Coverage with very limited attribute table; CAD-based annotation
Topography	CAD, Coverage

Metadata come with geodata from Vietnam do not have a standard structure and they are mostly absent from the dataset.

3. SHORTCOMINGS OF GEODATA IN VIETNAM

The quality of geodata in Vietnam varies depending on their sources and production date. Three examples of the typical shortcomings of CAD data and real examples of CAD data in Vietnam are visualized in Table 2. The advantages of geodatabase in solving such problems are also shown in the Table 2.

Table 2: Examples for shortcomings of CAD data in Vietnam

Typical shortcomings of CAD data	Real examples of CAD data from Vietnam	Advantages of Geodatabase
 Objects are fixed-size symbols. Accuracy is depending on the map scale. Query languages are not applicable. Calculations are done manually.	 Inaccurate locations of objects.	 Application of query languages for automatical calculation, geodatabase rules for error handling.
 No network information. Network analysis is impossible.	 Lack of network continuity, the road is cutoff at the bridge. A manual editing is required to fulfill the network.	 Network analysis: e.g. the shortest way between a home and different schools can be found automatically.
 Fixed texts for annotations. The texts are inputted manually, causing man-made errors.	 Gaps and overlapping polygons exist. Calculated areas and actual areas are different.	 Flexible, scalable annotations. Automatically calculated attributes avoiding man-made errors.

Real life problems will be described in the Chapter 4 with two case studies highlighting hands-on examples of working with such bad quality CAD data.

4 CASE STUDIES

Chapter 4 will show experiences from research projects featuring two case studies that describe shortcomings of CAD data and their impacts to real life applications.

4.1 Land Use– converting CAD data to geodatabase

The German Vietnamese joint research project „Research Association Mining and Environment“ (RAME), funded by the German Ministry of Education and Research, is developing a concept of post-mining land use planning for a coal mining area in Quang Ninh province, Northeast Vietnam. In this context, one of the main tasks is the implementation of a geodatabase capable of analyses including basic calculation of areas, intersection of different objects etc.

This case study is showing the conversion of a land use map drawn in MicroStation into a geodatabase, ready for further analysis steps. Main requirements for the geodatabase are nonoverlapping closed polygons seamlessly fitting to each other and attribute tables linked with the features.

The available CAD data (MicroStation format) are compiled in a map showing the land use of Quang Ninh. Objects in the original data are colored geometric elements without spatial or thematic classification. In order to analyze and process data within a spatial reference, it is necessary to convert the original CAD data into a geodatabase. The "data interoperability tool" of ESRI ArcGIS allows the direct data access and translation into a file geodatabase in ArcGIS. The conversion enables a further post-processing, where unnecessary data were manually deleted and essential data (polygons, lines, points) remained. Additionally, the CAD data did not have any projection attached. Thus, the converted data were spatially referenced (georeferencing).

Figure 1 shows the result of the imported MicroStation map into ArcGIS geodatabase. Errors (such as overlapping, interruption, doubling, open polygons, linear information in polygons etc.) became obvious and had to be corrected in order to use the geodatabase for analyzing. All errors might lead to wrong results during the analysis. For example, ArcGIS will not interpret an open polygon as an area thus rendering a correct calculation of the percentage of industrial areas in Quang Ninh is impossible.

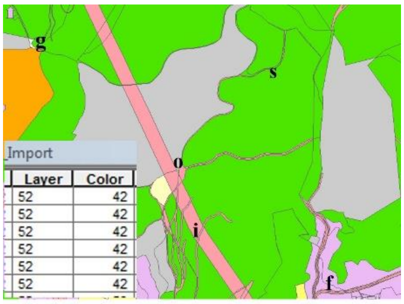


Figure 1. CAD data with obvious errors and no linked attributes

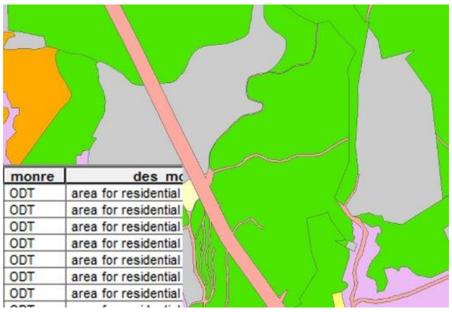


Figure 2. Corrected and attributed geodatabase

Figure 1 shows errors existing in the land use CAD data; i.e. missing data gaps (g), overlapping polygons (o), interruptions (i), seams at contact lines (s), fragmentations (f). Figure 2 shows the corrected and attributed geodata. Some errors remain e.g. width of streets (linear apricot elements) is larger than in reality (wrong representation of area percentage).

The Topology Tool of ArcGIS is used to check and verify any violations to user defined integrity rules. The detected technical errors are being corrected automatically with the provided solutions. Any undetected errors had to be corrected manually.

The CAD data include statements of the geometry and their representation only (lines, circles, points, line color, line width etc.) and they are not linked to attributes describing their characteristics. In order to make the data analyzable, all objects have to be associated with their attributes – the attributes have to be in tables rather than texts on the map. The attributes are semi-automatically transferred from the object information originated in CAD data (classified by colors and layers). However, parts are put in manually.

Figure 2 shows the resulting geodatabase ready for analysis. It can be displayed in maps according to thematic queries. The table in Figure 2 depicts the attributes attached to each element (polygon) of the geodatabase.

4.2 Hydrology – revising a DEM

The German Vietnamese joint research project „Land Use and Climate Change Interactions in Central Vietnam“ (LUCCi), funded by the German Ministry of Education and Research is developing land and water management strategies to adapt to climate change. Hydrological simulations are one crucial tool to develop such strategies.

For the description of the hydrological and geomorphological dynamics one of the most important information layers is the morphological description of the earth surface. The morphology is the governing factor to define the gradients and flow paths in the system.

Freely obtainable datasets include the globally available datasets like SRTM- (Shuttle Radar Topography Mission) and Aster-DEM (Digital Elevation Model). Also a DEM derived from contour lines and points from digital topographic maps (1:50.000) exists. Since the topographic maps have been primarily designed for visualization purposes the quality control of the attribute data was not very accurate. This leads in some parts of the area to errors in the altitude information e.g. confusion of digits, missing or additional digits. The solution used for the existing DEM was to delete the affected contour lines and points which leads to major information loss. This is especially the case because almost all contour lines with a lower hierarchy have been deleted. Some errors were not identified resulting in wrong surfaces (see Figure 3, left).

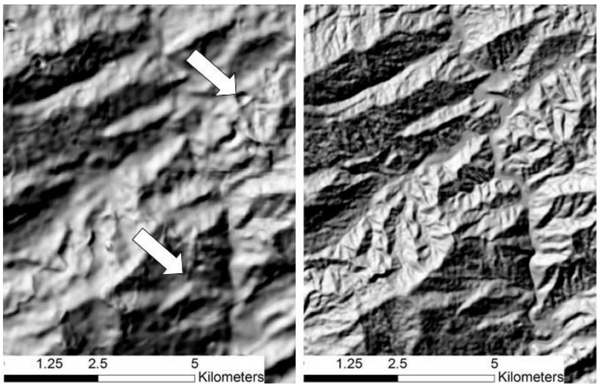


Figure 3. Example for the same display detail of the existing and new generated DEM. The arrows mark examples for artefacts caused by erroneous attribute data.

To overcome the described problems, a new DEM has been developed using the contour information of the topographic map. Since the contour lines were mapped by the US-Army during the American/Vietnam War, they are very detailed. On the other hand, they show weaknesses in the description of the valleys and lowlands because of their military origin. In order to improve this weakness/gap, the contour information has been aggregated with the data of the SRTM-DEM to form one combined dataset. The following steps have been performed: (i) Correction of the attribute data using the annotations of the contour lines in the map and application of plausibility rules of contour steps regarding the neighbor lines in a manual or semi-automatic manner. (ii) Transformation of the SRTM-DEM into points. Production of a buffer map of all contour information originated of the topographic map with a buffer distance of 90 m. Application of the buffer map to delete SRTM points and joining the remaining SRTM points with the contour points. (iii) Interpolation of a final raster (see figure 3, right).

The resulting DEM has been checked optically with the help of DEM derivations like “hillshade” and “slope”. No systematic inconsistencies or faults could be identified and the DEM is utilizable for further use.

5. CONCLUSIONS AND DISCUSSIONS

Above chapters give a fact of geodata quality in Vietnam and problems coming with it. In order to solve the problems, a standardization of geodata management need to be carried out. There are many steps need to be done, starting from including metadata into geodata sets, selecting geodata model and, finally, the whole process needs to be embedded in a technical and institutional framework. This Chapter will explain the importance of metadata and geodata model and give a suggestion on spatial data infrastructure management framework which is being used in the European Community.

5.1 The importance of metadata and geodata model

Metadata are essential for the inheritance and interchange of any kinds of geodata, thus, any pieces of geodata shall come with their metadata. Metadata should contain information about the digital data stocks themselves (semantics), access mechanisms (syntax), the data structures (structure) and storage sources (navigation). Using metadata with a standard structure helps to avoid redundant data entries and to reveal existing gaps in the data sets. Metadata also helps in the standardization of data and terms as well as in the quality assurance for the data sets. More information on the topic metadata can be found in the international standard ISO 19115-1:2014 "Geographic Information – Metadata".

Geodata are an integral part of many economic and development activities. Geodata are used for analyzing, representing and online deploying purposes rather than for printing maps only. Therefore, the CAD data model and even the coverage model are no longer suitable for an integrated geodata management system. Professional geodata applications are only applicable on the geodatabase data model. A consideration to use the geodatabase or an advanced data model is strongly suggested.

5.2 Spatial information infrastructure

Geodata are only one component of the spatial information infrastructure (SDI). There are many other components such as a computer system, technical standards, management entities etc. A good and sustainable management of spatial information depends on how such components will be defined and brought into one system. Vietnamese Government is spending out millions of dollars every year for geodata collecting and managing (Hung Vo Dang, Lam Son Hoang and Anh Tuan Vo, 2012). However, a complete SDI is still an on-developing idea. The INSPIRE (Infrastructure for Spatial Information in the European Community) Directive, which came in force in 2007, is the tool that EU countries use in order to build up a SDI. The main objective of the INSPIRE directive is to collect, document and share spatial information among member countries. INSPIRE provides technical guidelines for a whole spatial information infrastructure from the topmost framework (metadata, spatial data set, data services, network services etc.) down to the detail technical specifications of every single feature (rivers, sluices, buildings etc.) (European Parliament, 2007). INSPIRE is one of the most fruitful templates and should be adapted to fit into Vietnamese geodata handling and legal framework. Especially the data specification component for the starting phase can help to alleviate many problems that have been highlighted in Chapters 3 and 4.

6. REFERENCES

ESRI. *ArcGIS Desktop: Advance*: ESRI Press, 2013.

Etablishing an Infrastructure for Spatial Information in the European Community: INSPIRE, 2007. http://inspire.ec.europa.eu/index.cfm.

Hung Vo Dang, Lam Son Hoang, and Anh Tuan Vo. "Development strategy of national spatial data infrastructure for Vietnam." 2012, accessed November 2014.

Thông tư Quy định lập bản đồ chất lượng nước dưới đất tỷ lệ 1:50.000: 09/2014/TT-BTNMT, 2014.

Thông tư Quy định về thống kê, kiểm kê đất đai và lập bản đồ hiện trạng sử dụng đất: 28/2014/TT-BTNMT, 2014.

Thông tư Quy định kỹ thuật về mô hình cấu trúc, nội dung cơ sở dữ liệu nền địa lý tỷ lệ 1:2000 và 1:5000: 55 /2014/TT-BTNMT, 2014.

Zeiler, Michael. *Modeling our world: The ESRI guide to geodatabase concepts*. 2nd ed. Redlands, Calif.: ESRI Press, 2010.