STATISTICAL COMMISSION and Demonstration 1
ECONOMIC COMMISSION FOR EUROPE English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography
(Neuchâtel, Switzerland, 10-12 April 2000)

Topic (ii): Spatial database management and (geo-)data warehousing

GEOTOPOGRAPHICAL DATABASE (GTDB) A DATA-DICTIONARY-BASED GIS FOR PRODUCTION AND DISTRIBUTION OF SWITZERLAND'S DIGITAL LANDSCAPE MODELS

Submitted by the Swiss Federal Office of Topography[1]

Demonstration

ABSTRACT

The Swiss Federal Office of Topography (SFOT) has a long history of high quality products. Technical development increases the need for topographical information in digital form. The economic situation requires a wider use of digital landscape models. GTDB is a data-dictionary-based GIS, which aims at preserving the quality of vector datasets while improving support for an efficient update and distribution process. The data-dictionary is the core component of the GTDB. It also ensures quality assurance in the production as provides documentation of vector datasets.

Keywords: Conceptional datamodel, data dictionary, data production, data distribution, geodata, gis, information system, interlis, internet, quality assurance, rdbms, spatial data warehouse, topology.

I. INTRODUCTION

I.1 Economic situation

1. The Swiss Federal Office of Topography (SFOT) has been in charge of geodetic surveying and topographic mapping at a national level ever since it's foundation in 1838. SFOT has therefore a long tradition as a production site of high quality geotopographical products.

2. Since 1997 SFOT has participated as a pilot office in the Swiss Federal Administration's new Public Management Project (NPM). Thanks to NPM, SFOT has increased autonomy in allocating its resources. On the other hand it bears a higher responsibility to obtain technical and economic results. NPM intensifies the need to generate revenues from selling products and services.

[1] Prepared by Gaudenz Sonder.

I.2 Technical situation

3. In 1995 the Swiss Army lauched its FhrSim95 project. The purpose of this project is to provide a computer-based training system for military commanders. The system has to simulate the effects of military decisions in a realistic manner. This means that the simulation has to run based on a landscape model which is a proper reproduction of the real world.

4. To satisfy the need of FhrSim95, SFOT started its VECTOR25 project. The immediate goal of this project is to provide the high quality topographic data for the FhrSim95 within a short time. As the vector data is used for simulation purposes, topological correctness is a crucial succes factor.

5. Within one year SFOT succeeded in setting up a production process which is optimised for initial map digitising. In four main steps the raster maps of the national map of scale 1:25‘000 are transformed into vector format:

- vectorisation of raw data.
- building of node-edge-topology.
- buidling of area respectively overlay-topology.
- edge matching between neighbouring mapsheets.

6. The result of this process is called VECTOR25 Level 1 and it is a collection of 8 DXF-Files (e.g. traffic network) per mapsheet. Each object class (e.g. category of road) is represented by a DXF-Layer. Although DXF has no inherent topology, nodes are expressed by coincident line ends and areas by centorid points.

7. The production process for VECTOR25 is optimised for initial vectorisation of huge amounts of data, but it does not provide support for updating nor for distribution of the dataset.

I.3 Goals

8. To overcome this deficiency, SFOT started the GTDB project (geotopographical database). This project has a two-fold goal:

- to sustain the value of the digital landscape datasets in the long term
- to increase the usefulness of the geotopographical information.

9. GTDB has to combine the traditional strength of SFOT with the new requirements in a changed environment.

I.4 Sustainability

10. Sustainability means - among other things - that the data is updated periodically. It is important to make sure that the update-process does not compromise the quality of data: i.e. topological correctness, geometrical accuracy, consistency constraints and homogeneity are to be preserved. Assurance of quality therefore is a major issue of the GTDB project.

11. Besides ensuring quality, the GTDB also serves to document the datasets (metadata) and to provide an efficient tool for the back-up and archiving of data.

I.5 Usefulness

12. The political and economic environment of SFOT are changing. A federal office, which follows the NPM philosophy, has to permanently adapt the task it was given a long time ago by the legislation to

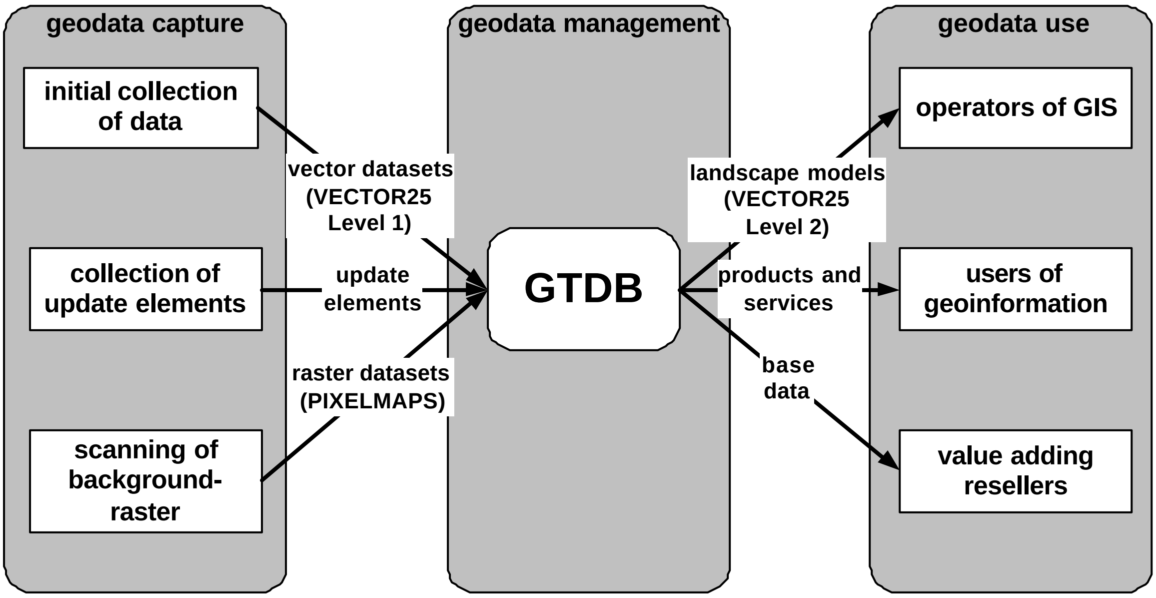
the current needs of its customers. Today, the demand for geotopographical information is shifting from printed form to digital form. To satisfy this demand, it is essential that products are distributed efficiently.

13. GTDB has to provide the tools which permit the provision and distribution of products derived from landscape models. This distribution process has to meet with customer needs (e.g. short delivery time, flexibility in selection criteria, etc.) and it has to be handled with limited resources on the part of SFOT.

II. SOLUTION

II.1 Context diagram

14. The context diagram specifies the system's interactions with its environment.



15. GTDB obtains its input data from other projects and processes:

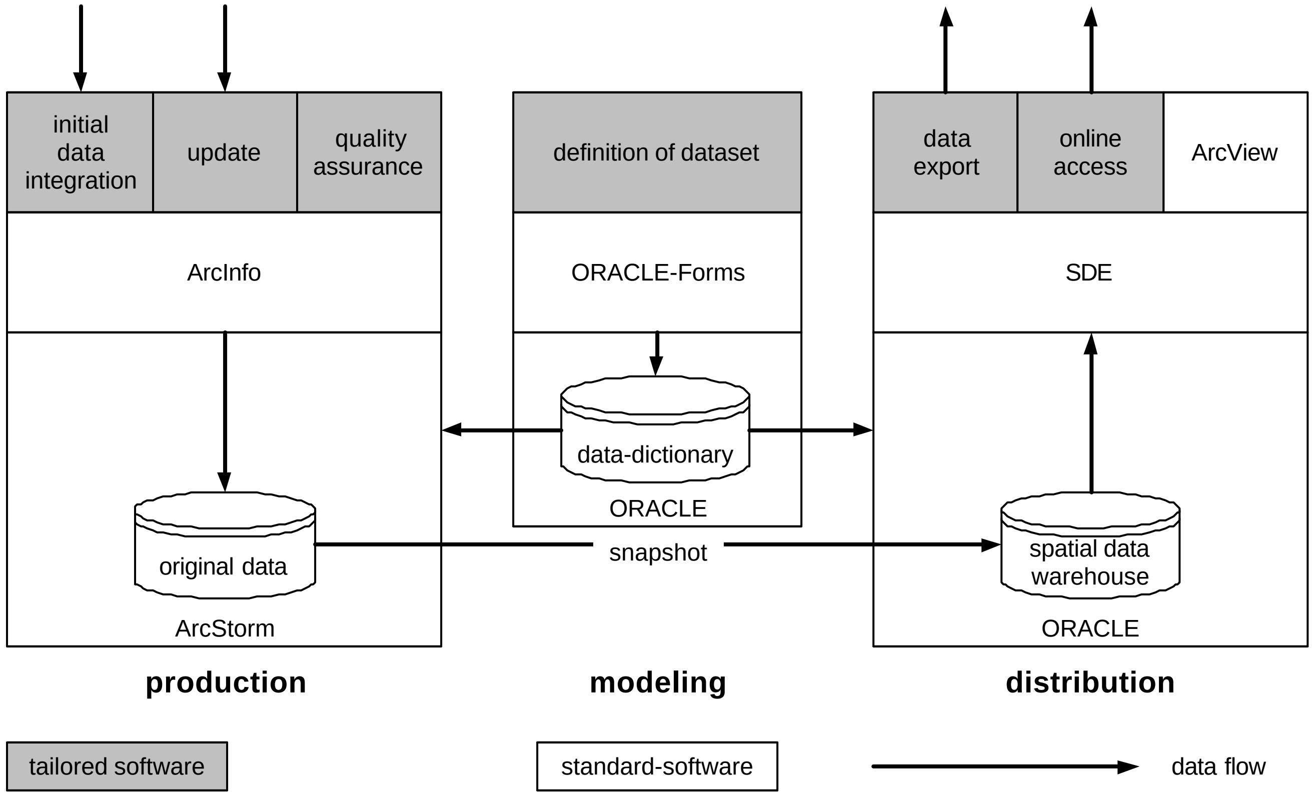
- vector datasets from projects which focus on initial collection of data.
- update elements from the photogrammetric and terrestrial surveys, which also serve to update the printed maps.
- pixelmaps and aerial photographs as background information.

16. GTDB provides and distributes products and services to different customers:

- datasets for operators of GIS (e.g. telecom and transportation industry).
- geoinformation (e.g. online querys, statistical results, maps on demand).
- base data for value adding resellers, that develop applications, provide services or publish information on the web.

II.2 System architecture

17. The following diagram describes the components of the GTDB system. The GTDB is based on standard-software from ESRI and ORACLE, which is extended by tailored software to better support the specific needs of SFOT.



18. The GTDB is made up of three main components:

- the production component supports the back-office-processes which are necessary to efficiently integrate and update datasets. The quality-checks are also part of this component. The tasks of this process are controlled by the data-dictionary.
- the distribution components support the client-side-processes which serve to provide and deliver off-line products (e.g. export files according to the customer demands) and online services (e.g. direct access to the data based on socket connections or HTTP).
- the heart of the GTDB system lies in its data-dictionary. It contains metadata to document the distinct datasets. Furthermore, the data-dictionary contains parameters for the quality-checks, the work with ArcInfo and the display of data.

19. At the time this paper is written, ArcStorm is used as database for the production component. As several operators work in parallel, ArcStorm has to ensure mutual exclusiveness.

20. A snapshot mechanism is used to replicate the original data into the distribution database. This database serves as a spatial-data-warehouse and is optimised for read-access. The distribution component uses ESRI's spatial database engine (SDE). SDE is a middle layer which provides for 2-dimensional indexing for geospatial data which is completely stored in an ORACLE database.

21. GTDB uses IBM RS/6000 Hardware with AIX as operating system on the server side and PCs with Windows-NT on the client side.

II.3 Data-dictionary

22. Before a new vector dataset is integrated into the GTDB, its structure (e.g. data-catalog: layers, objectValues, etc.) and its quality parameters are entered into the data-dictionary. GTDB's data-

dictionary is built using a relational database management system (RDBMS). Its ORACLE database is made up of 76 tables. The project information is entered by an ORACLE-forms-application.

II.4 Conceptional datamodel

23. The data-dictionary describes the conceptual data-model of datasets which are managed with the GTDB: A dataset is conceptionally made up of several thematic layers (e.g. hydrology network). Each thematic layer is composed of one or two topological layers (e.g. nodes and edges of hydrology network). Each topological layer has its topological type (point, node, line, area) and a defined set of attributes.

24. Each GTDB object has at least the following standard attributes:

- unique and stable objectID.
- objectOrigin which describes where the data comes from.
- ObjectValue which defines the type of object (e.g. category of road).
- YearOfChange.

Additional attributes can be defined for topological layers.

II.5 Quality parameters

25. As the vector datasets of the GTDB are to be exported and used on arbitrary GI- and CAD-systems, few assumptions about the capabilities of these target systems can be made. Therefore, conditions to be met by the datasets are defined to minimize the risk of incompatibilities between datasets and target systems. Quality check routines enforce these conditions by marking errors for interactive-graphical treatement.

26. Since GTDB is host to existing as well as future datasets, there is a strong requirement for a flexible design. In order to make the quality check routines run for future datasets without further programming efforts, a table-based approach was chosen. The following quality parameters (a selection of a total of two dozen) serve to control the quality check routines. They are part of the semantic description of vector datasets:

- permissible attribute values (enumerations)
- minimal geometrical distances between vertices and nodes
- maximal dimensions of objects
- objectValues (e.g. road categories) that are allowed to meet in a node (e.g. road crossing)
- conditions which must be fulfilled i.e. where two edges are permitted to cross without node (e.g. bridge, tunnel)
- no two objects having equal geometry
- for exchange formats which do not support explicit topology (such as DXF), topology must be preserved implicitly when coordinates are rounded to a defined number of digits to the right of the decimal point
- no two edges joining in a node may have equal azimuths
- to prevent sliver-polygons, the perimeter-to-area-ratio has to be limited
- there are constraints to the objectValue of the boundary between two areas, depending on the objectValues of the areas (e.g. „shoreline of lake“ cannot be the boundary between forest and built-up area)

II.6 Other features of the data-dictionary

27. Apart from the above-mentioned features the data-dictionary also contains further ArcInfo-specific parameters to support the production process:

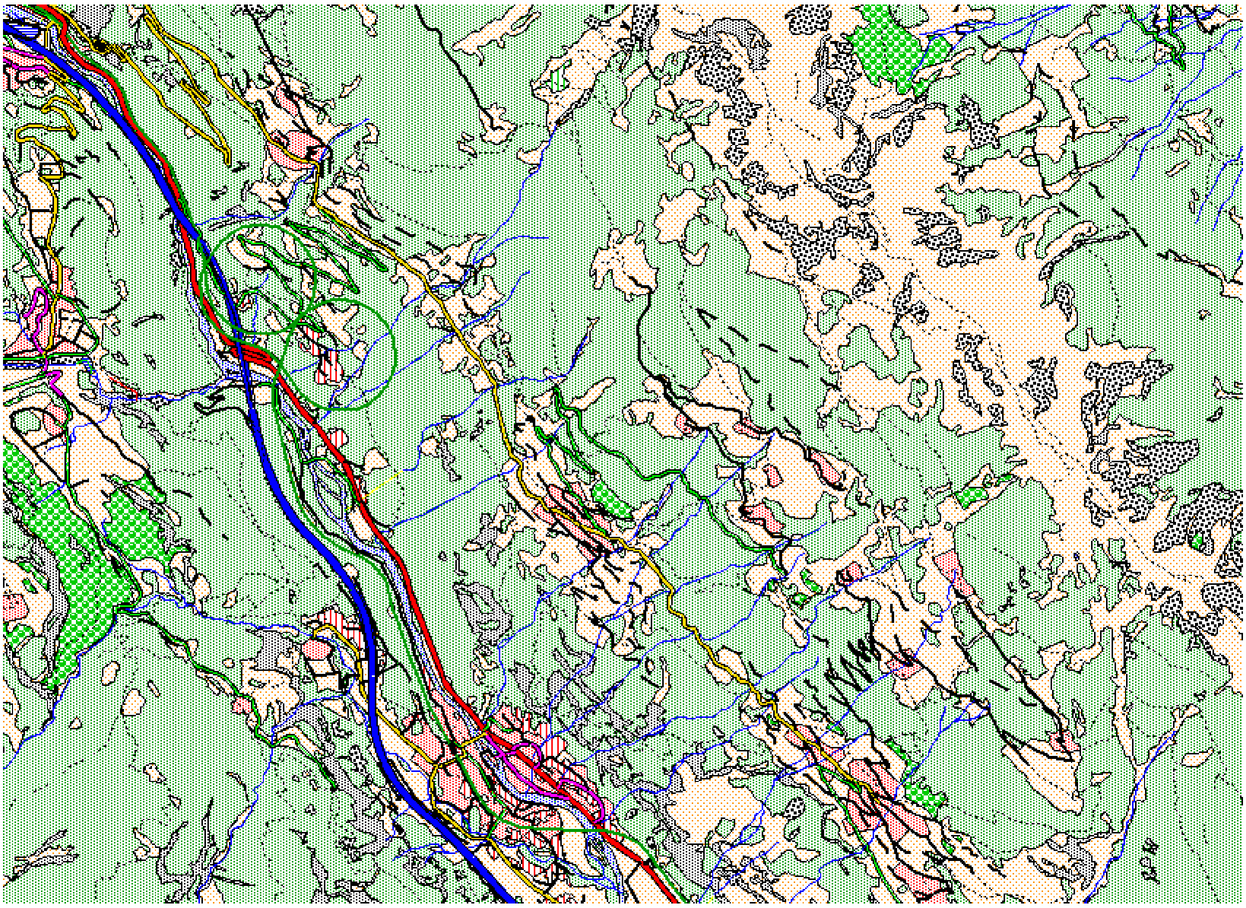
- parameters to control the system behaviour (e.g. snap tolerance)
- parameters to control the display of datasets (e.g. colors, line types).

The data-dictionary contains an ORACLE-sequence-mechanism which generates values for the objectID-attribute of every new object.

III. OUTLOOK

III.1 Amount of data

28. At the time of writing approximately 2 mio objects are integrated into the GTDB database (VECTOR25 Level 2). The following image shows VECTOR25 Level 2 south of the St. Gotthard area..



The system is expected to reach up to 10 mio objects. Further datasets (VECTOR200 Level 2) and additional layers in existing datasets (e.g. buildings) will be supplemented.

III.2 Improvements in the distribution system

29. The production and the modelling component of the GTDB are in productional use, whereas the distribution system at the moment still lacks a certain functionality. Tests have been undertaken to make GTDB-datasets available online within the Swiss Federal Administration's intranet (KOMBV). In the first stage this will happen using ESRI-specific protocols: Vector datasets in the SDE database are accessed from an ArcView-client. In the future, vendor-independent mechanisms such as the OpenGIS Data Model Architecture will offer open solutions. A HTML-based solution will be used to present and to provide an online ordering-system for geotopographical datasets in the Internet.

III.3 Support for INTERLIS

30. INTERLIS is a Swiss standard for semantical data modelling and exchange. It consists of a conceptual schema language and a system-neutral data transfer format. SFOT will extend GTDB by a mechanism to automatically derive an INTERLIS data description out of its data-dictionary. INTERLIS will then be used for distribution and long-term data archiving, as it is a vendor-independent, semantic-oriented, ASCII-based and well-documented format for exchanging and storing vector datasets.

III.4 Object comparison based on the unique objectID

31. The unique and stable objectID is a special feature of GTDB‘s datasets: It is the basis for comparison of different versions of a dataset; it allows for incremental updates and automatical detection of change in landscape.

32. The objectID will be preserved during the whole lifetime of an object. For this reason it may be used to reference GTDB-datasets on customer information systems. This will permit information from different providers to be mapped to the same vector base: one provider may have an information system to manage walking tracks, while another one has an information system for bicycle routes. If both systems contain GTDB’s objectID-attribute, it will be easy to find out on which road-segments hikers and bikers are likely to interfere.

III.5 Potential of GTDB datasets

33. So far each dataset has been treated individually. By combining GTDB-datasets and possibly datasets from other sources, there is a huge synergy potential: the hydrology network of VECTOR25 Level 2 already contains the key-attribute of Switzerland’s water information system (GEWISS).

34. Future development will include support for integration of semi-3D datasets (containing a height value per vertex) which originate in photogrammetrical measurements.

IV. CONCLUSIONS

35. GTDB is a GIS which provides a set of flexible tools to model, produce and distribute vector datasets. It stores landscape models in a nationwide seamless database and permits:

- an efficient update process
- integrated quality assurance
- safe backup and archive procedures
- improved documentation of datasets, which does not only contain structural but also semantic information (such as attribute values, geometrical constraints, topological constraints, textual description, visual presentation etc.)
- easy overview of available vector datasets
- decreased expense for distribution
- decreased time2customer.

The complete functionality of GTDB is available for future datasets without any additional programming efforts. Definition of the dataset in the data-dictionary will be sufficient.