PREPARING THE UPDATING OF THE
BELGIAN TOPOGRAPHICAL DATABASE ; A CHALLENGING PROJECT

J. Henrion, J. Beyen, J. Vanommeslaeghe
National Geographical Institute, 1000 Brussels, Belgium – (jhe, jbe, jvo)@ngi.be

Commission IV, WG IV/1

KEY WORDS: Photogrammetry, Updating, Three-dimensional, Database, Project

ABSTRACT:

There is more to preparing the updating of a topographical database than the conception of an appropriate data model. We needed to thoroughly study the available internal and external data sets. Further we needed to decide on how to get the best part from each data set, e.g. by writing a Visual Basic application for combining structured and identified 2D data with the z-values of the parent 3D data. This was the best solution for most object types. However, for some object types it was better to project the 2D data on a DEM. Moreover, it is very useful to know the number and kinds of modifications to be expected in the landscape. A good knowledge of both the available data and the modifications in the landscape and hence of the amount of work to do allows to define realistic requirements for the future: updating cycles, selection criteria, tolerances for absolute and relative accuracy. The confrontation of these requirements with the (continuously evolving) technical possibilities and with the production capacity of the available means for updating will determine detailed needs for development, investments and training. It is also the basis for outlining the procedures that should be followed in order to meet the requirements.

1. INTRODUCTION

In 2001, the National Geographical Institute (NGI) developed a strategical plan. In this context different strategical projects were started, one of which is called 'Seamless Geographical Information System of Reference' (SGISR). The objective of this project is to establish:
- a complete production line for collecting, treating and integrating update data in the topo-geographical reference data;
- a centralized, seamless GIS for stocking, managing, editing and distributing the topo-geographical reference data;
- the tools that are necessary for the NGI’s applications on basis of the topo-geographical reference data in the scales ranging between 1:10.000 and 1:50.000.

This objective is to be reached by the end of 2005. In view of the complexity of this project, nine associated projects have been started to deal each with some specific aspects, amongst which:
- External information sources, Change detection, Generalization and Updating. The present paper mainly deals with the updating project.

In 1988 the NGI started building extremely rich digital topographical reference data. These base data are to be complete in 2006. Most base data have been produced according to the following (simplified) workflow:

a) data classification: noting down in the field all semantical contents of the future base map on black and white aerial photographs
b) stereoplotting: drawing accurately the 3D geometry in a CAD-file, distinguishing about 40 different kinds of elements, i.e. a rough classification; the resulting ‘3D line’ is accurate and 3D, but not topologically cleaned
c) topological cleaning at the expense of loosing the z-values, detailed identification distinguishing over 300 codes + structuring, result: 'Top10V-GIS'.

A second production flow at NGI results in Top50V-GIS: seamless 2D data for the whole country on a 1:50.000 scale. 50% of these data result from the generalization of Top10V-GIS. The other 50% were produced separately. Top50V-GIS is being updated in a 5 years' cycle on basis of field survey and orthophotographs. Its second edition is to be finished in 2006.

One of the goals of the SGISR-project is to migrate from several, separately updated datasets to an object-oriented topographical reference database which is updated through one process. A major step to reach this goal will be to combine the best part from each data set into the database, by adding z-values to ‘Top10V-GIS’ and by eventually taking into account less accurate but more up-to-date information from Top50V-GIS.

The workflow for updating the reference data will differ considerably from the original datacollection. We will integrate external data that meet our quality requirements. Stereoplotting will be used to compare the database contents with recent images and to update the data. We intend to update the roads more frequently than the other objects and will use remote sensing for detection of most recent changes for roads. The in-the-field collection of data will be considerably different from the present data collection. The present in-the-field updating of the 1:50.000 seamless database by means of pen computers, provides valuable information on the performance of processes with pen computers and GPS; it will also soon be equipped with voice-command. The update rate and the in-the-field collection of data will also vary in function of the objects to be updated. Ideally the goal is a yearly update of roads, a two-yearly update of buildings and a five-yearly general update of the information in the database.

2. THE FUTURE DATA MODEL

The reference-data will be structured in an object oriented way. It takes good knowledge of the used definitions and selection criteria of all objects to be able to conceive an appropriate data model. The main difficulty lies in the conceptual choices to be made, part of which were treated by our project working group:
for answering the following questions we mostly based ourselves on the experience of the persons whom we thank at the end of this paper.

**a) How should we produce change only updates for clients?**

+ Which change metadata should be kept and on what level?

One solution for producing change only is by automatically comparing the data in the client’s old database with the data in the producer’s present database, as developed by Beyen & Henrion (1998) and by Badard (1999). IGN-France has continued in this way, refined the programs, adapted the output format to GML2 and built a system around it for distribution through the web. This method has many advantages: most important is that one needs neither unique identifiers, nor change metadata. On the other hand, this kind of software is not available on the market, so NMA’s have to develop it themselves and they have to provide with upgrades whenever one of the underlying softwares is upgraded.

The other, more classical solution for producing change only is based on the use of unique identifiers and change metadata, as developed by Galetto & Viola (1993, 1994) and Beyen (1994). OS-England, TDKadaster-Netherlands and CIRB-Belgium are amongst those who have chosen this solution. They differ in the amount and the structure of the change metadata.

For several reasons, our project working group has chosen the ‘classical’ solution, with unique identifiers and change metadata. We propose to only store change metadata that are necessary to keep record of ‘who modified which part of the geographical information, when and with what accuracy’ (multisource). We intentionally do not try to keep track of parent-child relationships, nor of detailed geometrical operations (like partial delete, move vertex etc.) Considering that the performance of the updating procedure itself is much more important to us than the query time for change only update data, we adapted the simplest possible structure allowed by ISO19108, and like CIRB-Brussels (structure indicated in bold in figure 1) we reduced the number of tables to a minimum:

The change only update data for customers can then be queried as follows:

- **Add** = the ID’s that appear only in the Active table
- **Mod** = the ID’s that appear both in the Active table and the Old table
- **Del** = the ID’s that appear only in the Old table (with end date > …)

If ever this solution shows not to work out as well as we hope (e.g. due to errors in the timestamping, loss of ID’s, double use of numbers, human errors caused by too complex procedures, clients being unable to restore topology after receiving change only updates, clients having modified the data structure …), we may always turn back to the solution comparing databases.

**b) Should priorities for performing operations on data be modelled?**

Here we take the advice of Prof. Peled to do so. We should not only follow the operations of the different updating operators, but also manage them. We intend to make a hierarchy of groups of operators for updating the geometry on the one hand and a second hierarchy for updating the attributes on the other hand: e.g. for identifying most objects a data classifier in the field should be higher ranked than a stereoplotting operator, but for drawing the right shape and position he should get a lower priority.

**c) May updates automatically propagate to neighbouring objects?**

We rely on Thierry Badard’s experience and his advice not to propagate updates automatically to neighbouring objects because of the high risk of propagating errors as well. A semi-automatic procedure is more appropriate.

**d) Where and when do we need consistency checks?**

It is obvious that the consistency checks during the data loading into the new database should be at least as severe as the ones during the updating later. Otherwise the updating operators who should concentrate on real modifications would be bothered by error messages for old inconsistencies that could already be solved before or by error messages due to unrealistic tolerances. Hakima Kadri-Dahmani (2001, 2002) distinguishes 4 different types of relationships (topology, composition, correspondence and dependency) and 2 types of constraints that allow to identify and solve conflicts (thematic and spatial constraints). We refer to her work for a more detailed answer to the question above.

### 3. STUDY OF THE AVAILABLE DATA

**3.1 The base map data**

The first edition of the digital topographical base map is to be completed in 2006, hence by then the first data will be 18 years old. It would be considered as reasonable if we could perform the first updating of the whole country in six years’ time. In that case the data would in average be eleven years old at the moment of their updating. It is beyond any doubt that in these conditions stereoplotting is the most economical method for updating (unless we obtain external update information that meets our quality requirements). However, for updating by stereoplotting the data have to be 3D.

We have two kinds of base map data at our disposal:

- **3D line** = MicroStation design files, resulting from the stereoplotting, with accurate 3D data that are roughly classified but not topologically clean and not structured; these data represent the situation of the landscape at the moment of aerial photography; the parcel boundaries are often snapped to roof borders, therefore their z-values are of poor quality, contrary to the other elements.
- **Top10V-GIS** = Arc Info coverages with data that have the advantage of being topologically clean and completely identified (with over 300 different codes) though that are 2D and a bit less
accurate in xy. It would be best to load these data in the future database, after recovering the z-values from the CAD-files.

The first 720 km², dating from 1988, had far less detailed contents than the data in the rest of the country. Furthermore, at the rate that the landscape has changed, there will be not much left unchanged. We therefore decided that the data collection in these areas should be redone completely.

During the following 15 years, from 1989 till now, the specifications for the data classification, the stereoplotting and the data structuring have evolved continuously. We retrieved nearly all instruction-notes with new specifications valid from a certain date, but not telling whether the map sheets that were being produced at that time already took in account the new specifications or not. Did the operators follow a new specification for the active or the next working area or only for the next map sheet? Even though we kept a record of all begin- and end dates of each major production step in each working area, we cannot answer this question for sure without verifying the data themselves.

The updating operators at every stage of the work will need to learn about the initial situation of the data in a certain area so that whilst performing the update they may pay attention to certain changes in the specifications. This illustrates the importance of metadata. The operators will need clear instructions on what they should adapt to the new specifications and what they should leave as it is.

3.2 The data for the 1:50.000 map

The 50K data cover the whole country and are already being updated on basis of field survey and orthophotographs before we start the ‘Seamless GIS of Reference’. By comparing the production years of the base map data with the first and second edition of the 1:50.000 map data, and sorting the results, we learned that because of the 15 years’ difference in age of the data, it would be most interesting to study the updates in the map 31/3-4. This study confirmed that it is very useful to glance over the 50K update data whilst updating the 10K database. Those data are also an important source of information on the number and kind of modifications that we may expect in the landscape.

3.3 The external data

At the moment, there are but very few external data sets that could directly be useful for our project. One of these is the high tension lines file from ELIA, the public utility which manages the transport of the high tension power in Belgium. The file was originally made by us, but it was accurately kept up to date by ELIA. Hence, it would be best if we could load their data into the future database. A similar co-operation is being organized with ELIA. Hence, it would be best if we could load their data into the future database, after giving z-values to the data. For many object types, the best way to do so would be to project the Top10V-GIS data on a very good DTM, e.g. produced by laser scanning. For some other objects, as well as for nearly all objects in the many areas where we do not have a laser scanned DTM, the most accurate method is by recovering the z-values from the parent CAD-files. This way, we combine the advantages of the two most important data sets: being topologically structured, identified and 3D.

For recovering the z-values from the parent CAD-files our colleague Hugues Bruynseels wrote a VBA-script that uses some predefined functionalities of ArcMap. In a first step, it converts the elements of the CAD-files into a cloud of 3D points in an ArcGIS 8.3 Geodatabase. Then the Top10V-GIS-coverages are imported into the same Geodatabase and a z co-ordinate (with default = 0) is given to each pair of xy co-ordinates. Since the original data were subject to manipulations during the structuring: small displacements, splittings, creation of additional nodes due to topological cleaning, the script needs to perform a proximity search: in a third stage, the application goes over every element of Top10V-GIS and for each encountered vertex it runs a proximity search and selects the corresponding triplet (x,y,z) from the point cloud. The distance in this search is a parameter; we have limited it to 10 meter. At 10 m the search is interrupted. If no corresponding triplet was found in the point cloud, the z-value is set to $z = -99$, in order to make it easy to detect the problem and correct it manually. The efficiency of this method varies with the object categories. It works extremely well for buildings, which solves our major problem. It does not work as well for the linear elements, though we hope to improve the results by generating additional points at intersections in the parent data. There will always remain a number of objects that have to be projected on a DEM. The best available DEM has to be chosen in function of the object type and the area.

5. THE FUTURE DIGITAL PHOTOGRAMMETRICAL WORKSTATION CONFIGURATION

For updating the future ‘Seamless GIS of Reference’ database by stereoplotting, we need our digital stereoplotters to be linked with a GIS. This is one of the most important requirements for the updating implementation; it may be a limiting factor in the choice of the underlaying DBMS. C. Heipke (2004) described how the ideal GIS should look like in general, from a photogrammetrical point of view. We would like to point out some more detailed requirements for the future stereoplotter configuration. It should answer positively to the following requirements. Many of them seem obvious, but we point them out explicitly because they are not as often available as one would expect.

It should be possible to (N=Necessary, D=Desirable):
N1) visualize the DBMS-contents in 3D
N2) edit the DBMS-contents in 3D
N3) automatically check input attribute values against their list domain or range domain and give warnings when not OK
N4) automatically timestamp each operation
N5) automatically source stamp each operation with a user group
N6) clean overshoots, undershoots and intersections, making a vertically projected 2D topology without messing up the z-values nor the db-linkages: in batch
D7) automatically adapt the corresponding stereo-images when performing a queued locate on the DBMS-contents
D8) shift the stereomodel in $x,y,z$ in order to make it locally coincide with the vector data
D9) visualize all elements of a 2D-file in 3D at the height of the...
floating point (cursor), e.g. for verifying external 2D data files
D10) automatically perform user defined consistency checks
D11) clean overshoots, undershoots and intersections, making a
vertically projected 2D topology without messing up the z-values
nor the db-linkages: dynamically – online, refreshing the cleaned
elements in the view without refreshing the images.

It should also be possible to easily create the following update functions:
N12) replace (or delete) the geometry of an object completely,
without loosing its attributes and automatically adapt ‘length’ or
‘area’ and ‘perimeter’
N13) update according to the following process:
- for deleting an object: copy the record from the Active table to
the Old table and fill in the end date and increment the record
number; then delete the corresponding record from the Active
table
- for adding a new object: add a record to the Active table and fill
in attributes values and geometry etc.
- for modifying an existing object: copy the record from the
Active table to the Old table and fill in the delete date and
increment the record number; then modify the geometry or the
attribute values of the corresponding record in the Active table
and adapt the load date, the source(s) and the source date(s).
D14) change a (group of) selected graphical element(s) from one
object class to another by one single mouse click; e.g. having
selected a group of paths, click on ‘become a dirt road’
D15) adapt the geometry of an object by ‘move_vertex’ or
‘move_the_whole_graphical_element’ and set the attribute
‘big_movement’ to 1 if the vertex or the whole graphical element
was moved by more than 5 meter.
We do not have the time nor the means to evaluate all possible
combinations of digital stereoplotters with databases. (In 2001 P.
Plugers already described 34 different models of digital
photogrammetrical workstations.) At present, our data are being
collected by digital stereoplotters SSK (ZI) using the CAD-
software MicroStation SE or J (Bentley). After transforming the
3D MicroStation designfiles into 2D ArcInfo coverages, the
topological cleaning and further identification are being
performed with ArcInfo 7 (ESRI). Regarding the licences and
maintenance contracts that we have, we started by examining the
solutions proposed by Bentley, Intergraph and ESRI.
1) MicroStation GeoGraphics (Bentley)
If it were possible to go on using the available Intergraph SSK
stereo software together with the used MicroStation CAD
software and realize the above points 1 till 14 by connecting
Microstation J through GeoGraphics 7 to a DBMS, we would not
have to purchase any supplementary licences (apart from the
DBMS), because the GeoGraphics module is included in the
maintenance contract. GeoGraphics 7 can be linked to Oracle
8iSpatial (It does not support Informix). For working with 3D-
data however, we need Oracle 9iSpatial or higher, which is only
supported by Microstation 8 + GeoGraphics 8, having a format
that is different from our design files. It is therefore impossible for
us to maintain the present data format for stereoplotting.
2) ISSD (ZI ) + MicroStation J (Bentley)+ Dynamo (Intergraph )
+ Oracle 9iSpatial
3) Geomedia Stereo (Intergraph ) + Geomedia (Intergraph )
4) Socet Set (BAE) + ArcGis 8.3 (ESRI)
5) LPS (Leica) + ArcGis 9 (ESRI)
The comparison of the different configurations is still in
progress.

6. HOW MANY AND WHAT KIND OF
MODIFICATIONS MAY WE EXPECT IN
THE LANDSCAPE ?

We should distinguish both the frequency of changes and the
importance that users attach to these:

We would like to distinguish 2 groups of objects: those of which
the updating can be considered as important on the one hand and
the less important group on the other hand. The ‘important’ group
(indicated in bold in table 1) contains as well objects of (at least)
low importance that appear in very great number in the landscape
as objects of very high importance that rarely change. For this
group we envisage shorter updating cycles than for the less
important group. The latter contains much information that can
only be collected in the field; hence it is more labour-intensive.
One could ask oneself whether it makes sense to go on
distinguishing pastures from arable land, seeing that changes
between these soil- ids are so frequent in Belgium that these
differentiation cannot be very useful. It might be better to abolish
this distinction in the future data catalogue.

Some numbers from our National Institute for Statistics: In
average there are 31 500 new buildings a year, which means that
during the first updating about 350 000 buildings will have to be
added, along with probably as many modifications to existing
buildings. There are also about 825 km of new public roads a
year, so we may expect more than 9000 km of new public roads
and an unknown number of private roads.

<table>
<thead>
<tr>
<th>Importance</th>
<th>Very low</th>
<th>Low</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very frequent</td>
<td>Arable vs. pasture</td>
<td>Gardens</td>
<td>Paths vs. dirt roads</td>
<td>Houses</td>
</tr>
<tr>
<td></td>
<td>Individual trees</td>
<td>Small watersurfaces</td>
<td>Road width</td>
<td>Roads, Woods</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hedges &amp; Tree-Rows</td>
<td>Churches, Town halls</td>
<td>Factories, Supermarkets</td>
</tr>
<tr>
<td>Average frequency</td>
<td>Forest type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare</td>
<td>Culverts</td>
<td>Slopes</td>
<td>Point symbols</td>
<td>High tension, Water courses</td>
</tr>
<tr>
<td></td>
<td>Sources</td>
<td>Water course width</td>
<td></td>
<td>Railroads, Churches, Town halls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hospitals, Schools</td>
</tr>
</tbody>
</table>

Table 1
7. DEFINING REALISTIC REQUIREMENTS FOR THE FUTURE

It is important to set tolerances for absolute and relative accuracy high enough to avoid that we should do over more than a small fraction of the measurements for the sake of accuracy. This means that we have to forbid the operators to replace data that are but inaccurate by less than 3.29σ (reliability 99.9%). σx, y, and σz are still to be determined by the geodesy department. Having 22.7 millions of objects, even then there will be 22 700 objects to be replaced for the sake of accuracy. Stereo-operators should understand that the latest aerotriangulation is not necessarily better than the previous one.

In practice the number of possible scenarios for updating is limited. We can either quickly update a limited number of objects or slowly update (nearly) all objects, or we can look for a combination. Depending on this choice, the need for personnel and means in the different services will take different proportions. The present field teams need at least 15 years to completely revise the whole country. It is not an option to adopt a 15 years’ updating cycle. Several studies show that users would like to receive updates of roads every year or even every half year, updates of buildings every 2 years, and updates of the other objects every 5 years. In our situation, it would be more realistic to adapt the updating cycle to the limitations in aerial photography: due to the atmospheric conditions, we may be happy if we succeed in covering the whole country with aerial photographs every 3 years. It needs no study to know that with the present production capacity we cannot update all objects within 3 years. On the other hand, 6 years is clearly not acceptable for a first update of the roads. Therefore we shall examine whether public roads, buildings and perhaps some other very important objects may eventually be updated every 3 years, and the other important objects every 6 years: Table 1 shows that apart from roads and buildings, the presence and position of big rivers and channels, railroads, high tension lines and woods are considered very important information. All this information can be updated by stereo interpretation. Hence these objects are ‘natural candidates’ for a quick revision. The only very important information that needs to be gathered by other means is the function of some buildings (factories, supermarkets, churches, town halls, hospitals, schools). In this scenario we give priority to the updating of very important and important objects. The field teams are only available for updating the less important objects when they are not needed elsewhere.

For assuring a constant map production level at long term, it would be useful to maintain a balance between a (very) quick revision of all (very) important objects in some areas and a slow but complete revision in some other areas.

8. CONFRONTATION WITH THE AVAILABLE MEANS FOR UPDATING

As mentioned in 3.1, it will not be possible to adapt all data within a reasonable time to the specifications that are being used at the moment. If we want to perform the first updating of the important objects in a reasonable time we shall have to temporarily accept a certain amount of heterogeneity in the data. For the less important objects, that will be updated in a very long cycle, we need to strive for homogeneity from the beginning, so that at the end of the first complete revision all data may be homogeneous, provided that we make the important objects homogeneous during their second or third updating cycle.

Up to now, only very few areas were partially updated for the most important objects that are visible on aerial photographs (about 700 km2). From this we learned that in practice a first partial updating by stereoplotting of all major objects is about 4 times quicker than the original datacollection and, if restricted to all roads and buildings it is even 6 times quicker. This means that our stereoplotting capacity is big enough for updating the roads and the buildings in 2–3½ years: by the end of 2008 we can have up-to-date data for all roads and buildings in the country, as defined in the requirements. On the other hand, we estimate that updating all other ‘important’ objects by stereoplotting would take at least eight months for half of the country. This means that in the first 3 years, we cannot succeed in maintaining a constant map production covering half of the country along with the updating of all buildings and roads, without using the additional surveying capacity. In the best case we may free a stereoplotters’ capacity of 6 months by updating all roads by surveying, after having detected the areas to be measured by remote sensing. This would leave us a whole year for updating the remaining ‘important’ objects in half of the country. The remaining surveying capacity may even be used for keeping the roads up-to-date in a shorter cycle, depending on the availability of satellite images. This programme may be a little bit too ambitious, but it is not very far away from feasibility either.

If the surveyors are needed for other tasks so that we cannot count on them for measuring roads, we may be forced to postpone the updating of some ‘important’ object types to the years 2009–2011. We expect the external data to become more important during (and after) this period.

Another possible scenario for 2006–2008 is of course maintaining a constant map production by increasing the ‘important’ objects at the expense of making the first updating cycle for roads and buildings a few months longer. However, this would not reduce the long updating cycle of the remaining objects that have to be collected in the field. The latter can only be reduced by enlarging the expensive field teams and by finding new sources for a part of the information. Which scenario NGI will follow has not been decided yet. We do not recommend to slightly increase the stereoplotting capacity for only a short period due to the considerable training efforts that this would imply.

Figure 2. R = roads, B = buildings, VI = very important objects, I = important objects, O = other objects; different colours represent different updating cycles.
The exact needs for investments and training will be determined in function of the chosen hard- and software for every stage of the updating process. We foresee to purchase a number of pencomputers, GPS-RTK receivers and a DBMS. The data classifiers will have to learn to use all these. The training effort will be most important for this group. On top of the new technology, like all other operators in the production line, they should learn to deal with the three-dimensional and the temporal nature of objects.

**CONCLUSIONS**

We made choices for the updating aspects of the data model and determined criteria for choosing the future digital photogrammetrical workstations. We wrote a VBA-script to give z-values to the structured and identified data. The data and especially their accuracy are still being studied in order to define the future tolerances. The confrontation of the expected amount of work with the available means shows that, on condition of investments and training, our total stereoplotting and surveying capacity suffice to reach reasonable updating cycles. The exact needs for investments and training will be determined in function of the chosen hard- and software for every stage of the updating process.

**REFERENCES**


**ACKNOWLEDGEMENTS**

We would like to thank our colleagues at NGI who revised the present paper, as well as all colleagues from other National Mapping Agencies who shared their experience and their ideas with us, especially Thierry Badard and Hakima Kadri-Dahmani (France), Bert Kolk (the Netherlands), Ammatzia Peled (Israel), Sue Sleath (England). We would also like to thank Claude Hannecart and Yves Lardinois (CIRB – Brussels) for allowing us to copy some of their ideas.