GIS VERSIONING MANAGEMENT – THE APPROACH OF THE SURVEY OF ISRAEL

Yaron Felus[ab], Haim Srebro[a], and Yoav Tal[a]

[a] Survey of Israel, Tel Aviv Israel - (felus, haim, tal)@mapi.gov.il
[b] Ferris State University, MI , USA

ICWG IV/VIII

KEY WORDS: Versioning, Unique Spatial Identifier, Spatial Data Revision, History tracking, GIS, NSDI, Survey of Israel

ABSTRACT:

In the last two decades, there has been considerable progress in mapping and spatial data collection and today many developed countries are completely mapped. The three main challenges for National Mapping Agencies in these countries are:

- How to effectively update and revise these databases with new information about our rapidly changing environment
- How to integrate these geospatial data in the national databases utilizing standard relationships for data interchange
- How to manage the history of changes and to manage past versions of datasets

The following report describes the new approach of the Survey of Israel (SOI) for spatial data consolidation and for map revision. This approach is based on a unique spatial identifier given for every spatial feature in the database. The properties and maintenance of this identifier will be discussed along with various options that were investigated. Using this unique identifier SOI developed an algorithm to create Add, Delete and Update tables. These tables present an efficient method to store and distribute revisions and updates of the national spatial database to various organizations. The approach also allows for efficient documentation of past changes and the maintenance of historical spatial data thus creating a spatial-temporal framework for databases that includes a time-stamp versioning options. The significance of this approach to the Israeli National Spatial Data Infrastructure will be assessed using examples and a case study.

1. INTRODUCTION

1.1 Background on the Israeli National Spatial Data Infrastructure

The Survey of Israel (SOI) national spatial database (NSDB) was established in the early 1990s (Peled and Adler, 1993, Peled, 1994) and the basic principles that were set at the time for the creation of this database are still relevant for modern applications. These principles include:

- A seamless and homogenous data base spanning across the entire country with equal content and uniform positional and attribute accuracy.
- The NSDB is acquired using a combination of photogrammetric methods from aerial photography at 1:40,000-scale, and field surveys collecting information not interpreted on aerial photographs and attribution
- The data is divided into the following thematic layers or datasets: Buildings, Transportation, sites, Infrastructures, Land cover, Relief, Hydrography, Geographic Entities, Nautical Information, and Cadastral data. Each dataset is further divided into feature classes at the time by using different codes.
- The NSDB will be updated and revised in predetermined cycles.

To ensure the uniform quality of the database, a comprehensive process was developed to create this database that included the following steps: aerial photography, aerial triangulation, stereo feature-collection, quality control (Interpretation, accuracy, and topology), data revision and integration of new data into existing NSDB, data enhancement using field surveys, and data distribution.

This process was successfully used for almost twenty years and the changes that transpired in it were mostly in the technologies that are implemented in the operations. The aerial triangulation procedure became more automated with new digital photogrammetric stations (DPS) collecting tie-points and performing the adjustment more efficiently. The stereo feature collection procedure turned to be more efficient with DPS that employ superpositioning techniques to display vector data on top of imagery while updating the data. Quality control, data integration and enhancement has transformed into a set of computerized batch process performed with new GIS software, and the distribution of data became completely electronic using on-line geographic portals (see also Srebro, 2009).

1.2 Redesigning the NSDB

While much of the core principles of the NSDB remained valid, technological advancement and the development of new spatial applications necessitated the implementation of major modifications to this database. These technological advancements include:

- The integration of spatial data in organizational IT systems such as Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) systems
- The emergence of Location Based Services (LBS) technologies that employ efficient navigation software, and mobile computing
- The development of popular Web geospatial visualization and mapping applications, such as Google Earth and Microsoft Bing Maps Platform
- The global development and considerable expansion of national and sub-national SDI technologies and regulations.

The Survey of Israel followed the latest technological development and the changing needs of the geospatial community and consequently initiated a set of activities aimed at redesigning its spatial database structure, specifying a new set of geospatial products and modifying the spatial and attribute content of the Israeli NSDB. While these activities are still in progress the following key guidelines were agreed upon for the implementation of the new NSDB design:

1. The transportation and geographic entities datasets will be designed in the spirit of the Geographic Data Files (GDF) standard, referred to as GDF version 4.0 (ISO 14825:2004).
2. The database will be structured according to modern principles and will be stored in the latest computer format (ArcSDE geodatabases in ST-Geometry data type which is OGC/ISO 13249-3 compliant)
3. The land cover data set will be adapted with respect to the classification methods to meet international standards such as the Food and Agriculture Organization of the United Nations specification (FAO, 2009 or the work of ISO/CD 19144-2).
4. A new set of database procedures for incremental updating and versioning will be used. These procedures will take advantage of a unique Spatial Identifier (SID) which will be assigned to every feature in the NSDB and a comprehensive system of rules that will be used for the assignment of the SID.
5. A modern metadata system was designed and currently being implemented, to support the NSDB documentation and integration. The metadata was developed in cooperation with the Israeli Institue of Standards, according to the ISO specs 19115.

In the following sections, we will further describe the unique Spatial Identifier, how it will be assigned and the effect that the inclusion of this field will have on the SOI NSDB updating procedures and on the users of the SOI NSDB. This paper is organized as follows: Section 2 reviews the key issues involved in the creation of a spatial identifier. Section 3 explains how the concept of SID supports efficient procedures in incremental updating and versioning. Conclusions and final remarks are presented in Section 4.

2. A UNIQUE SPATIAL IDENTIFIER (SID)

2.1 Introduction to the SID

The concept of a unique identifier for every record of information is inherent in the application of databases and was developed in the early days of relational databases where the term primary key was used (Codd, 1970).

The same concept of primary key was used by SOI in the development of the Israeli NSDB and every record in the database got a unique primary key to identify it. Nevertheless, relying on the Database Management System (DBMS) an identifier alone proved to be problematic both from a technical and a conceptual point of view. Technically, the DBMS unique key was unstable and changed with major initializations of the system. Moreover different spatial operations such as split or copy created duplications. These technical problems necessitate the SOI to develop a stable and supervised method to create and maintain the uniqueness of a primary key termed as a Unique Spatial Identifier. Conceptually, the spatial Identifier was not recognized by all levels as a fundamental component for spatial data revision, integration, and analysis with the possibility of historical examination.

A few organizations have developed methodologies for the creation and maintenance of SIDs, some examples include:

- the UK Ordnance Survey with the Topographic Identifier (TOID) consisting of a 4-character prefix followed by 16 “numeric” values as described in DNF &OS (2005)
- the Swiss Geo-Forum with the object identifiers (OID) consisting of 8 character prefix followed by 8 numeric values as outlines in the INTERLIS 2 specification for data exchange at INTERLIS (2009)
- Yahoo GeoPlanet that employ a 32 bit ~ 4×109 number termed WOEID (Where On Earth ID) as a unique identifier for geographic entities as described in Yahoo GeoPlanet (2008).

For completeness, we will note also the ISO 19112 model which sets geographic Identifiers for place names in a gazetteer system. In the non-spatial context it is important to mention the globally unique identifier or GUID identifier used in software applications (Microsoft, 2010) which is unique in any context (hence, "globally"). While each generated GUID is not guaranteed to be unique, the total number of unique keys (2128) is so large that the probability of the same number being generated twice is extremely small. Nevertheless, since the GUID number is so large it is difficult to manage a NSDB using GUID and this option was immediately rejected.

These various efforts were studied and the following methodology for the creation of the Israeli SID was developed.

2.2 Properties of the Israeli SID

The following guidelines were set by SOI for the establishment of the SID:

The SOI SID will be a Long Integer data type (i.e., 4294967294 values) and will be assigned and created separately for each dataset in the NSDB. This makes the SID number large enough to accommodate for future growth in the number of features in the NSDB and still each one to be handled by software search and sort functions.

A 4-character prefix will be attached to the SID in some applications where the datasets are merged. The prefix will be assigned specifically for an organization and a dataset, for example, SIBL will be assigned for the Survey of Israel Building dataset.

The assignment of the SID will have no intelligence or logic in it (e.g. location indication).

The SID will remain with the feature throughout its existence and will not be reassigned to a new feature when the feature is deleted. In general, changes in the spatial feature will not change the SID. This principle requires that in most cases the SID will be stable throughout the feature lifetime. If a certain topographic feature such as a building is demolished and reconstructed, but keeps the same attributes and usage, it will maintain the same SID.

Only in exceptional cases, for example, in a complete modification of a neighborhood, will changes in the features instigate a change in the SID.

SID will be assigned to the entire database including: topographic features (e.g, buildings, road segments, point of interests), virtual features (e.g., parcels, statistical zones, turn restrictions and contour lines) and aggregated features (blocks which are composed of parcels, or building complexes).

In order to maintain continuity for spatial objects that were split or merged, it was necessary to add another field, termed the PARENT field. The PARENT field is used if a certain feature is created from another feature for example a road that was split into two roads or two parcels that were united into one. In both cases the new features will get a new SID, however the PARENT field will preserve the associated link to the old feature by having its SID.

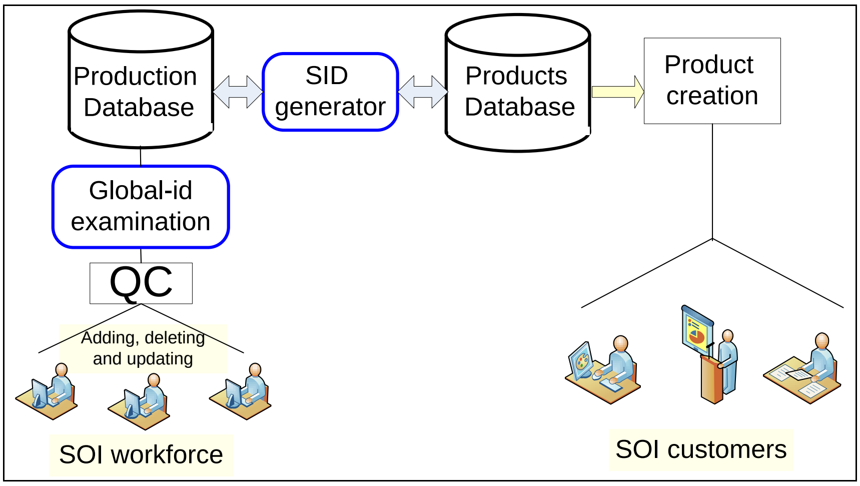

Figure 1: The relationships between the production and the product database.

2.3 SID Creation and Management

During the discussions on the implementation of the SID, it became evident that the creation of new SIDs should happen at the end of the data production process. The SID generation function will be installed as part of the process of converting the production database (where all the editing and revision operations are being performed) into the product database from which data products are derived (See Figure 1).
The SID generation process should be autonomous and verified by a few independent functions.
The SID generation algorithm works by comparing the old database with the new database and using the SID field (see Appendix I). The algorithm creates the Add, Delete, and Update tables and identifies four options for a spatial entities:

1. Spatial entity is created and thereafter a new SID should be created, the entity should be added to the Add table with a null PARENT field.
2. Spatial entity is abolished and therefore the entity SID is marked as non-active and the entity should be added to the Delete table
3. Spatial entity is modified, the SID is not changed and the entity should be added to the Update table
4. Spatial entity is created from the split or merge of an old entity, thereafter a new SID should be created, the entity should be added to the Add table with a PARENT field that is equal to the SID of the old entity.

The Add, Delete, and Update tables are the Change-Only update files which will be given to the SOI NSDB users who wish to incrementally update their database.
The creation of the SID only at the end of the database production process presented a problem because index numbers play an important role also in the data production and creation process. These index numbers are used to associate additional data to topographic features. For example turn restriction information is associated with two road segments using their spatial index, or addresses are associated with buildings. Therefore, it was agreed that another index number should be used within the production database. This index will be automatically generated by the GIS software. This is a registry style string within a geodatabase which is termed the globalID. Consequently, SOI surveyors, data collection operators or GIS editors will use the globalID for their work and leave the SID as is or set it to NULL for new features. They will be responsible for the assignment of the PARENT field which will be carried out using efficient data collection tools.
Another important reason for the need for the globalID index is the requirement to work in a distributed database environment. Distributed database environment will allow the SOI to update its database using multiple contractors, mobile editors, and work stations that may be physically disconnected from the NSDB but may work on the same area during the same time.
In the past, SOI used the Cut and Paste approach to perform multiple and parallel edits on the NSDB. In the cut and paste approach each updating operator/ contractor used to get a chunk of the data (1:50K sheet size), lock it to prevent other from using this chunk and update it. The globalID mechanism will allow the SOI to use Application Replication approach (also known as Geodatabase replication) where changes are controlled and synchronized by the GIS software. Application Replication enables incremental updating of all the spatial database content including topologies and geometric networks but it requires that all the parties use the same advanced GIS software (ESRI, 2004). For that reason the Application Replication approach will be used only internally in the SOI and by its contractors.
The presented algorithm for the generation of the SID and the Change-only Updates will be comparable to the Database Replication approach. While there are some limitations in terms of the ability to update topologies and geometric networks, the Database Replication allows incremental updating which is independent of the GIS application. In the following section we will list a few other benefits of the SID concept and the COU methodology..

3. UPDATING, VERSIONING AND HISTORY TRACKING PROCEDURES

3.1 Creating a national common language using the SID

The spatial identifier is the key tool for data sharing and data integration. Today many information systems and governmental databases use the address as a location identifier. Nevertheless, an address is not unique, some buildings have a few addresses (i.e., corner buildings), some have no address (i.e., in remote places) and a couple of buildings may share the same address (i.e., in a university

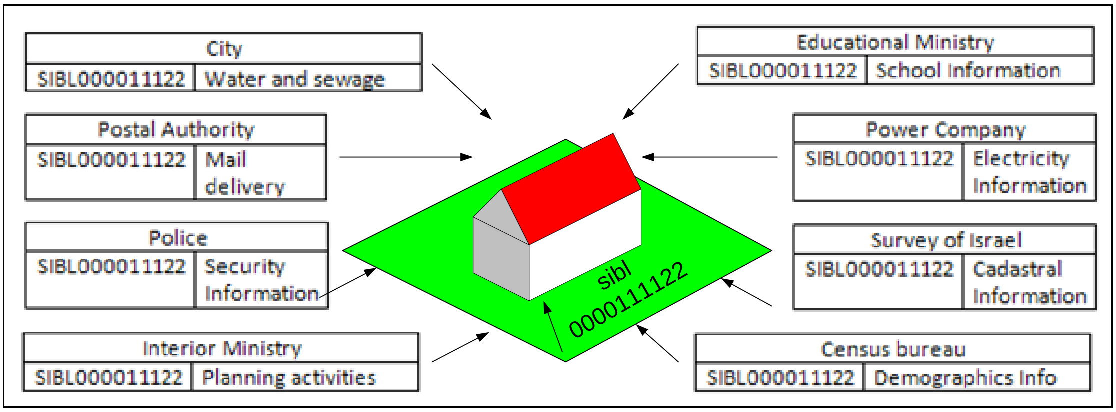

Figure 2: The SID creates a common identifier to bridge across databases and platforms.

complex). The need for a common identifier for structures, road segments, junctions was brought up by many governmental organizations (security, post-office, ministry of interior) as a necessity for daily operations.
Implementation of the national SID concept will solve these problems and create a common language to link data between organizations in the SDI-NSDI. Namely, government agencies would be able to combine individual spatial databases into a system that contains a comprehensive, multidisciplinary file on every structure, road segment, or any other topographic feature (Figure 2).

3.2 Updating and versioning

Implementation of the national SID concept will enable better procedures for data updating and versioning. SOI users compile, enhance, and build their own value-added, specific application database using the NSDB. The user's primary concern is their specific application data. In the past SOI provided updates in bulk, with the updated data set replacing the old one. Users should have had re-integrated the update in their systems rebuild the links to their value-added, specific application data.
The SID will allow users to easily integrate updates of the NSDB without losing the critical link to their data. Furthermore, the addition of the PARENT field will allow the users to attach the data to features that have changes in split or merge operations.
The SID management procedure creates the Change-Only updates files. These files are much smaller than the complete database (in an order of a few gigabytes) and therefore will allow the SOI to send the Change-Only updates files via internet communication.

3.3 Historical tracking

The ability to track the history of a certain location was always important as a mean for understanding the environment and diagnosing the sequence of events that created the current situation. An existing environmental hazard, for example, can be created by a pollutant that was created by a gas-station that was abolished many years ago. Nevertheless, only recently, it has become possible to visualize time and to track chronology (the order or sequence of features through time) using off the shelf commercial GIS (ESRI, 2010).
The SOI is prepared to meet this challenge and supply its customers and the State of Israel with the ability to track history. The Change-Only Updates (COU's) and the PARENT field will be the keys to this ability, by storing and archiving these COU layers and maintaining the PARENT field it will be easy to create a time series that presents various changes (Additions, Deletions and Updates) that occurred in a certain location.

4. CONCLUSIONS

The SOI has initiated a series of major steps to change its data products to better meet the needs of its customers in the government and in the public and private sectors. Each step will require considerable efforts in developing software, assimilating new procedures, and managing new operations. Moreover, the SOI will need to work with the customers to enable them to take advantage of these improvements and to assimilate this advanced spatial database in their systems.
It is believed the SID concept will be a corner stone in a much stronger National Spatial Infrastructure for the State of Israel

ACKNOWLEDGEMENT

Part of the presented work came from the fruitful discussions and joint work with Dr. Basheer Haj-Yehia, Moshe Yaniv, Hadas Sagi, Baruch Shirion, Yaron Levi, and Adi Neuman. The assistance of Yohanan Gavis is greatly appreciated.

REFERENCES

Beyen J. and Henrion, 1998. Updating Topgraphic Databases with Arc Info; Client-fitted creation of change-only information. In D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany pp 59-64, also at: http://www.ifp.uni-stuttgart.de/publications/commIV/beyen9neu.pdf

Codd E. F., 1970. A Relational Model of Data for Large Shared Data Banks, in Communications of the ACM, Vol. 13(6), pp. 377-387

Cooper A.K. and A Peled, 2001. Incremental Updating and Versioning, 20th International Cartographic Conference, Beijing, Vol 4, pp 2804-2809, summary at: http://researchspace.csir.co.za/dspace/bitstream/10204/1782/1/Cooper3_2004.pdf

DNF& OS, 2005. Unique Object Identifiers within the Digital National Framework (DNF), DNF Expert Group and Ordnance Survey at: http://www.dnf.org/Papers/20051114%20Unique%20Object%20Identifiers%20v200.pdf

ESRI, 2004. Versioning, An ESRI Technical Paper at: http://downloads2.esri.com/support/whitepapers/ao_/Versioning_2.pdf

ESRI, 2010. Time Awareness in ArcGIS 9.4 Leads to Better Understanding of Complex Geographies Visualizing Time in GIS, ArcNews Winter 2009/2010 at: http://www.esri.com/news/arcnews/winter0910articles/visualizing-time.html

FAO - Food and Agriculture Organization of the United Nations, 2009. Land Cover Classification System dichotomous at: http://www.africover.org/LCCS_dichotomous.htm

INTERLIS, 2009. Object Identifier (OID) – A Data Exchange Mechanism for Land-Information-Systems at: http://www.interlis.ch/oid/oid_e.php

ISO 14825, 2004. Intelligent Transport Systems –Geographic Data Files – Overall Data Specifications, ISO TC204 at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=30763

Microsoft, 2010. GUID Structure, Windows Developer Center webpage at: http://msdn.microsoft.com/en-us/library/aa373931(VS.85).aspx

Peled, A., 1994. Remapping a Country - The Israel National GIS, Conf. Proceedings, ACSM/ASPRS 1994 Annual Convention, Reno, NV., USA, Vol. 1 (ASPRS), pp. 245-249. at: http://libraries.maine.edu/Spatial/gisweb/spatdb/acsm/ac94059.html

Peled A., Adler R., 1993. A Common Database for Digital Mapping and GIS. International Journal of Geographical Information Systems 7(5): 425-434.

Srebro H., 2009. A Status Report of the Activity of the Survey of Israel, FIG Article of the Month March 2009 at: http://www.fig.net/pub/monthly_articles/march_2009/march_2009_srebro.html

Yahoo GeoPlanet, 2008. Where On Earth Identifier (WOEID), Developer Network API Documentation at: http://developer.yahoo.com/geo/geoplanet/guide/concepts.html#woeids

APPENDIX A

A flow chart of the algorithm for SID and ADD, Delete and Update tables generation is presented in Figure 3. The input for the process are the old database sorted by SID, the new database sorted by SID and LN the largest SID number in the old database.
The example in Figure 4 demonstrates how the algorithm is employed.

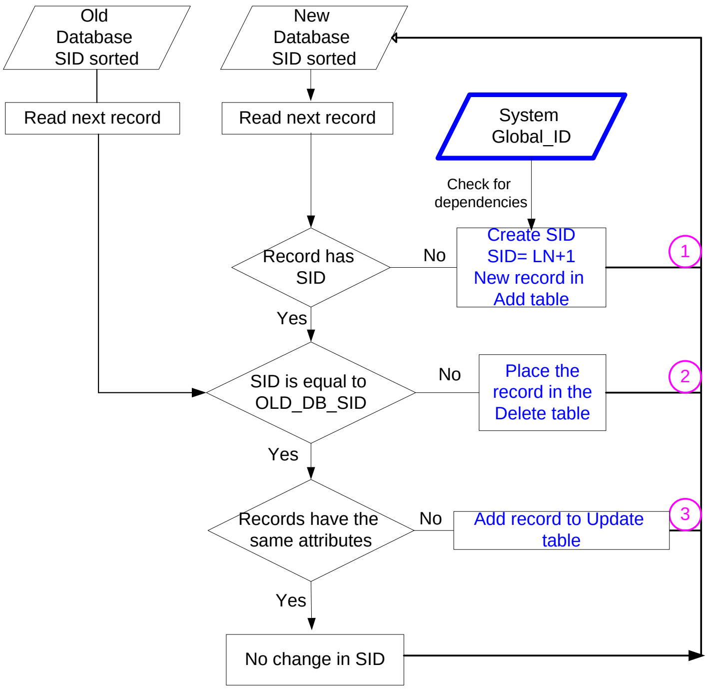

Figure 3: The SID generation algorithm

Old Database (01/01/09)

SID	Parent	Area	X	Y	Z	Street _name
11	-	250	8000	370	820	Gefen 11
12	-	90	8200	320	890	Deror 7
13	8	380	8120	310	870	Herzel 1
14	9	170	8400	370	810	No name
15	-	450	8800	360	850	Sharon 4

Revised Database (30/03/09)

SID	Parent	Area	X	Y	Z	Street _name
11	-	250	8000	370	820	Gefen 11
13	8	380	8120	310	870	Herzel 1
14	-	430	8100	380	815	Zait 78
15	-	490	8900	370	850	Sharon 4
0	12	220	8000	300	830	Tamar 9

New Database (01/04/09)

SID	Parent	Area	X	Y	Z	Street name
11	-	250	8000	370	820	Gefen 11
13	8	380	8120	310	870	Herzel 1
14	9	430	8100	380	815	Zait 78
15	-	490	8900	370	850	Sharon 4
16	12	220	8000	300	830	Tamar 9

Update Table (01/04/09)

SID	Parent	Area	X	Y	Z	Street _name	Geometry
14	9	430	8100	380	815	Zait 78	shape
15	-	490	8900	370	850	Sharon 4	shape

Delete Table (01/04/09)

SID
12

Add Table (01/04/09)

SID	Parent	Area	X	Y	Z	Street name	Geometry
16	12	220	8000	300	830	Tamar 9	shape

Figure 4: Example that demonstrate the generation of SID and the creation of Change only Update files