

QUALITY CONTROL FOR CROWDSOURCING LARGE SCALE TOPOGRAPHIC MAPS

Irit Bekker¹, Yaron Felus¹

¹Survey of Israel, Lincoln 1, Tel-Aviv, ISRAEL, (iritb , Felus)@mapi.gov.il

KEY WORDS: Crowdsourcing, Mapping Standards, Map Conflation, Quality Control

ABSTRACT:

Crowdsourcing is an emerging methodology for spatial data collection. This methodology has been used extensively to create small scale maps such as in OpenStreetMap. However, it has rarely been used in acquiring large scale, accurate engineering data. Expert-sourcing uses methods similar to crowdsourcing except that it collects information from experts and professionals. This article describes the development steps of novel expert-sourcing methodologies for updating the national geo-spatial database (cadastre and topography) in Israel. The input data for the process are building permit maps at a scale of 1:250 created by professionals. The first step in the process was the adaptation of a national standard for digital map contents, cartography and computer format. The second step was the development of the crowdsourcing system to get the maps from mapping professionals, assess them and verify that they meet the criteria. The two challenges in this step were to design an engaging User Experience (UX) that will not require training or a user manual and to develop automatic quality control tools. The third step was the creation of a data integration process to update the national database from the various sources. The system (aka TopoCad) provides a comprehensive set of tools to support the practice of obtaining information from all the surveyors and mapping experts in Israel. The system aggregates all the information of every mapping activity (geodesy, cadastre, construction permits, mapping and GIS) in a semi-automatic process. This novel platform improved governmental procedures and made services more efficient. In addition, this innovative tool improved the national databases by updating them with new data on an on-going basis.

1. INTRODUCTION

Internet technologies have affected the way National mapping Agencies (NMAs) operate and work. At first the internet provided platform for NMAs to distribute their data and share their maps on-line (Web 1.0). An overview of the Wikipedia list of NMAs shows that many of them have an active geospatial portal (Wikipedia 2018). Currently, many NMAs are investigating the use of Web 2.0 to improve their operation using crowdsourcing methodologies. Some examples include: The US Geological Surveys (USGS) Volunteered Geographic Information (VGI) project that encourages citizens to collect and edit data using the National Geospatial Portal about man-made structures to improve the USGS authoritative spatial database (McCartney et al. 2015); The Dutch Cadastre in the Netherlands and the Finnish Geospatial Research Institute which are looking into the use of crowdsourcing methods to enrich their topographic database as reported in Bol et al. (2016); and the Vicmap Editing Service (VES, at: <https://ves.land.vic.gov.au/login>) encourages registered public users to notify the Australian state of Victoria of changes required to the Vicmap core spatial data products.

In the same manner the Survey of Israel is using its national geospatial portal to get citizens feedback on its national map at <https://www.govmap.gov.il/?lang=en> and investigate crowdsourcing methods to map defibrillators at <https://medman.co.il/>.

However, none of the reported NMA's activities fully employs crowdsourced data in a systematic manner and most NMAs are cautious about integrating crowdsourced data with authoritative data as this may reduce the quality and consistency of their national datasets (McLaren, 2012).

A new approach is presented in this paper. This approach (expert sourcing) employs experts (professional surveyors) instead of the crowd (Non-Professionals). This new approach is implemented using a computerized system (Topocad) which will be presented in this article. The system consistently collects spatial data from mapping experts to update the national cadastral and topographical database. The previous version of the system had tools for searching and retrieval of geodetic, cadastral and topographic information of the Survey of Israel (Felus et al., 2013). The second version of the Topocad system includes expert-sourcing tools to upload, verify and integrate all types of surveying data and maps. In order for the approach to work, these surveying data and maps should be submitted by the crowd of mapping experts in a uniform machine readable format. A new set of mapping standards were published by the Survey of Israel to ensure that every map in every step of an engineering project is made in a computer readable manner. The topocad system reads these crowdsourced files and integrates these in the national databases following a rigorous quality control process.

The system with its crowdsourcing tools is now working and will be described in this article. The continuation of this article is organized as follows:
Section 2 reviews the institution of new mapping standards in Israel. These standards have created of a common language as a basis for expert sourcing.
Section 3 describes the development of expert sourcing tools with their user interface.
Section 4 explains the quality control process and the data conflation tools. Section 5 provides an overview of Conclusions and final remarks are presented in Section 5.

2. CREATING A COMMON LANGUAGE BY MAPPING STANDARDS

Engineering projects are characterized by a large number of spatial data files and large amount of information which is generated in every step. These data files are generated by engineers in different fields (architecture, construction, transportation, water, electricity...) but mostly by licensed surveyors who provide the basic map to overlay the data layers. Moreover, different steps in the development of an engineering project create different types of data at different scales and content. According to the Israeli Planning and construction Law and regulations there are 5 steps in an engineering construction process which requires the involvement of a licensed surveyor. Those steps are:

1. Planning, where a master plan of the project area is made for the local authority approval (a map at a scale of 1:1250)
2. Registration, where a subdivisions plat used to register the cadastral division of parcels at a scale of 1:1250
3. Building permit, where a building permit map at a 1:250 scale is prepared with topographic data, precise cadastral information, above and underground infrastructure and the exact location of property corners and construction limits.
4. Construction, where the following plans are prepared
 - a. A stakeout plan with coordinates of the foundation markers and building columns.
 - b. Ground floor plan used to monitor the location of the ground floor
5. City approval of the constructed project where an as-built survey map is drawn to document the engineering activity.

In addition, most of the local authority are preparing maps for their Geographical Information system. These GIS's are supporting efficient management of city infrastructure (sewage, lighting, water, power, gas, communication...) and operation (garbage collection, public safety, street cleaning...).



Figure 1: The life cycle of an engineering project that require map submission.

Each of these steps in the life cycle of an engineering project (as presented in figure 1), require a surveyor or mapping expert to supply a map for the authorities for validation and approval. So far these maps were submitted in different format,

sometimes even as paper documents, making it impossible use them for updating geo-spatial databases.

In order to be able to use these maps to update national database, the survey of Israel, the Israeli National Mapping Agency, led the Inter-Government GIS committee to approved the national mapping standard. The Inter-Government GIS committee was established in 1991 by Governmental order as the committee with representatives from 45 governmental ministries and public organizations.

The national mapping standard is based on list of surveying and mapping regulations and instructions. It provides a comprehensive set of definitions and descriptions for the content, data acquisition accuracy, computer format, and cartography of each map in the life cycle of an engineering project. The file format is CAD based format (DXF) and a list of layers and blocks detail the exact structure of the file (see Figure 2). For example, building polygons are placed in layer 2200, with block 2200 which has 5 attributes (building name, street address, building no., no. of floors, type).

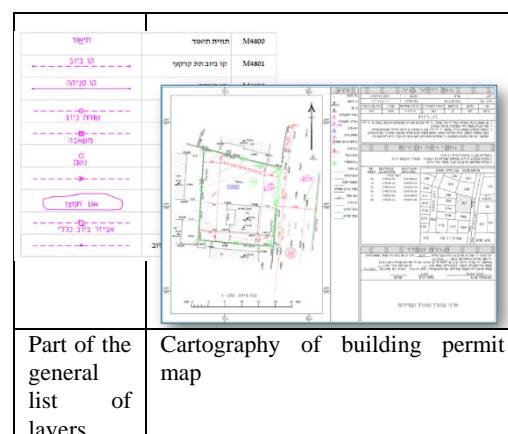


Figure 2: Excerpt from the national mapping standard

The national mapping standard is based on the Surveying Regulations which defines the accuracies and the data collection methods also presented in Table1 (Felus, Keinan, Regev, 2013)

Accuracy level	RMSE in Easting or Northing (m)	Largest scale	RMSE of Elevation at well-defined points	Contour line interval	Use
3	0.06	1:250	0.08	0.10	Building permit and As-build map
4	0.13	1:500	0.15	0.25	Detailed planning
5	0.25	1:1,000	0.30	0.50	Subdivision plat
6	0.30	1:1,250	0.38	1.00	Detailed Zoning map
7	0.63	1:2,500	0.75	1.25	City master plan

Table 1: accuracies of mapping products from the Surveying Regulations

In order to make the national mapping standard widely acceptable the Survey of Israel held a large number of training

sessions and workshops for the surveying and mapping community as well as with CAD software developing firms (that created tools to use the standard). Moreover, an online help desk was established to assist any professional who had troubles in adopting the national mapping standard. These efforts were successful and in the last two years most of the maps in Israel conform to the computer and logical format, cartographic layout and layer structure of the national mapping standard.

3. DEVELOPING THE CROWDSOURCING TOOLS

Following the approval and the adaptation of the national mapping standard, the SOI started to develop the tools and technologies to be able to use the crowdsourced maps and information for updating the national database. The system was developed on an existing platform called the TopoCad.

A key challenge in developing a crowdsourcing system is the design and development of user interface or the software user experience (UX). The user interface should be simple and intuitive but still meet all the functional requirements. It should be one that requires no training or a user manual. The project approach was to place the user in the centre, study the user and his work habits. Based on these insights, the project team came up with the design of the software user-interface. The design includes four steps (see Figure 3) and is based on the following rules:

- Simplicity – functions should perform in an intuitive manner that require minimum steps and minimum number of command buttons.
- Unified approach – all the function where designed in a similar style.
- Identification – is needed only for specific function using a card otherwise most of the functions are open for public use.
- Visualization - Graphic tools to view results on a map.

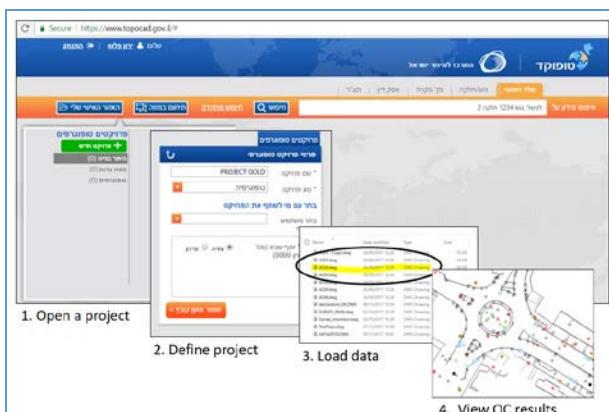


Figure 3: the four steps in the system user interface – open, define, load data, view QC results.

Internally, the system has two processes or modules: The Quality Control module and the Conflation module. These will be described in the following two sections.

3.1 The quality control module

Data from the TopoCad system is acquired from mapping experts and surveyors that are identified using a smart

identification card. Nonetheless even expert data have errors and mistakes and it is imperative to validate the quality of the data. Quality control for crowdsourcing data is a critical issue and was described in many publications (e.g. Barron et al. 2013). QC is specifically important when combining crowdsourcing files with an authoritative national database. The TopoCad system Quality control processes are based on the ISO19117 (2003) standard principles and quality components. The six quality components in the standard are: logical consistency, positional accuracy, completeness, thematic accuracy, temporal accuracy, and usability. The system performs semi-automatic QC process as presented in Figure 4 and performs the tests 5 components from the standard as follows:

- Computer format and logical consistency tests are automatic procedures in which the software verifies that indeed the submitted files are in DXF or DWG cad format and that the fidelity of the relationships in the data set meets the requirements. Some examples for the logical consistency tests include:
 - Topological checks, such as search for overshoots or undershoots within a thematic layer.
 - Domain rules tests in the database fields and attributes, such as a test that the elevations are not smaller than -500 m.
 - Logical tests between thematic layers, such as a test that unreasonable intersection of building layer with road layer occurs
- Completeness tests are done in an interactive manner by a human operator which samples the files and search for omissions (missing features) and commissions in 20% of the project area. The tests are done by superimposition the vector map on top of a high resolution (10 cm) orthophotograph or by field inspection. The selection criteria for missing and unrequired features is described in (Felus, Keinan, Regev, 2013).
- Positional accuracy assessment is performed by field surveys. During the survey selected features on the map are measure accurately in the field and the surveyed coordinates are compared with the map coordinates. The RMSE is calculated on the check points. The number of check points is given in Table 2. The calculated RMSE should meet the accuracy requirements as given in Table 1.
- Thematic accuracy assessment is performed by field surveys which scan 20% of the project area. The field surveyor verify that the identification of entities and assignment of attribute values are correct in the data set.
- Temporal quality is the test that verify the map is updated to the published date it is performed in the same manner as the completeness tests.

These tests are performed on a sample of the submitted maps. Maps submitted by new mapping experts and surveyors are fully tested (100%). Following five maps that have passed the QC tests, the QC process is performed on only 50% of the

submitted maps and then after five more successful submissions the process is performed on 30% of the maps (i.e.; 1 out of 3 maps is checked)

Map scale	Sample size
1:250	10 points + 2 point per 0.5 hectare
1:500	10 points + 1 point per 0.5 hectare
1:1000	10 points + 1 point per 2.5 hectare
1:2500	10 points + 1 point per 5 hectare

Table 2: number of check point for RMSE calculation.

3.2 The Conflation module

A map that passed the QC test should be consolidated and combined with the national topographical database (NTDB). The NTDB and its properties is described in Srebro et al (2010). The aim of this conflation process is to merge the data of the small map (e.g. a building permit map) with the NTDB. Through the conflation process, individual strengths of the source datasets can be combined, namely the output of a conflation process should be a map with superior positional accuracy and more attributes and data associated with the map (ref: Saalfeld, 1988). The TopoCad system has the following characteristics for its conflation process:

1. The NTDB is the main database and the small maps are merged into it with high accuracy and rich content.
2. The process is semi-automatic which means that the system performs the conflation but presents it to the human operator for approval. This way every modification to the authoritative NTDB is double-checked. The process is very fast and the operator can scan quickly through the modifications with the background of a high-resolution orthophotograph to assist him.
3. Every feature gets a unique metadata code which describes it source (data of the map, scale, accuracy, creator).
4. The process has three operations to the NTDB as depicted in Figure 4:
 - a. Add – A new feature is added to the NTDB if there is no existing feature in a buffer of a given threshold. The threshold is specified by the operator but the default is 2 meter which is the double estimate RMSE accuracy (2σ) of the NTDB.
 - b. Delete - A feature is removed from the NTDB if the new map contains no feature in a buffer of a given threshold.
 - c. Update – A feature which exist in the NTDB and in the new map (in a buffer of a given threshold) but has changed shape, position, height will be replaced by the new feature.
5. The conflation process on point features and building is performed using the Algorithm 2 in Keinan et al (2016). The conflation process is more challenging with lines, polylines and polygons. Therefore, the algorithm by Filing and Doytsher (1999) is used.

This algorithm employs counterpart nodes to match and transform corresponding segments and objects (see for example Figure 5-7).

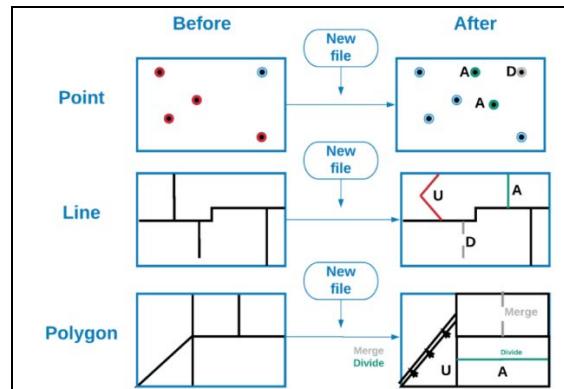


Figure 4: the three operations of the conflation process.



Figure 5: Full adjustment (1) and updating new data (2)

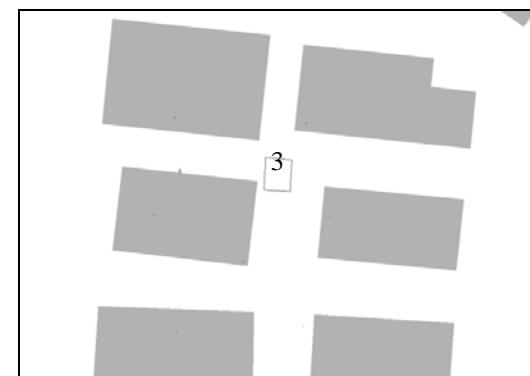


Figure 6: Extreme case. New data is found and requires updating. Quality parameters and policies should be set for deciding if the database should be Supdated.

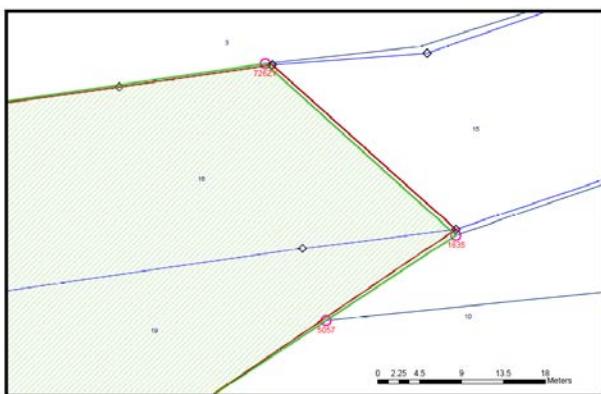


Figure 7: Conflating two parcel files using counterpart nodes.

4. CONCLUSION AND FURTHER WORK

This article describes a new approach for the creation of accurate national spatial database from large scale topographic maps. The new approach employs maps made by professional surveyors according to that national standard. The maps are submitted as part of a building construction process but the data is preserved and used to update the national topographic databases. The process is working successfully and more than a 600 maps have been integrated in every month of the initial phases of the system. It is expected that the number will grow to more than 5000 maps a month which represents all the changes in the country. Thus the process will provide a complete methodology to update every layer in the NTDB (buildings, transportations, infrastructure,).

The process is not fully automatic and requires human experts to do part of the quality control and to validate and assist the conflation process. It is hoped that these stage will be improved and become less dependent on human operators.

REFERENCES

Barron, C., Neis, P., Zipf, A. (2013): A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS at <https://doi.org/10.1111/tgis.12073>.

Bol, D.; Grus, M.; Laakso, Mari (2016); Crowdsourcing and vgi in national mapping agency's data collection. Proceedings of 6th International Conference on Cartography and GIS, June 13–17, 2016, Albena, Bulgaria at:
<https://cartography-gis.com/docsbca/iccgis2016/ICCGIS2016-50.pdf>

Capineri, C et al. (2016): European Handbook of Crowdsourced Geographic Information. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bax>

Felus, Y. A.; Sagi, Y.; Regev, R.; Keinan E; (2013); TopoCad - A unified system for geospatial data and services; International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-7/W2, 2013 at: <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W2/93/2013/isprsarchives-XL-7-W2-93-2013.pdf>

Felus, Y.; Keinan, E.; Regev, R. (2013) Regulations in the field of Geo-Information. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-7/W2, 2013, pp.87-92 at: <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-7-W2/87/2013>

Filin S., Doytsher Y., (1999): Linear Approach to Map Conflation: Matching of Polyline. Surveying and Land Information ISO/FDIS 19157 (2003), Geographic information - Data quality at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=32575

Keinan, E., Felus, Y. A., Tal, Y., Zilberstien, O., and Elihai, Y. (2016): updating national topographic data base using change detection methods, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B7, 529-536, <https://doi.org/10.5194/isprs-archives-XLI-B7-529-2016>.

McCartney A. E, Craun J.K, Korris, E., Brostuen, A. D., Moore R. L. (2015). Crowdsourcing The National Map. Cartography and Geographic Information Science. 42. S54-S57. 10.1080/15230406.2015.1059187.

McLaren R. (2012) National mapping agencies: Between a rock and a hard location, Geospatial World at: <https://www.geospatialworld.net/article/national-mapping-agencies-between-a-rock-and-a-hard-location/>

Shadbolt N.; Berners-Lee T.; Hall W. (2006) The Semantic Web Revisited IEEE Intelligent Systems archive, Volume 21 Issue 3, P. 96-101, DOI: 10.1109/MIS.2006.62

Srebro H., Felus Y, Tal, Y., 2010, New Initiatives in ISDI (Israeli NSDI), ISPRS Archive Volume XXXVIII-4-8-2/W9, Core Spatial Databases - Updating, Maintenance and Services – from Theory to Practice 15-17 March 2010, Haifa, Israel at: http://www.isprs.org/proceedings/XXXVIII/4_8_2-W9/papers/final_232_ISPRS_SFB_030810-srebro-Felus-.pdf Systems, Vol. 59(2), pp. 107-114.

Wikipedia (2018), National mapping agency at https://en.wikipedia.org/wiki/National_mapping_agency