Comments on three papers presented at FCSM meeting on November 15th, 2001

Shoreh Elhami, GIS Director
Delaware County Auditor

The three papers presented at "Issues in Geospatial Data Collection" session at FCSM conference were of various natures and therefore, the following comments will address each one separately.

1) Using Digital Geospatial Information to Locate Sample Units in the Field

With todays increasing use of mobile devices in a variety of applications, determining the effectiveness of these equipments in the field has become a critical issue. This paper has investigated the efficiency of two types of mobile devices for locating hard to find outlets in an urban area.

The paper discusses 3 stages of data collection (orientation, navigation, identification) using a handheld device loaded with a commercial mapping software, as well as a tablet PC also loaded with a commercial mapping software and connected to a GPS receiver.

The study has concluded that using commercial mapping software along with a GPS will increase the efficiency of planning a route, navigating to addresses and identifying outlets specially when the staff is unfamiliar with the territory. Their team was comprised of staff with varying skill levels and from different age groups which assisted in identifying important issues such as: the role of having prior familiarity with computers/new technology, mapping software programs, and also the role that the age of the data collector may play.

It was particularly determined that an easy to read interface, a larger display and larger fonts, as well as displaying minimal information would be crucial in increasing the efficiency of the process. Written directions and graphic display of the route played an important role and the interactivity of the mapping software provided for a flexible environment thru which the driver had a better chance of recovering from errors. In regards to the tablet pc and the GPS, it was determined that this setup increased the confidence level of the staff especially where they were unfamiliar with the area.

While the study has successfully analyzed a variety of issues regarding the use of the mobile devices for determining locations of hard to find addresses, the following suggestions may further compliment this study:

- Conduct identical tests (on the same outlets) in the same areas by people of similar demographic and expertise in a traditional fashion (using strictly paper maps) and compare the results and productivity rates with when the handheld

device was used alone and when the tablet pc and GPS receiver was utilized. It would especially be interesting to find out if the mapping software would be as useful at the planning and navigation stage if the staff is familiar with the area.

Experiment the process with the following variations:

- In areas with varying address types (rural addresses, multi-use units)
- In fast growing areas (containing many new streets) with commercial street databases and more accurate local datasets
- In areas containing inaccurate/misaligned road network (from the geometric standpoint); especially find out how useful the GPS is in these areas
- With devices equipped with audio instructions
- By comparing the affects of providing audio instructions vs. sending one or two individuals to the field
- By testing other map software programs, especially those that may combine the GPS and mapping application in the same package
- With adding imagery to the dataset (on the tablet pc) and analyzing its impact on identification stage
- By adding an application that provides for automated routing and even reporting live traffic information
- By using a dual display system (one to display directions and the other, the map and the best route), both installed in the vehicle
- By testing the process in areas containing detailed supporting data (point coverage, parcel data, landmarks) and study its impact on the orientation and navigation stages
- By using the GPS unit for identifying individual addresses in multi-unit residential or commercial uses
- Add data entry stage to the existing 3 stages and experiment with other software programs that accommodate that task. Test the process:
    - With GPS receivers of various accuracy levels
    - By applying differential Correction (either real-time or post correction), and further research the skills and training required for this task

2) Geospatial Data Collection and Analysis as Crucial Process in an Integrated Census

This paper proposes an ambitious plan for Israel's 2006 Census. It is evident that the use of GIS as an integral part of the entire process will provide for boundless opportunities, however, the success of this project will heavily depend upon several key elements.

The first and most important element is having a comprehensive and high quality dataset. A high quality dataset means: accurate, current, and well-documented datasets, either in spatial or non-spatial formats.

Some base spatial data layers very important for this process can be listed as:

- A road centerline (populated with current and accurate address ranges to allow for effective geo-coding)
- A point coverage of units including their xy coordinates and other unique identifiers (such as a unique parcel number, address, and etc. so it can be used as a relate/link item)
- Other supporting data layers at a more detailed scale such as cadastral/property ownership layers whenever they are available

There can be a variety of administrative datasets such as:

- Utilities
- Voter registration
- Welfare recipients
- Population registration
- And etc.

The Israeli's spatial database includes a comprehensive GIS dataset including a point coverage of all structures within the country. The point coverage in particular will play an integral role in the success of their project. Israel's nationwide cadastral coverage and their accurate street database are only complimentary to this rich dataset.

The second important element is establishing a level of confidence (which is mentioned in this paper as a Reliability Index) in every dataset used in the process. Close examination of every datasets' Metadata is an absolute necessity. When examining the Metadata, each datasets' currency, accuracy level, geographic extent, any embedded

standards, contact points, and most importantly any potential spatial items must be identified.

Thirdly, it is very critical to select administrative files appropriate for targeted areas or purposes. This could provide for more focused and efficient results mainly due to dealing with a smaller geography.

For example, in fast growing areas and also areas with a large number of multi-use units, the voter registration file may be the most suitable file to determine the locations of residents in an area. The addresses of these units are often missed if only cadastral datasets are used.

The fourth element deals with increasing the number of administrative files. Using a large number of administrative files and linking them can become very tricky and can create extremely complex scenarios. Additionally, lack of standardization (specially for addresses) could become a great hindrance in the process.

The success resulting from the last element may become the most challenging one since it requires collaboration and reaching agreements among various governmental and at times non-governmental entities.

Israel's innovative approach is impressive and detailed results of the country's 2006 Census including: data gathering and creation, creating linkages between various datasets, and effectiveness of GIS technology should be carefully examined.

3) Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey

The conclusion of this paper indicates that using residential mailing addresses and specifically, the correlation between the results of the Half Open Interval procedure and actual missing addresses for developing a metropolitan survey are beneficial in comparison to traditional methods.

The project is conducted in Dallas County, Texas and on a database containing 818,000 mailing addresses. It is worth to mention that the quality of the Delivery Sequence File and the road centerline utilized for this project, especially in a large and densely populated area such as this, is quite unique and exceptional. Achieving 99.4% hit rate is quite impressive and very difficult to achieve in many other areas of the country particularly in areas with rural addresses or fast growing regions.

Therefore, entities utilizing this method must first examine the currency, and completeness of DSF. DSF files in other parts of the country may not be as complete as Dallas County's database and that can adversely impact the effectiveness of the HOI methodology.

Another important issue to consider is to closely examining the quality of the road centerline used for geo-coding. Even enhanced TIGER-based road centerlines include missing roads and in some cases even though the roads exist, the address ranges are absent or incorrect which could result in low hit rates when geo-coding the files.

Utilizing HOI in conjunction with DSF is a natural match and GIS is a perfect tool for identifying and displaying locations of missing addresses in a sequence. However, the delivery sequence must be closely examined for unusual patterns.

Finally, it is worth to mention that HOI procedure could be effectively used for periodical Quality Control of any entity's DSF and for comparing the results with localities' Master Address Files for achieving a more complete coverage.

Literature Cited:

Nusser, S. M. and Fox, J. (2001). Using Digital Geospatial Information to Locate Sample Units in the Field, 2001 FCSM Paper.

Blum, O. and Calvo, R. (2001). Geospatial Data Collection and Analysis as Crucial Process in an Integrated Census, 2001 FCSM Paper.

Iannacchione, V. G. and Staab, J. and Redden, D. (2001). Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey, 2001 FCSM Paper.

Kish, L. (1965). Survey Sampling, John Wiley & Sons, New York.