REPORT OVERVIEW

In the Fall of 2010, the Bureau of the Census, Geography Division contracted with independent subject matter experts David Cowen, Ph.D., Michael Dobson, Ph.D., and Stephen Guptill, Ph.D. to research five topics relevant to planning for its proposed Geographic Support System (GSS) Initiative; an integrated program of improved address coverage, continual spatial feature updates, and enhanced quality assessment and measurement. One report frequently references others in an effort to avoid duplication. Taken together, the reports provide a more complete body of knowledge. The five reports are:

1. Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data
2. Measuring Data Quality
3. Reporting the State and Anticipated Future Directions of Addresses and Addressing
4. Identifying the Current State and Anticipated Future Direction of Potentially Useful Developing Technologies
5. Researching Address and Spatial Data Digital Exchange and Data Integration

The reports cite information provided by Geography Division staff at "The GSS Initiative Offsite, January 19-21, 2010." The GSS Initiative Offsite was attended by senior Geography Division staff (Division Chief, Assistant Division Chiefs, & Branch Chiefs) to prepare for the GSS Initiative through sharing information on current procedures, discussing Initiative goals, and identifying Initiative priority areas. Materials from the Offsite remain unpublished and are not available for dissemination.

The views expressed in these reports are the personal views of the authors and do not reflect the views of the Department of Commerce or the Bureau of the Census.

Department of Commerce
Bureau of the Census
Geography Division

Task 3 - Identifying the Current State and Anticipated Future Direction of Potentially Useful Developing Technologies

A Report to the Geography Division, U.S. Bureau of the Census

Deliverable #6 – Syneren Technologies Contract – Task T005

David Cowen, Ph.D., Lead Author,
Michael Dobson, Ph.D., and Stephen Guptill, Ph.D.

January 14, 2011

Topic Page

Executive Summary 1
Current Sources of Digital Imagery and Sensors 2
Image Processing and Geoprocessing Tools 3
Positioning Technology 5
Finding Candidate Areas for Targeted Address Canvassing 6
Citizen or User Generated Data 7
3.0 Purpose 9
3.0.1 Scope 9
3.0.2 Harris (2008) Rural Structure Change Detection Report 9
3.1 Investigation of Current Sources of Digital Imagery and Sensors (Aerial and Space Borne) 11
3.1.1 Assessment of the Applicability of Digital Imagery and Spectral Data Sources To Meet GSS Initiative Requirements 11
3.1.1.1 Spatial and Temporal Resolution 11
3.1.1.2 Spectral Resolution 16
3.1.2 Update on current and future Federal programs that provide funds for acquisition 18
3.1.2.1The National Agriculture Imagery Program (NAIP) 18
3.1.2.2 Imagery for the Nation (IFTN) 19
3.1.2.3 The National Digital Orthophoto Programs (NDOP) 20
3.1.2.4 The Multi-Resolution Land Characteristics (MRLC) Consortium 20
3.1.3 Digital Globe / Microsoft Advanced Ortho Aerial Program 21
3.1.4 Oblique Imagery 23
3.1.4.1 Pictometry 23
3.1.4.2 Fugro EarthData 26
3.1.5 Light Detection And Ranging (LiDAR) 26
3.1.5.1 Overview of LiDAR Technology 26
3.1.5.2 Federal Organizational Issues Relating to LiDAR 31
3.1.5.2.1 NGA's LiDAR Data for Greater Washington DC 32
3.1.5.3 Terrestrial and mobile LiDAR systems 35
3.1.5.4 Research Findings regarding LiDAR 37
3.1.6 Analysis of Existing and Planned Earth Observation Satellites 41
3.1.6.1 Overview of Existing Satellite Based Sensors 41
3.1.6.2 Planned Programs That Will Impact GSS Initiative 43
3.1.6.2.1 SPOT 6 and SPOT 7 43
3.1.6.2.2 Schedules For Other Earth Observation Satellites 43
3.1.7 Summary 44
3.2 Investigate the Current State of the Art in Digital Imagery Processing 44
3.2.0 Harris Report 44
3.2.1 Image Classification to Meet GSS Initiative Requirements 45
3.2.2 Feature Extraction 48
3.2.2.1 Feature Extraction Research 49
3.2.2.1 Object Based Image Analysis 54
3.2.2.2 Image Processing Trends 55

3.2.3 A Review of Change Detection Procedures....57
3.2.3.1 Review of European National Mapping Organizations Criteria and Findings Regarding Change Detection 58
3.2.3.1.1 Change Detection Research at The Ordnance Survey 62
3.2.3.1.2 European Experiments with Building Level Change Detection....62
3.2.3.2 Binary Change Detection Research 64
3.2.3.3 Object Based Change Detection 67
3.2.4 A Review of Commercial Software Solutions....69
3.2.4.1 Image Classification within ArcGIS....69
3.2.4.2 Commercial Image Processing Systems 70
3.2.4.2.1 Overwatch Feature Analyst, LiDAR Analyst and Building Collection Toolkit 71
3.2.4.2.2 Geoeye Road Tracker 73
3.2.4.2.3 ERDAS 74
3.2.4.2.4 ENVI 76
3.2.4.2.5 eCognition....76
3.2.4.2.6 LP360 (QCoherent Software) 78
3.2.4.2.7 Observa Change Detection Server (CDS) for Aerial and Satellite Imagery 78
3.2.4.2.8 IDRISI 79
3.2.4.2.9 Hitachi/ Pictometry ChangeFindr™ 80
3.2.4.3 Summary....83
3.2.5 Geoprocessing Procedures....84
3.2.5.1 General Procedure for Discovering New Roads....84
3.2.5.2 Procedure To Generate Street Centerlines From Parcel Data....85
3.2.5.3 Other Recent Geoprocessing Research....85
3.2.6 Summary....87
3.3 Investigate the Current and Future Technology to Support Precise Location Measurements (Advanced GPS Positioning) 88
3.3.0 Introduction....88
3.3.1 Overview of Positioning Technology 92
3.3.2 Sources of Error in GPS....96
3.3.3 Augmentation....96
3.3.3.1 Real Time versus Post Processing Augmentation 97
3.3.3.1.1 SBAS Satellite (or Space) -Based Augmentation System 98
3.3.3.1.1.2 Global Satellite Based Augmentation System 98
3.3.3.1.1.3 Wide Area Augmentation System (WAAS) 98
3.3.3.1.2 Ground Based Augmentation System (GBAS)....99
3.3.3.1.2.1 Differential GPS....99
3.3.3.1.2.1.1The Network RTK 99
3.3.3.1.2.1.2 Global Differential GPS (GDGPS)....100
3.3.3.1.2.1.3 The Local Area Augmentation System....100
3.3.3.1.2.1.4 Nationwide Differential GPS System 100
3.3.3.1.2.1.5 International GNSS Service (IGS)....100
3.3.4 Federal Programs for Ground Based Augmentation....101

3.3.4.1 National Spatial Reference System (NSRS) 101
3.3.4.2 Continuously Operating Reference Stations (CORS) 101
3.3.4.3 The Virtual Reference Network (VRN) 102
3.3.5 Global Navigation Satellite System (GNSS) 102
3.3.5.1 An Analysis of the Prospects and Timeline for Additional Constellations 103
3.3.6 Assisted (Versus Augmented) GNSS Based On Communication 103
3.3.7 Prospects for Future Developments from Nokia and Other Cell Phone Manufacturers about Autonomous Cell Phone Positioning 104
3.3.8 Summary 106
3.4 Identification of Approaches to Finding Candidate Areas for Targeted Address Canvassing 106
3.4.1 Proposed Model 106
3.4.1.1 Discussion of Components of the Decision Support System 110
3.4.1.1.1 Level of Geographical Analysis 110
3.4.1.2 Data Elements 113
3.4.2 Public Sources of Roads 113
3.4.2.1 Crowd Sourced Streets 114
3.4.3 Address Lists 116
3.4.3.1 Use of Mass Mailers 118
3.4.4 Temporal Analysis of DSF Data 118
3.4.5 Parcel Data 121
3.4.6 Decision Support System - Analytical functions 122
3.4.7 Decision Making 123
3.4.8 Other Related Topics 124
3.4.8.1 National Urban Change Indicator (NUCI) 124
3.4.8.2 NOAA Digital Coast Coastal Change Analysis Program C-CAP 124
3.4.8.3 Population Growth Projection Research 125
3.4.8.4 Day And Night Time Population Estimates 125
3.4.9 Summary 126
3.5. Investigate the Current and Potential Use Of Crowdsourcing/User Generated Data for Collecting Spatial Data 127
3.5.1 An Examination of Crowdsourcing in the Form of Active and Passive Community Input (User Generated Content) For Tracking Map Changes Including Roads and Addresses 131
3.5.1.1 Active Community Input 131
3.5.1.2 Passive Community Input 133
3.5.1.3 Summary 135
3.5.2 A Description and Evaluation of the Data Exchange Mechanisms Used To Receive, Ingest, and Authenticate Data from Volunteers 137
3.5.2.1 TomTom/Tele Atlas Data Ingest and Evaluation 138
3.5.2.2 NAVTEQ Data Ingest and Evaluation 141
3.5.2.3 OpenStreetMap Data Ingest and Evaluation 143
3.5.2.3.1 Edit Maps and Data 143
3.5.2.3.2 Error Correction 144
3.5.2.3.3 More on OpenStreetMap 145

3.5.2.3.4 Summary 148
3.5.2.4 Google 148
3.5.2.5 Summary 152
3.5.3 A Report on the Status of Crowdsourcing Efforts by Federal Agencies 156
3.5.3.1 United States Geological Survey 156
3.5.3.1.1 USGS VGI 158
3.5.3.2 The FCC, NTIA and Crowdsourcing 161
3.5.3.3 FEMA 161
3.5.3.4 Summary 162
References 163

Executive Summary

This report provides an assessment of a wide range of potentially useful developing technologies that could be deployed by the Geography Division to meet the goals of the GSS Initiative. The information assembled in this report provides an assessment of the state of the art as well as a glimpse into the future. The general sense is that the current set of image and sensor data and tools have advanced to a mature stage. Access to many of these resources is now available through common web services. By surveying the experiences of the other national mapping organizations as well as local and state partners we have assembled assessments of best practices as well as examples of some cutting edge technology. The report can serve as a valuable reference document as well as a starting point for internal discussions about how to proceed with the GSS Initiative.

The report is specifically targeted at systems, data, and software solutions that would provide efficient and cost effective procedures to guide a targeted address canvass. This objective requires the harvesting of the most accurate and current geospatial data that exists, then implementing procedures that can classify, extract, and format those data into useful information. We believe that the GSS Initiative requires a spatially explicit decision support system that provides surveillance and resource allocation functions. Such a system could continuously monitor the entire nation in terms of the quality and completeness of the MTdb. In areas where the data were deficient, the system could automatically alert the management about conditions on the ground, provide access to finer resolution data, and provide critical information required to justify and allocate resources such as a targeted address canvas. The analogy has similarities to tactical operations on a battlefield. In fact the objective is exactly the same – minimize the need to send personnel into the field.

The nature of the task required us to cast a large net to discover the current state of the art with respect to geospatial data around the world. In fact, many of the topics, such as the evolution of the Global Navigation Satellite Systems (GNSS) are inherently global. Developments in imagery, earth observation satellites, Light Detection and Ranging (LiDAR), and the tools to process such data are evolving in research centers and commercial firms across the globe. Our analysis spans a wide range of environments. These include work by local governments as part of their routine business processes as well as current research projects underway in national mapping organizations in Europe. The analysis provides an update on several federal programs and reports on anticipated events that are likely to impact the GSS Initiative over the next decade. The emphasis of the analysis is on technical feasibility while recognizing the practical aspects regarding implementation on a nationwide basis. The analysis covers positioning technology, data sources (including user generated content), tools to identify, capture, and update features, and spatial decision support systems. By understanding the current status of each of these components the Geography Division should be able to move forward with the GSS Initiative with a clear view of technical and logistical alternatives.

The report is organized around five major topics:

1. The current sources of digital imagery and sensors (aerial and space borne).
2. The current state of the art in digital imagery processing procedures for image classification, automated feature extraction, and change detection to identify, represent, and geographically position new roads and structures.
3. The current and future technology to support precise location measurements (advanced GPS positioning).
4. Identification of approaches to finding candidate areas for targeted address canvassing.
5. The current and potential use of citizen or user generated data.

Current Sources of Digital Imagery and Sensors

The section on digital imagery and sensors concerns the data sources that can support the GSS Initiative objectives and emphasizes the technology that can detect features such as roads and structures. One finding is that as a result of "content wars" among the large web mapping or location based service providers (Google and Microsoft) high resolution aerial images are now easily accessed using web browsers. These two providers have also entered into long term relationships with GeoEye and DigitalGlobe to continuously update their image resources. Google has partnered with GeoEye to access satellite imagery with sub meter resolution. Microsoft has teamed with DigitalGlobe in the Advanced Ortho Aerial Program that has begun to acquire one foot aerial imagery for the country.

Our analysis suggests that the requirements of the GSS Initiative demand very recent imagery with a spatial resolution of less than a meter and reflect a recent view of the situation on the ground. Local governments that need to assess property are now routinely acquiring orthographic imagery with pixels of 4 or 6 inches on an annual basis. Many local governments that are conducting virtual property inspections have chosen to acquire oblique imagery from Pictometry. While the Pictometry imagery, at a restricted scale, is currently incorporated into the Microsoft Bing map "Birdseye view", the full resolution imagery with a parcel overlay is available as a web service. This oblique imagery can be visualized within a desktop GIS environment as single images or oblique mosaics.

From a federal perspective the Department of Agriculture's National Agriculture Imagery Program (NAIP) is the only program to acquire imagery of the nation on a regular basis (currently a 3 year cycle). This program, collecting ortho imagery with a resolution of one meter, is oriented to agricultural inventories; therefore, it acquires imagery during the growing season (leaf on). The Imagery for the Nation program supported by the Federal Geographic Data Committee appears to be stalled due to lack of funding support.

Both GeoEye and DigitalGlobe operate satellites (GeoEye-1 and Worldview-2) that collect sub meter resolution imagery that could support identification of structures and road features. As part of National Geospatial-Intelligence Agency's Enhanced View program both companies are planning to launch new satellites in the next few years. (WorldView-3 is scheduled for 2014 and GeoEye-2 should be operational by 2013.)

The most significant advancement in image-related technology has been the explosion in the acquisition of Light Detection and Ranging (LiDAR) data. LiDAR technology utilizes active laser sensor pulses at rates up to 500,000 per second to capture X, Y and Z coordinates. While the majority of the data are acquired from aircraft, there are also advances in LiDAR collection from terrestrial and even mobile devices. While designed to generate high resolution digital surface models for applications involving flood plain mapping, many users at local, state and even national levels have determined that LiDAR represents the best source for automatic detection of surface features, such as buildings and roads, and may be used to detect change. Given the range information acquired from the different returns and intensity measures, it is possible to generate detailed representations of the natural and built environment. The fact that the data are produced in GIS formats makes it easy to incorporate the imagery into an analytical environment. Over the past decade the processing of LiDAR has evolved from an experimental to a commercial off the shelf environment. In addition, LiDAR has become the basis for a high resolution national elevation data base. With funding from FEMA, USGS, and local partners, many states are collecting comprehensive coverage with LiDAR. Thus LiDAR data are increasingly available not just in flood plains, but for all areas. The use of mobile LiDAR has been widely implemented by NAVTEQ and will be part of its JourneyView program scheduled for 2011.

Key Points:

- The GSS Initiative requires one meter or better spatial resolution.
- Many web image services are currently available.
- Both GeoEye and Digital Globe operate commercial satellites with sub meter spatial resolution.
- Significant amounts of LiDAR data are being collected for a variety of applications.
- The Imagery for the Nation initiative appears to be stalled.
- The Digital Globe/Microsoft Advanced Ortho Aerial Program, designed to create 30 centimeter imagery on a three year cycle, will complete the first collection of new imagery by mid 2012.
- The National Agriculture Imagery Program (NAIP) is the only federal program to acquire imagery of the nation on a regular basis.
- The Census Bureau participates in federal imagery activities through its membership in the FGDC National Digital Orthophoto Program (NDOP) subcommittee.

Image Processing and Geoprocessing Tools

Another objective of the research was to survey the current state of the art in feature classification, extraction, and change detection. It is important to note that the same tools can be deployed to process different scales of spatial data. Therefore, the starting point to any image processing process must include data of appropriate spatial and temporal resolution. Our analysis focused on methods being employed to capture individual streets and structures.

Classification is the fundamental process of labeling pixels in an image. The software to perform supervised and unsupervised classification is very mature. There are also detailed sets of

protocols and best practices to guide the process. It is interesting to note that most of these image processing tools are now available as “raster processing” tools in ArcGIS. We also noted that there are three promising new techniques for image classification: (1) Classification and Regression Tree (CART) analysis, (2) Artificial Neural Networks (ANN), and (3) Support Vector Machines (SVM).

The forefront of classification research has focused on Object Based Classification procedures. These procedures utilize a wide range of data sources (LiDAR, parcels, etc.) to segment the study area into polygons and incorporate rules that take into consideration the shape and texture of data to differentiate features that may have similar spectral characteristics. The eCognition software, now owned by Trimble, is widely used by national mapping organizations in Europe to perform large scale land cover classifications using this method.

The report includes a brief description of the most applicable commercial image processing systems that perform image classification, feature extraction, and change detection. In addition to eCognition, this includes ERDAS Imagine and Delta Cue, ENVI, Openwatch Feature analyst and LiDAR analyst, QCoherent LP360, Observa Change Detection Server, and IDRISI. In addition to handling traditional remotely sensed data, software solutions to process the huge amounts of data in LiDAR point clouds have also matured and also include functions to automatically extract features, such a building edges, and convert them into polygons.

GeoEye’s RoadTracker® appears to be a particularly interesting new software system for semiautomatic road extraction from imagery. It has functions to control the width and smoothness of the lines, support topological snapping, perform automated topology cleaning, and make adjustments based on 1 and N node detours. This tool is being used by NGA to create road networks in places without good base maps. We believe that this process might be used effectively in the GSS Initiative to capture roads from imagery.

It could be argued that a key focus of the GSS Initiative is change detection. Therefore, the report includes a detailed assessment of change detection procedures. An important component of this analysis stems from work on this problem by the national mapping organizations in Europe. Their findings strongly suggest that the object oriented procedures that incorporate LiDAR are the best for detecting changes at the scale important to the GSS Initiative. This includes the ability to identify individual structures over large geographical areas. The eCognition software is widely used to conduct this analysis. A review of academic research also confirms the importance of object oriented change detection. It is also suggested that binary (change / no change) measures are valuable as surveillance tools to generate a spatially explicit alert.

While many European national mapping organizations are conducting large scale land cover classifications and change detection efforts, many local government in the United States are focusing on monitoring changes in individual buildings. We discovered fairly wide spread use of ChangeFindr™ to perform this task. ChangeFindr™ is a combination the Hitachi HouseDiff™ software and the Pictometry change analysis module. Two case studies reveal that this system can efficiently extract the outline of buildings and analyze them between two different time periods. These users provide data about how this process has been used to conduct virtual field

inspections to find new buildings and improvements, leading to a new assessed value that generates a significant return on investment. The same process has been used by local governments to find and document additional housing units for the LUCA process.

We believe that every effort should be made during the GSS Initiative to harvest road files from local and state partners. Many states (South Carolina, New Mexico, Indiana, Arkansas, etc.) have established state offices to support emergency response efforts. As a result, they have a central repository of "E911 street centerlines" (as well as address points, parcels, etc.). As part of this research we developed a simple geoprocessing system that utilized E911 roads to discover and separate new roads from the existing TIGER roads. This suggests that the Geography Division could extract new roads from an entire road file provided by a county without requiring the county to preprocess the file. We also demonstrated that street centerlines can be automatically created from parcel polygons. This procedure would provide a useful basis for generating roads from an accurate base map, evaluating the classification of roads that fall within a parcel, and to even insert roads from planned subdivisions.

Key Points:

- Pictometry's ChangeFindr™ and GeoEye's RoadTracker® are potentially useful tools for creation of structures and roads.
- There are several robust commercial GIS software tools to process LiDAR data.
- Several national mapping organizations in Europe are conducting valuable research on feature extraction and change detection of roads and buildings.
- Three promising new techniques for image classification are: (1) Classification and Regression Tree (CART) analysis, (2) Artificial Neural Networks (ANN), and (3) Support Vector Machines (SVM).
- The NGA LiDAR collection of 3D buildings for the Washington DC metro area is a good indicator of the value of LiDAR for the GSS Initiative.

Positioning Technology

The third section of the report focuses on current and future trends in positioning technology. From a GIS perspective, the MTdb is an assemblage of point, line, and polygon features. Over the last decade the positional accuracy of these data was substantially improved to support the use of handheld devices to capture accurate features in the field. Since the time that the original hand held devices were designed and deployed, we have seen an emergence of a host of locationally aware devices. These GPS enabled cellular phones and smartphones have dramatically altered the ways citizens and professionals utilize geographic information. The location based services market has exploded and personal navigation systems are now inexpensive commodities. Our report *Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data* provides an analysis of current mobile technology and makes several recommendations about how to design and deploy the next generation of field oriented devices. It is important to note that the FAA's Wide Area Augmentation System (WAAS) system for augmenting GPS data remains the most cost effective technology to support the accuracy requirements of the GSS Initiative. The findings of our research focus on changes that will emerge in this decade. The findings indicate that we are in the midst of a convergence

of technology from the traditional surveying community and the mobile communication industry. The current US NAVSTAR (NAVigation Satellite Timing And Ranging) system will be expanded to the Global Navigation Satellite Systems (GNSS) with the addition of the following:

- Russian Federation: GLONASS – during 2010;
- European Union: Galileo – 5 to 10 years;
- China: Compass – 5 to 10 years;
- Plus at least 2 Regional Systems:
    - India: Regional Navigation Satellite System (IRNSS);
    - Japan: Quasi-Zenith Satellite System (QZSS).

As a result, users around the world will have access to many more satellites. These will support real-time sub meter accuracy without augmentation. For example a NOAA scientist predicts that by 2027 we will have autonomous (standalone) positional accuracies of less than 1 meter. With access to more than 20 satellites, users would always have access to at least 6 thus eliminating the problem of multipathing in urban canyons.

Concurrent with the expansion of the GNSS have been advancements in the determination of locations based on assistance from communication devices. By using the location of cell towers, Wi-Fi hot spots, and even the location of other cell phones, smart phones are able to fix locations even inside buildings. These devices are also “continuously aware” meaning they do not have to wait for a “time to first fix”. According to some experts we might soon see centimeter level accuracy from cell phones. The major finding is that the GSS Initiative will have access to highly accurate and inexpensive devices to support field based operations.

Key Points:

- The Global Navigation Satellite System (GNSS) will greatly expand the number of satellites available to calculate positions.
- The location based services market has exploded and personal navigation systems are now inexpensive commodities.
- There will be a dramatic expansion in alternatives for augmenting or assisting GPS in quickly fixing and determining locations.
- The FAA’s Wide Area Augmentation System (WAAS) system remains the most cost effective technology for the GSS Initiative.

<u>Finding Candidate Areas for Targeted Address Canvassing</u>

The fourth section focuses on ways to combine data and tools to create a spatial decision support system to guide the targeted address canvass part of the GSS Initiative. Here we discuss the benefits of assembling data from the MTdb, the 2000 and 2010 census, the American Community Survey, other federal agencies, public partners, and even commercial data providers to establish a nationwide monitoring system of indicators of change. These indicators could include: information about vacancy rates from Department of Housing and Urban Development (HUD), Census Bureau/HUD new residential construction statistics, Home Mortgage Disclosure Act (HMDA) mortgage activity from the Federal Financial Institution Examination Council, and

the DHS National Urban Change Indicator (NUCI) program. It could include population and housing estimates and forecasts from CACI and Geolytics. It could also include estimates of the length of new roads from partners and crowd sourcing efforts. These data, available in a variety of formats (GIS, raster, tabular), would be assembled or fused at the tract level. As an interactive web based mapping system, it would serve as a decision support system. Bureau statisticians could develop models that would incorporate significant indicators that would determine thresholds for binary classification of tracts that have a high probability of missing or inaccurate elements of the MTdb. The system would also support exploratory data analysis and data mining using visualization tools such as the GeoVIZ system from Penn State University. Most importantly, it would provide an interface to access additional web services that could be used to "drill down" in a virtual navigation system to reveal the most accurate and current geospatial data for any part of the nation. In a final step, the system would be used to assemble new data to incorporate into the MTdb or to provide the relevant sources to a mobile device for field based inspection and update.

Key Points:

- Targeted address canvassing could be guided by a spatial decision support system that provides surveillance, analytical visualization, and resource allocation functions.
- An extensive array of indicators could be assembled at the tract level from both internal and external data sources.
- The GSS Initiative should incorporate procedures to harvest road files and address lists from local and state partners.
- Longitudinal analysis of the USPS Delivery Sequence File and Zip + 4 data could serve as a basis for monitoring change.

Citizen or User Generated Data

Crowdsourcing, also known as Volunteered Geographic Information, represents a potentially important compilation source for collecting address and spatial data often not captured by more traditional methods. Utilizing an architecture of participation, crowdsourcing can harness the power of mass collaboration in an attempt to capture accurate geospatial data. Used to collect both active and passive community input, crowdsourcing is the power behind the OpenStreetMap project, as well as the primary method of collecting data used by Tele Atlas, the commercial navigation database provider owned by TomTom.

The data quality aspects of crowdsourcing are based on the notion that these geospatial databases are editable by anyone and that, over time, as more eyes examine the data, more errors will be caught and corrected. Various research suggests that crowdsourced data are not yet quite as accurate or comprehensive as the data provided by authoritative government or commercial sources of map databases. In addition, it appears that current efforts aimed at developing crowdsourced databases are focused on collecting the geometry of streets and roads at the expense of attribute and address collection. However, it is still early in the development of crowdsourced systems and we believe that the methods employed in capturing crowdsourced address and spatial data could be of significant benefit to the Geography Division in the continued development of the MTdb.

As a consequence, we examined the techniques of active and passive (probe-based) crowdsourcing as used by OpenStreetMap, Tele Atlas/TomTom, NAVTEQ and Google. Each of these organizations utilizes different approaches to crowdsourcing and while we found the OpenStreetMap approach quite powerful, we believe that some aspects of the probe data collection utilized by the Map Share component of the TomTom personal navigation devices to be an elegant and compelling method of incenting and measuring the contributions of users. Further, we note that the major users of crowdsourced data differ in how they evaluate, edit, and correct these data.

Volunteered Geographic Information appears to be of significant interest to various agencies of the Federal Government, although it is not currently utilized in a production mode by any agency. For example, our examination of the efforts of USGS revealed that this limited test was exploratory in nature and focused on the potential use of the OpenStreetMap technology stack for joint editing between agencies, rather than on the use of crowdsourcing as a data compilation tool.

We believe that the use of crowdsourcing within the framework of a more formal compilation and edit process (a hybrid approach to crowdsourcing) may provide a powerful tool for collecting and augmenting spatial databases. We urge the Geography Division to consider adapting crowdsourcing as it moves ahead with the concept of the continuous updating of the MTdb.

Key Points:

- As systems mature crowdsourced address and spatial data could be of significant benefit to the GSS Initiative.
- The probe data collection strategy utilized by the Map Share component of the TomTom personal navigation devices is particularly interesting.
- OpenStreetMap incorporates a number of powerful tools to collect and edit crowdsourced data that could be of use if the Geography Division pursues a hybrid approach to crowdsourcing.

3.0 Purpose

This report assesses a wide range of potentially useful developing technologies that could enable the Geography Division to meet the goals of the GSS Initiative. It is specifically targeted at systems, data, and software solutions that would provide efficient and cost effective procedures that would guide a targeted address canvass. The developing technologies investigated included: change detection using imagery as a method of capturing housing structures, the use of improving GPS technology and post-processing techniques to further refine positional accuracy, the impact of collecting elevation (third dimension) data, innovative approaches to finding candidate areas for targeted address canvassing, and potential role of user generated geographic data content. The goal is to provide the Bureau with an assessment of the most promising technologies, a list of providers or partners, an estimate of the future directions and developments in the technology, and recommendations for a course of action to test and implement new processes for data capture.

3.0.1 Scope

The scope of this research is potentially very broad. The methods we employed cast a large net and then focused on those data sources and solutions that are most promising in terms of Census' needs. This study includes a review of current state of the art, financial and technical feasibility, and best practices in local government and national mapping organizations. At the same time we examined recent research projects that may evolve into practical and affordable solutions in the near future.

Our working assumption is that Census wants to examine innovative technological solutions that will support of targeted address canvassing. In essence, the Bureau would like to identify those geographic areas that are likely candidates for finding new (missing) residential dwellings and then locate and extract specific structures, record addresses, and collect any new infrastructure such as roads.

3.0.2 Harris (2008) Rural Structure Change Detection Report

It is important to note that this report builds on the Harris Corporation 2008 *Rural Structure Change Detection Report for the Census Bureau*. A summary of the scope of that report is portrayed in Figure 1. Our report provides an update on the image sources and processes that were reviewed by Harris, expands on potential methods to support targeted address canvassing, and adds an entirely new section on crowd sourcing as a source of spatial data.

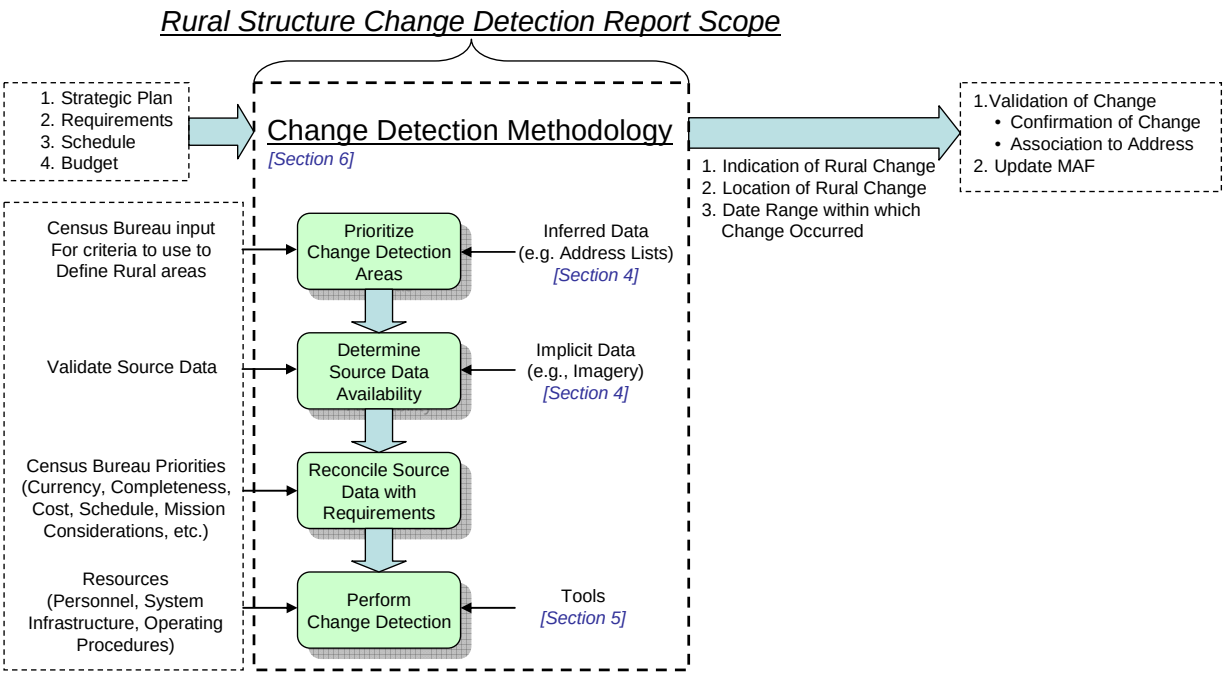


Figure 1 Harris change detection flow chart (Harris, 2008)

It is worth noting the Harris provided findings that are still relevant two years later:

- A MAF that contains a temporal baseline and geo-code for each address is more valuable when considering how to perform the ongoing maintenance of the MAF.
- Using an image-to-image Change Detection approach results in an indication of change that is temporally inconsistent with the MAF and, therefore, would require a means to de-conflict inconsistencies in order to use it to maintain the MAF.
- Using a temporally baselined MAF-to-Image approach forces the definition of guidelines to associate MAF addresses with structures in an image, eliminates temporal inconsistencies and requires only a current image.
- Visual Review Productivity – How much time does it take an analyst to visually scan an entire extent to identify structure changes using an image-to-image and address list-to-image approach?
- Change Detection Tools – How much training is required to maximize the use of an image analysis tool to perform Change Detection?
- Macro Level Change Detection – How effective is the use of low-resolution imagery to perform macro-level Change Detection to identify areas of change?
- Completeness – How much effort will it require to review an image analysis tool's Change Detection results to ensure structure changes are valid and complete (none were missed); can an image analysis tool be sufficiently reliable to reduce the "eyes-on" review time?
- Image Management – What type of imagery should be utilized; how much effort is required to manage these images and understand the relationships between overlapping images?

- Address Association – Can acceptable guidelines be established by which to associate an address with a structure in an image?
- System Infrastructure – What constitutes the support resources that would be needed to scale Change Detection activities?
- Update – How will Change Detection results be used to achieve the update of the MAF?

3.1 Investigation of Current Sources of Digital Imagery and Sensors (Aerial And Space Borne)

The objective of this research task is to provide the Geography Division with an evaluation of aerial and space-borne (orbiting) sensor systems that would assist the division with developing inputs to a targeted address canvass activity under the GSS Initiative. It specifically addresses the need to monitor regional land cover changes and to identify new structures. It includes an assessment of the applicability of digital imagery and spectral data sources to meet these requirements. It includes an analysis of existing and planned earth observation satellites in terms of spatial, temporal, and spectral resolution. It provides an update on current and future Federal programs that will provide federal funds for imagery acquisition including LiDAR. Another level of investigation reports on the status of national imagery programs such as The National Aerial Photography Program (NAPP), The National Agriculture Imagery Program (NAIP) and Imagery for the Nation (IFTN) that would provide the Bureau with products that could meet their needs. We also provide a discussion of current activities by Google and Microsoft, two firms that are making significant investments in high resolution imagery.

3.1.1 Assessment of the Applicability of Digital Imagery and Spectral Data Sources to meet GSS Initiative Requirements.

3.1.1.1 Spatial and Temporal Resolution

Given the focus of the GSS Initiative on roads and structures the evaluation of remotely sensed data can be significantly confined based on spectral, spatial, and temporal dimensions. Spatial resolution is defined by the pixel size of the imagery. A pixel or picture element is the smallest two-dimensional area sensed by the remote sensing device. A comparison of spatial resolution is provided in Figure 2. Clearly individual buildings do not come into focus with resolutions larger than 5 meters. Jensen (2007) suggests that building oriented analysis requires a minimum spatial resolution of 0.25 – 5 m. In order to identify such features requires a minimum of four pixels within an object to identify. With a guideline of one-half the width of the smallest dimension the pixel resolution for a 5 meter wide mobile home would be no larger than 2.5 meters.

It is important to note that the spatial requirement for the GSS Initiative can now be met with imagery produced by commercial satellites. For example, the two images from GeoEye 1 with a resolution of 0.41 meters and 1.65 meters (Figures 3 and 4) would typically provide the necessary basis for point level placement of address points on a front yard or rooftop. However, for applications that require inspection of characteristics of individual structures, finer resolutions are required. A common specification for imagery resolution is one foot, however; many local governments are now collecting four inch data (Figure 5).

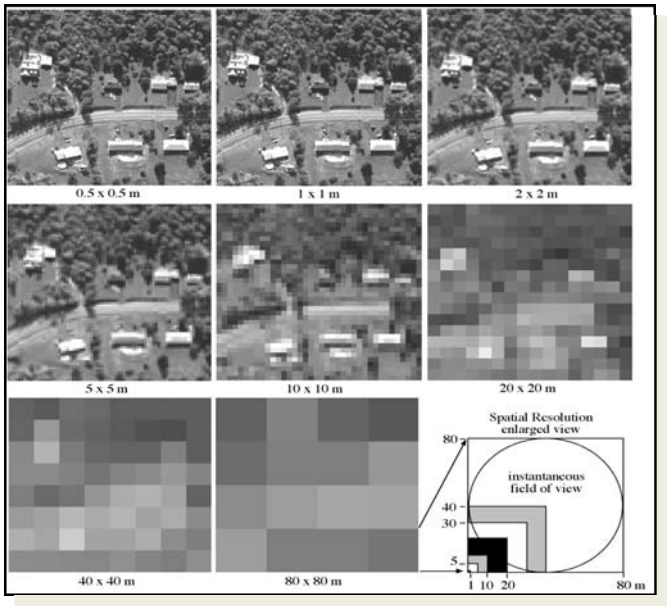


Figure 2 Comparison of different spatial resolutions (Jensen, 2007)

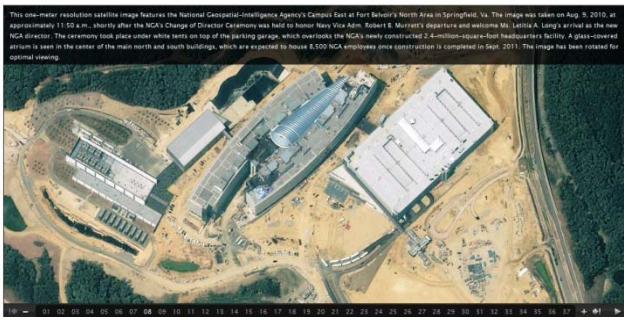


Figure 3 GeoEye One meter resolution data (www.geoeye.com)

Figure 4 GeoEye .5 meter resolution data (www.geoeye.com)

Figure 5 Four inch pixel Pictometry image (generated for this report)

The temporal dimension of any image is also critical to the GSS Initiative. Temporal resolution is defined by how frequently a sensor can collect an image of a particular area of interest. Sensors in airplanes and helicopters can acquire imagery on as needed basis while sensors on satellites are limited by their orbit and ability to capture images in an “off nadir” angle. For example, Landsat is a nadir sensor that can only visit the same location every 14 days. The relationship between spatial and temporal dimensions is portrayed in the chart prepared by Jensen (2007) (Figure 6). This chart demonstrates that the ideal imagery for building level analysis is one foot or better on an annual basis. In fact, many local governments are now scheduling yearly flights.

The value of any imagery is its ability to identify features. In the context of the GSS Initiative these features typically involve urban land cover and, ideally, specific types of residential land use. The examples (Figures 7and 8) from the National States Geographic Information Council (NSGIC) provide a useful way to consider the difference between land cover and land use.

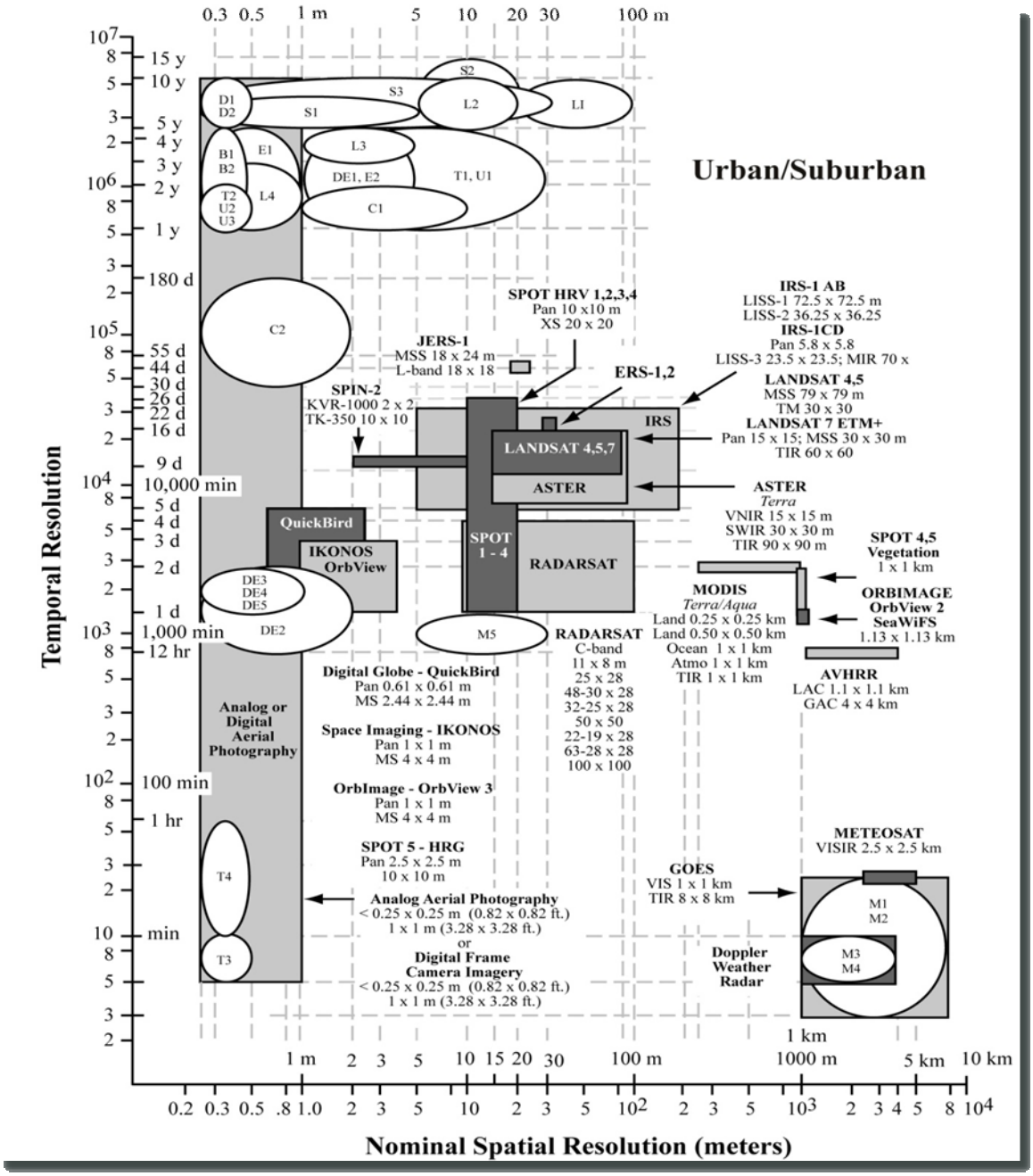


Figure 6 Sensor platforms with different spatial and temporal resolutions (Jensen, 2007)

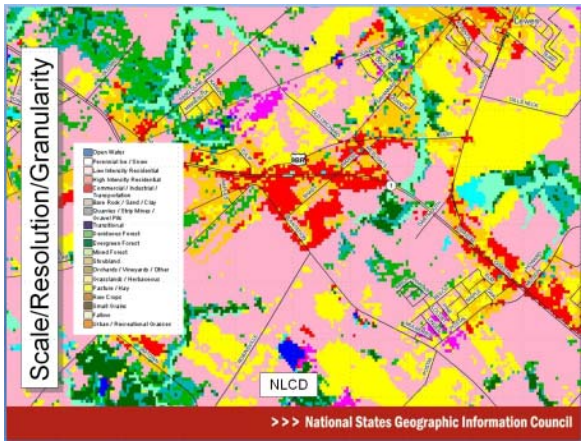


Figure 7 Land cover data from the NLCD (NSGIC)

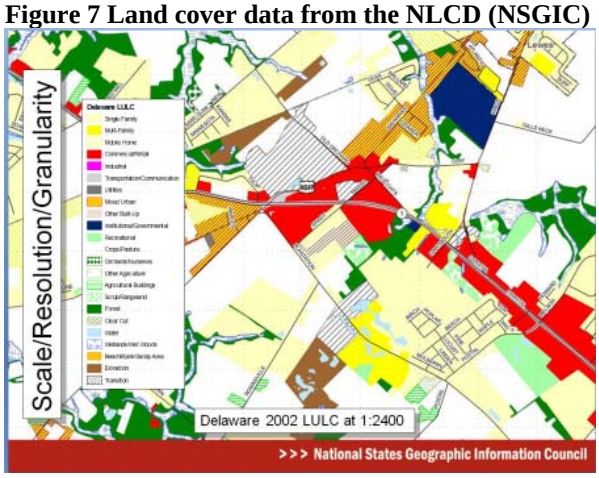


Figure 8 Parcel based land use data (NSGIC)

Traditionally, the distinction between land cover and land use is reported in terms of the original land cover classification system developed by Anderson (1976). As one moves from level I to level IV (table 1) in this scheme there needs to be a significant increase in the information. For example, while Landsat data can be used to determine urban land cover, the ability to identify a specific form of urban land use such as a mobile home would require high resolution resources or field inspection. Jensen provides a useful summary of the sensor requirements to meet the various levels of this hierarchy. (Figure 9)

Table 1 Examples of types of urban land cover at different hierarchical levels

1 Urban or Built-up
- 11 Residential
  - 111 Single-Family Residential
    - 1111 House, houseboat, hut, tent
    - 1112 Mobile home
  - 112 Multiple-Family Residential
    - 1121 Duplex
    - 1122 Triplex
    - 1123 Apartment Complex or Condominium
    - 1124 Mobile home (trailer) park

	Minimum Resolution Requirements		
Attributes	**Temporal**	**Spatial**	**Spectral**
Landuse/Landcover			
L1 – USGS Level I	5 – 10 years	20 – 100 m	V - NIR - MIR - Radar
L2 – USGS Level II	5 – 10 years	5 – 20 m	V - NIR - MIR - Radar
L3 – USGS Level III	3 – 5 years	1 – 5 m	Pan - V - NIR - MIR
L4 – USGS Level IV	1 – 3 years	0.25 – 1 m	Panchromatic (Pan)
Building and Property-Line Infrastructure			
B1 – building perimeter, area, volume, height	1 – 5 years	0.25 – 0.5 m	Stereo Pan - V, LIDAR
B2 – cadastral information (property lines)	1 – 6 months	0.25 – 0.5 m	Stereo Pan - V
Socioeconomic Characteristics			
S1 – local population estimation	5 – 7 years	0.25 – 5 m	Pan -V - NIR, LIDAR
S2 – regional/national population estimation	5 – 15 years	5 – 20 m	Pan -V - NIR
S3 – quality of life indicators	5 – 10 years	0.25 – 30 m	Pan -V - NIR
Energy Demand and Conservation			
E1 – energy demand and production potential	1 – 5 years	0.25 – 1 m	Pan -V - NIR, LIDAR
E2 – building-insulation surveys	1 – 5 years	1 – 5 m	TIR
Transportation Infrastructure			
T1 – general road centerline	1 – 5 years	1 – 30 m	Pan -V - NIR
T2 – precise road width	1 – 2 years	0.25 – 0.5 m	Pan - Visible
T3 – traffic count studies (cars, airplanes, etc.)	5 – 10 min	0.25 – 0.5 m	Pan - Visible
T4 – parking studies	10 – 60 min	0.25 – 0.5 m	Pan - Visible
Utility Infrastructure			
U1 – general utility line mapping and routing	1 – 5 years	1 – 30 m	Pan -V - NIR
U2 – precise utility line width, right-of-way	1 – 2 years	0.25 – 0.6 m	Pan - V, LIDAR
U3 – location of poles, manholes, substations	1 – 2 years	0.25 – 0.6 m	Panchromatic, LIDAR
Digital Elevation Model (DEM)			
D1 – digital surface model and/or digital terrain model	5 – 10 years	0.25 – 0.5 m	Stereo Pan -Visible, LIDAR
D2 – slope map (derived from DTM)	5 – 10 years	0.25 – 0.5 m	Stereo Pan -Visible, LIDAR
Meteorological Data			
M1 – weather prediction	3 – 25 min	1 – 8 km	V - NIR - TIR
M2 – current temperature	3 – 25 min	1 – 8 km	TIR
M3 – clear air and precipitation mode	5 – 10 min	1 – 4 km	WSR-88D Radar
M4 – severe weather mode	5 – 10 min	1 – 4 km	WSR-88D Radar
M5 – monitoring urban heat island effect	12 – 24 hr	5 – 30 m	TIR
Urban Hydrology			
H1 – impervious surface mapping	1 – 5 years	0.25 – 10 m	V - NIR - MIR, LIDAR
H2 – stream centerline and flood plain delineation	1 – 5 years	1 – 30 m	V - NIR - MIR, LIDAR
Critical Environmental Area Assessment			
C1 – stable sensitive environments	1 – 2 years	1 – 10 m	V - NIR - MIR
C2 – dynamic sensitive environments	1 – 6 months	0.25 – 2 m	V - NIR - MIR - TIR
Disaster Emergency Response			
DE1 – pre-emergency imagery	1 – 5 years	1 – 5 m	Pan -V - NIR
DE2 – post-emergency imagery	12 hr – 2 days	0.25 – 2 m	Pan - NIR - Radar
DE3 – damaged housing stock	1 – 2 days	0.25 – 1 m	Pan - NIR
DE4 – damaged transportation	1 – 2 days	0.25 – 1 m	Pan - NIR
DE5 – damaged utilities, services	1 – 2 days	0.25 – 1 m	Pan - NIR

Figure 9 Minimum resolution requirements for different feature attributes (Jensen 2007)

3.1.1.2 Spectral Resolution

While the spatial and temporal dimensions are most critical to the GSS Initiative it is also valuable to access the information content provided by data from different bands of the electromagnetic spectrum. Interesting examples are infrared and thermal band images. For example, infrared imagery is often collected along with natural color orthophotos. Since infrared imagery highlights healthy vegetation it is useful for distinguishing the built environment (Figure 10). Thermal imagery also provides a valuable source of urban information, however it is

generally considered to be too expensive and difficult to process to be a practical solution for large areas (Figure 11). With current technology it is possible to capture imagery with hundreds of narrow bands of the electromagnetic spectrum. An example of such a hyperspectral image is illustrated in Figure 12.

Figure 10 Color infrared image (Richland county SC GIS Department)

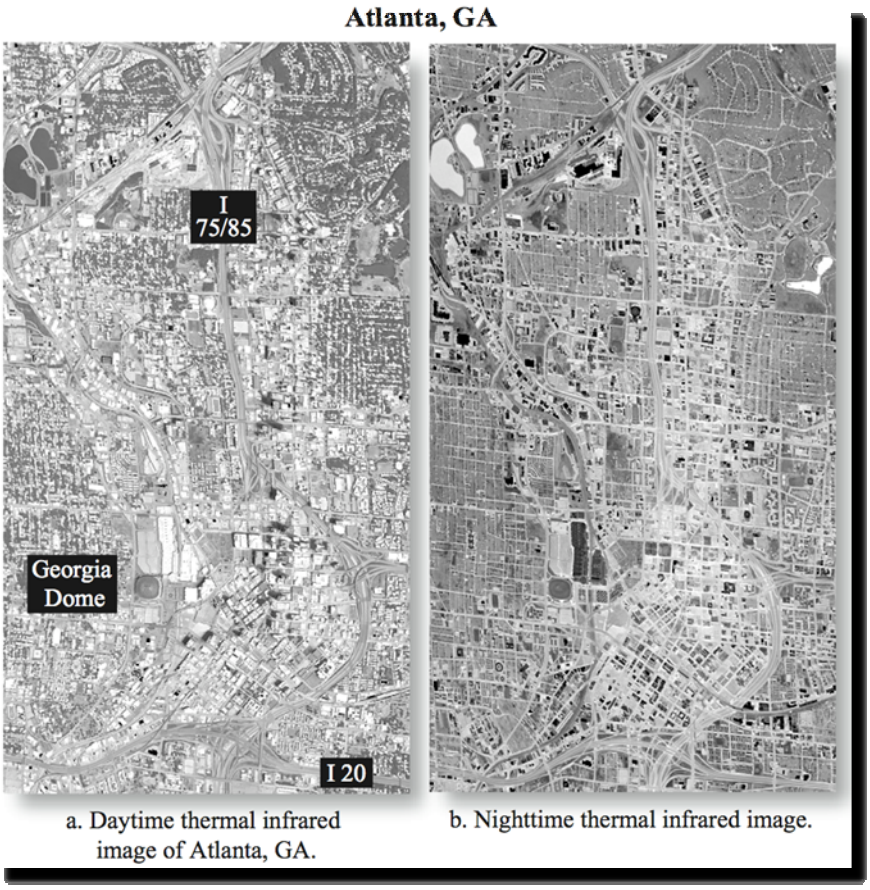


Figure 11 Day and night time thermal images of Atlanta (Quattrochi and Luvall in Jensen, 2007)

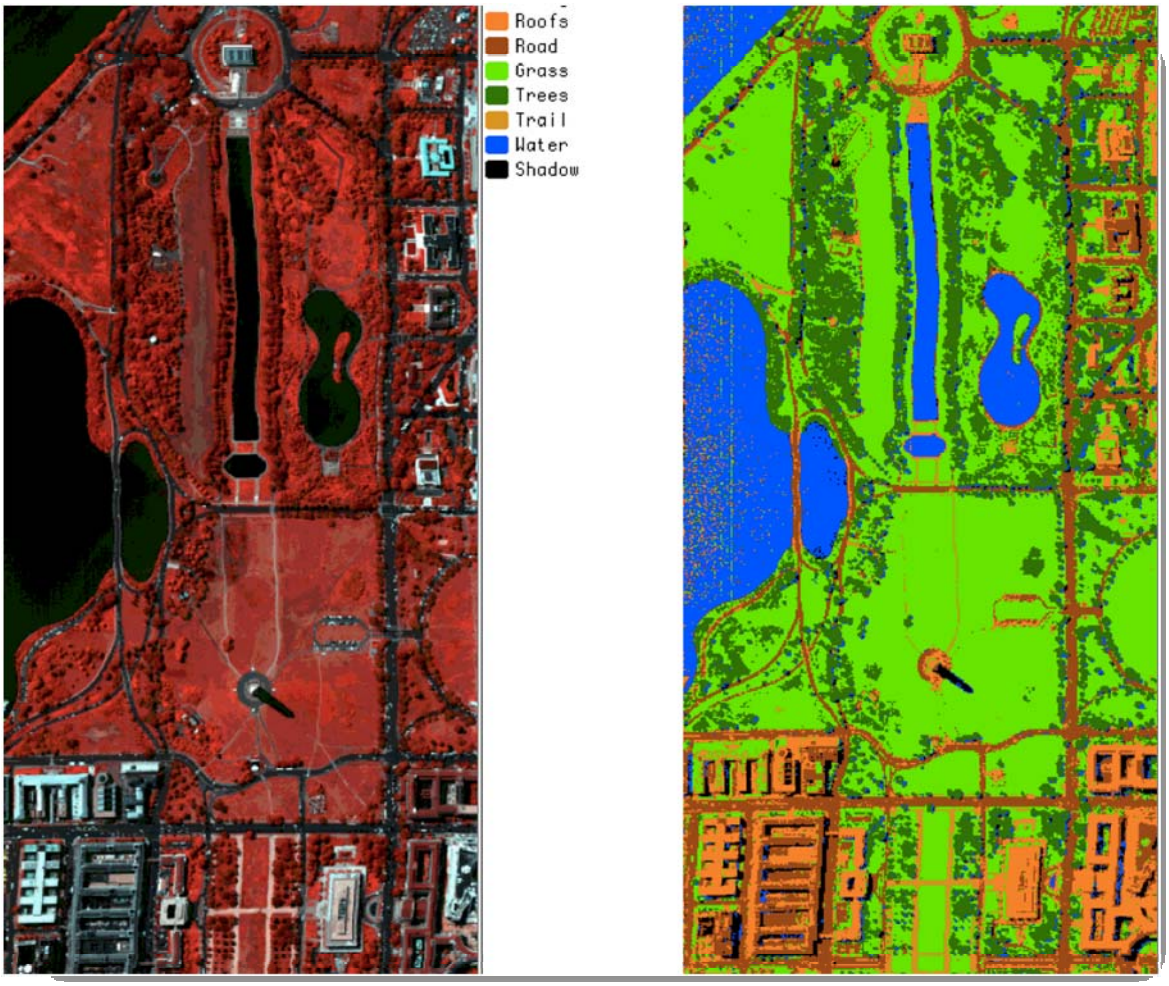


Figure 12 Hyperspectral image of Washington DC (Landgreb in Jensen, 2007)

3.1.2 Update on Current and Future Federal Programs that provide Funds for Imagery Acquisition.

As part of this analysis it is important to understand the federal image acquisition programs. This section provides an update on these programs.

3.1.2.1The National Agriculture Imagery Program (NAIP)

www.apfo.usda.gov/FSA/apfoapp?area=home&subject=prog&topic=nai

The National Agriculture Imagery Program (NAIP) acquires aerial imagery during the agricultural growing seasons in the continental United States. NAIP imagery is acquired at a one-meter ground sample distance (GSD) with a horizontal accuracy that matches within six meters of photo-identifiable ground control points, which are used during image inspection.

A primary goal of the NAIP program is to make digital orthophotography available to governmental agencies and the public within a year of acquisition. NAIP is administered by the USDA's Farm Service Agency (FSA) through the Aerial Photography Field Office in Salt Lake City. This "leaf-on" imagery is used as a base layer for GIS programs in FSA's County Service Centers, and is used to maintain the Common Land Unit (CLU) boundaries. A summary of NAIP plans for 2011 and 2012 were included in the minutes of the November 2010 meeting of the National Digital Orthophoto Program (2010)

3.1.2.2 Imagery for the Nation (IFTN).
(www.fgdc.gov/iftn)

Imagery for the Nation (IFTN) is an initiative originally developed by the National States Geographic Information Council (NSGIC) in 2005. The original goal of the IFTN was to expand on existing federal imagery programs such as NAIP and the National Geospatial-Intelligence Agency (NGA) / U.S. Geological Survey Urban Area Imagery Partnership (UAIP). The objective was to partner with states to create a coordinated program that would generate 1-foot resolution imagery once every three years for 50% of the land mass in all states except Alaska. In addition, 50% matching funds would be available for partnerships to acquire six-inch imagery over urban areas that have populations of at least 50,000 and overall population densities of at least 1,000 people per square mile. The initiative was supported by the FGDC Steering Committee and Executive Committee, and earned the support and endorsement of a number of key professional and advisory groups, including: the National Geospatial Advisory Committee (NGAC), ACSM, ASPRS, AAG, CaGIS, GITA, GIS Certification Institute, IAAO, NSGIC, UCGIS, and URISA. In 2006, NSGIC conducted a survey of its users that indicated that at least thirty states have orthophoto programs. The survey also provides a valuable set of findings about the uses and requirements for high resolution imagery in state and local government. (http://www.nsgic.org/hottopics/iftn/natl_survey_results.pdf)

Following an extensive review that has included a 90 page benefit cost analysis and a detailed record of decision prepared by the FGDC Executive Committee, the IFTN project is stalled. The latest action by the FGDC was to issue a Request for Information (RFI) to industry in the summer of 2010. This RFI is being coordinated by NGA. NGA reported on the RFI at the November 2010 Meeting of the National Digital Orthophoto Program. They included the following conclusions:

- No single vendor is currently capable of meeting both 1m and 30cm requirements.
- Multiple vendor responses could meet either the 1m or 30cm through teaming.
- Existing vendor libraries do not currently meet needs.
- Archive/retrieval capabilities are redundant with current government capability – EROS Data Center and USDA are excellent archival sites; local government organizations could be used as COOP sites by partnering with cooperative agreement.
- Program Management should be handled by the Federal government – Pros and Cons: Outsourcing PM: potential to reduce competition; loss of government expertise; should be accomplished by an independent third party vendor.

In-sourcing PM: current NAIP and Urban Areas model is successful; need additional government resources; advisory committee oversight recommended (NDOP); FAR compliance fosters competition; multi-agency approach ensures buy-in.

- Licensed data with universal access may be cost prohibitive – the desire is for non-proprietary but, the cost is too high.
- 30cm imagery will not meet state or local government needs in urban areas – locals need higher resolution such as 6" or 3"

(National Digital Orthophoto Program, 2010).

Given the current status of the program, it is our opinion that it is highly unlikely that the IFTN will serve as a valuable source of imagery for the GSS Initiative.

3.1.2.3 The National Digital Orthophoto Programs (NDOP)
(www.ndop.gov)

The National Digital Orthophoto Programs (NDOP) was chartered in 1993 as a consortium of Federal agencies with the purpose of developing and maintaining national orthoimagery coverage in the public domain by establishing partnerships with Federal, State, local, tribal, and private organizations. It is now a standing committee of the Federal Geographic Data Committee. Consortium Members consist of the following Federal Agencies

Bureau of Land Management (BLM)
Farm Service Agency Farm Service Agency (FSA)
Federal Emergency Management Agency (FEMA)
National Oceanic and Atmospheric Administration (NOAA)
National States Geographic Information Council (NSGIC)
Natural Resources Conservation Service (NRCS)
U.S. Census Bureau (USCB)
U.S. Department of Agriculture, Forest Service (USDA-FS)
U.S. Geological Survey (USGS)

While the Bureau does not contribute funding to NDOP it is an active participant and has benefited greatly from the program. For example, the Bureau has received imagery for the entire state of North Carolina as a result of its membership (O'Connor, 2011).

3.1.2.4 The Multi-Resolution Land Characteristics (MRLC) Consortium
(www.mrlc.gov/)

The Multi-Resolution Land Characteristics (MRLC) Consortium is a group of federal agencies who first joined together in 1993 (MRLC 1992) to purchase Landsat 5 imagery for the conterminous U.S. and to develop a land cover dataset called the National Land Cover Dataset (NLCD 1992). In 1999, a second-generation MRLC consortium was formed to purchase three dates of Landsat 7 imagery for the entire United States and to coordinate the production of a comprehensive land cover database for the nation called the National Land Cover Database (NLCD 2001). They have also created a National Land Cover Database (NLCD)–Land Cover Change Retrofit (LCCR) product and the NLCD 1992–2001 Land Cover Change Retrofit product that enables one to discover accurate and useful land cover changes than was possible by

direct comparison of NLCD 1992 and NLCD 2000 (Figure 13). In general, most historical Landsat derived land cover data is of limited value to the GSS Initiative. At the same time comparison of Landsat scenes for recent time periods such as the MacDonald Dettwiler and Associates (MDA) Correlated Land Change product is being used for the Department of Homeland Securities National Urban Change Indicator (NUCI) may be useful as an indicator of recent land cover changes. (See further discussion later in this report)

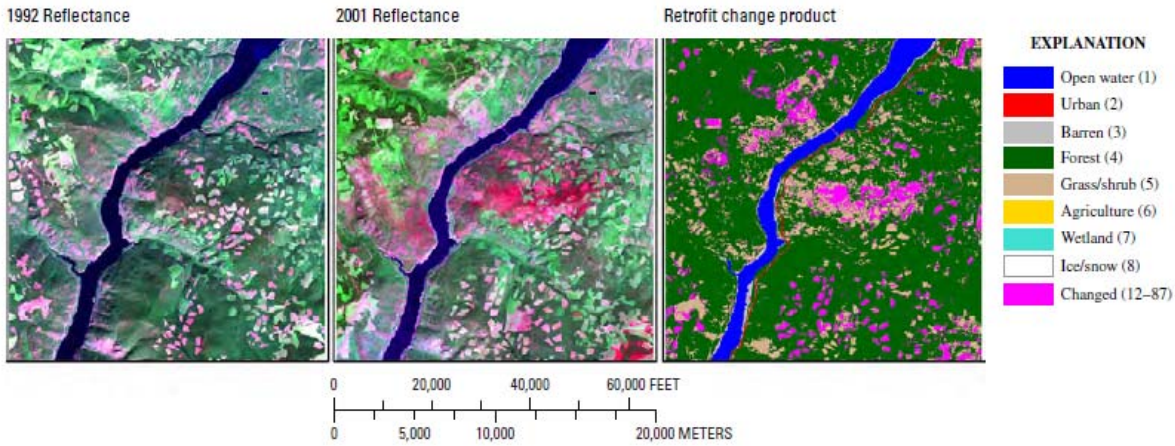


Figure 6. Subset 1. Fire scar and forest harvest cycle near Libby, Montana. Landsat images are displayed in a 6–4–2 band combination. Approximate image chip upper left corner: 115°25′13.21″W, 48°51′29.61″N.

Figure 13 National Land Cover Change retrofit data (MLRC)

3.1.3 Digital Globe / Microsoft Advanced Ortho Aerial Program

While the federal government is still trying to develop a plan for Imagery for the Nation (IFTN), DigitalGlobe has teamed with Microsoft to create a “cloud-free, leaf-off, seamless 30 cm” image base for the nation under its Advanced Ortho Aerial Program. This imagery will be collected in 1 degree by 1 degree cells (~ 10,000 km2) in a grid pattern for complete country coverage, with priority placed on highly populated areas. They expect to complete the nation within 2.5 years and to refresh the data based on priorities every 2.5 years. It is also interesting to note that Microsoft maintains its own photogrammetry division, Vexcel Imaging GmBH. The coverage map for the program and some technical details are included as Figures 14 and 15.

At the November National Digital Orthophoto Program (NDOP) meeting Digital Globe provided a status report on this program (previously called Clear 30) that included the following points:

- Off-the-shelf available as of November 1, 2010
- DG web server will be available around April, 2011
- Standard format and platform (UltracamG), uniform resolution
- Partnership with Microsoft
- Microsoft is responsible for data collection (contracted to aerial companies), aerial triangulation, color correction, ortho rectification, mosaicing, and export

- Digital Globe is responsible for data verification, ingest, product ordering, delivery
- Wall-to-wall coverage by second quarter 2012: CONUS, plus Fairbanks, Juneau and Anchorage (Alaska areas will be summer 2011)
- Flyer Priority

  High Value – most populous areas; High Value Metro areas

  Standard Value – also significant population

  Remote Value – rural

  (National Digital Orthophoto Program, 2010)

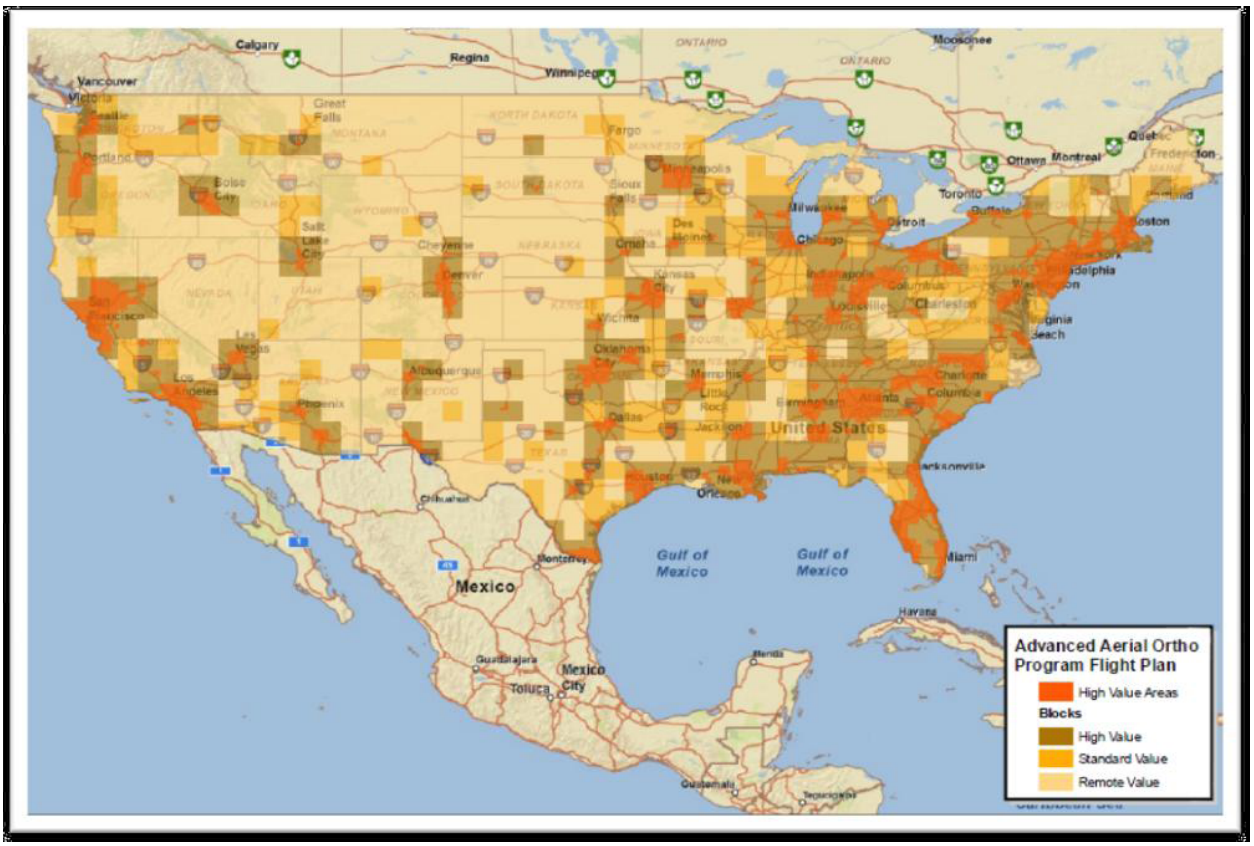


Figure 14 Advanced Aerial Ortho Program Flight Plan (NDOP November 2010)

SPECIFICATIONS

PARAMETERS	VISION	VISION PREMIUM
Product Options	Panchromatic or Natural Color	
Resolution	50 cm, 60 cm	
Bit Depth	8 bit	
Customer Framing	Area-based	
Resampling Kernel	Enhanced (natural color) or MTF (panchromatic)	
File Format	GeoTIFF 1.0, NITF 2.1, NITF 2.0	
Mosaic	Yes if applicable	
Dynamic Range Adjustment	ON	
Color Balancing	Optional for mosaics	Required for mosaics
Projections & Datums	Any supported datums and projection	
Tiling	8k X 8K (default) other options available	
Delivery	DVD, External Drive. FTP	

ACCURACY SPECIFICATIONS			
Orthorectified Product Levels	Display	Mapping	Precision
CE 90%	25.4 m	10.2 m	4.2 m
RMSE	16.5 m	6.6 m	2.8 m
Geographic Availability	Worldwide	Worldwide	Worldwide

Figure 15 Advanced Aerial Ortho Program technical specifications (Digital Globe)
http://www.digitalglobe.com/digitalglobe2/file.php/875/AerialProgram-DS-AP.pdf

3.1.4 Oblique Imagery

3.1.4.1 Pictometry
www.pictometry.com

Many local governments are routinely using very high resolution (6 or 4 inch) oblique imagery to conduct virtual property inspections. The leading supplier is Pictometry that has a fleet of aircraft that utilize five or nine cameras to simultaneously generate vertical and oblique views. For several years, these images have incorporated into Microsoft Bing Map Birdseye view (Figure 16). The data may also be accessed as either an on-line service with parcel boundaries (Figure 17) or as part of a desktop GIS (Figure 18). Pictometry has recently developed an oblique mosaic product (Figure 19). The value of Pictometry data is exemplified by Delaware County, Ohio which acquires standard orthographic and oblique images on alternative years. In

2009, by using its imagery, parcel data, and address points, the county reported that it found more than 10,000 missing addresses (out of 77,000 total addresses in the county) and was able to document them to the Census Bureau (Elami, 2010).

Figure 16 Bing Map Birdseye view bases on Pictometry data with (Bing.com)

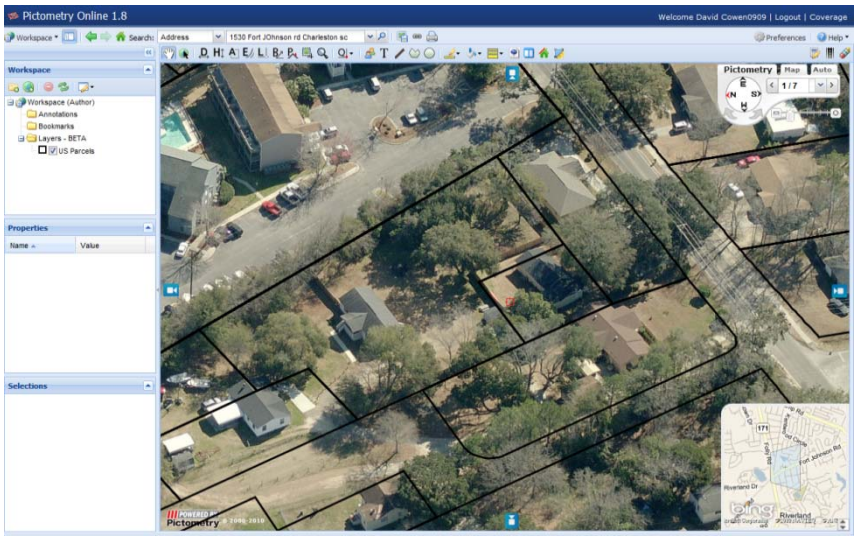


Figure 17 Pictometry on line service with parcel boundaries for a neighborhood in Charleston County SC This view from west corresponds with the image in figure 16. Parcel boundaries would align with ortho view. (Pictometry)

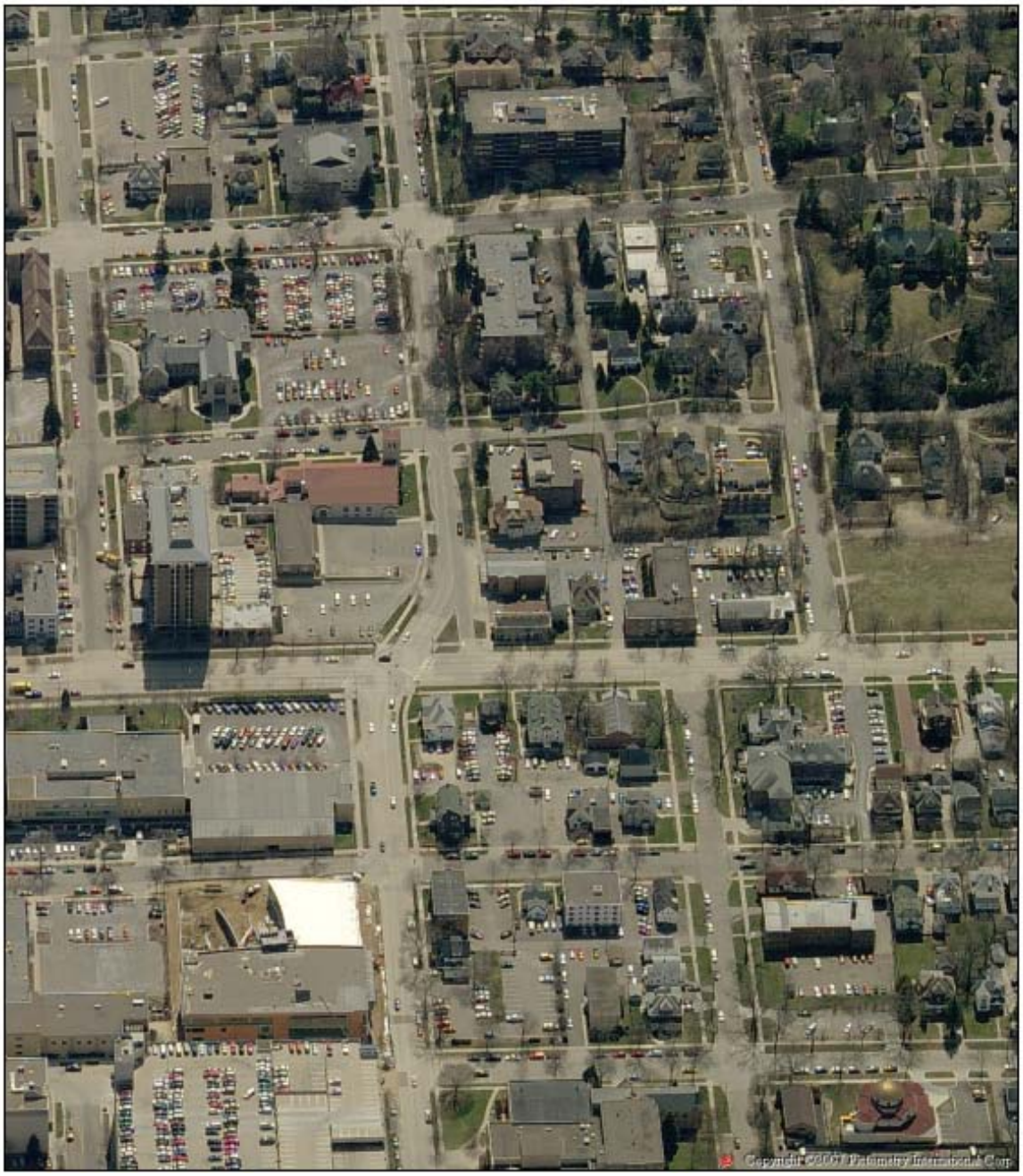


Figure 18 Pictometry oblique mosaic (Pictometry)

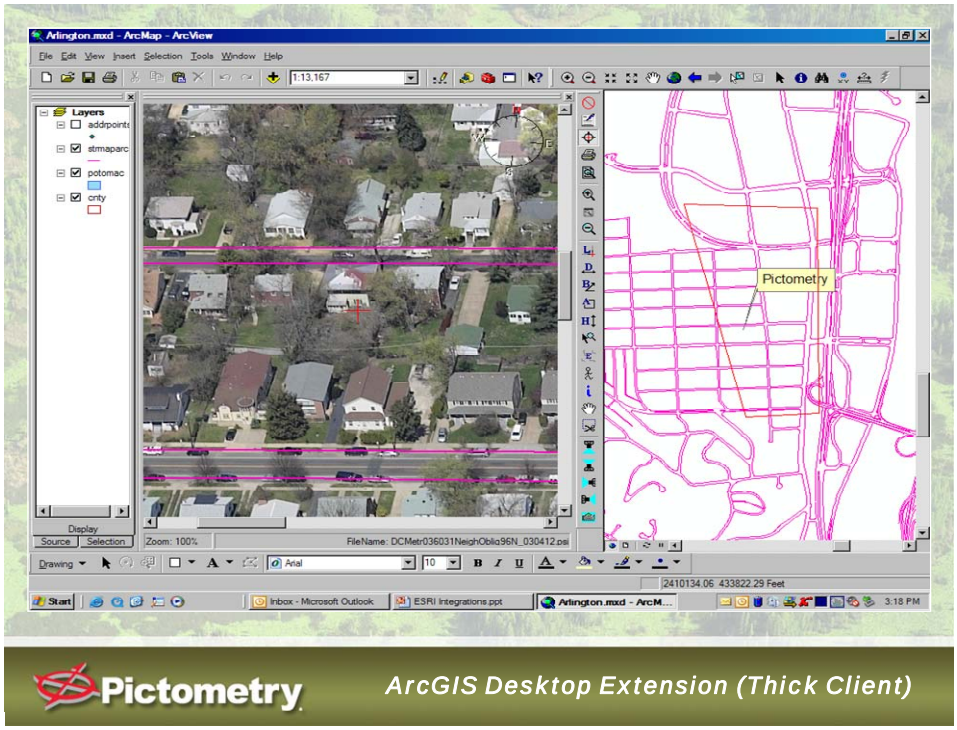


Figure 19 Pictometry ArcGIS Interface (Pictometry)

3.1.4.2 Fugro EarthData

www.fugroearthdata.com

While Pictometry has been the dominate supplier of oblique imagery, other aerial photography firms are entering the market. For example, Fugro EarthData has announced its PanoramiX product that will collect "imagery of the ground with a 104-degree panoramic field of view."

3.1.5 Light Detection and Ranging (LiDAR*)*

According to a recent white paper by Congalton (2010): "Light Detection and Ranging (LiDAR*)* has revolutionized the collection of elevation data and is an incredible source of information that can be used in creating thematic maps. In the last few years, these data have become commercially available and are being used as a vital part of many mapping projects." Prompted by the need to improve flood plain mapping airborne LiDAR has became a standard commodity. It is being used effectively to automatically identify and extract road and building features. Although it may not be available for extensive coverage areas and may not have the proper temporal dimension, LiDAR is an important new geospatial resource that will expand rapidly in the next decade.

3.1.5.1 Overview of LiDAR Technology

LiDAR is a special form of digital imagery generated by an active sensor in the same fashion as radar. Since it captures X, Y, and Z values from thousands of laser pulses it is having a significant impact on how surveying is conducted. Hodgson and Jensen (2010) indicate that in

the past decade the number of pulses per second produced by the sensor has increased from about 25,000 to 500,000 (Figure 20). They also note that LiDAR data represent a discontinuous sample of the Earth's surface while imagery is a wall-to-wall continuous sample. Multiple observations in imagery (e.g. bands) are at same *X-Y* position while the different LiDAR returns represent different locations and features. Various LiDAR returns and intensity measures can be used for image classification and automated feature extraction (Figures 21and 22). Since the LiDAR data are primarily used to generate elevation data it is useful to understand some current definitions.

- *Digital elevation model* (*DEM*) is defined as a file or database containing elevation points over a contiguous area. DEMs may be subdivided into:
    - *Digital surface models* (DSM) that contain elevation information about all features in the landscape, such as vegetation, buildings, and other structures, and
    - *Digital terrain models* (DTM) that contain elevation information about the bare-Earth surface without the presence of vegetation or man-made structures.

Digital surface models are an excellent source for change detection. For example, any change in the Z dimension between time T1 and T2 is a good indication of change on the ground.

Jensen (2007) provides following points regarding the collection of LiDAR (p 336 – 349):

> The first optical laser was developed in 1960 Hughes Aircraft Inc. Laser instruments were soon used to compute distance by measuring the travel time of light from a laser transmitted to a target and back to a laser receiver. The synergistic use of kinematic GPS and inertial measurement units on airborne LiDAR scanning systems has allowed the technology to mature rapidly. LiDAR derived horizontal and vertical accuracies and cost of operation are now similar to that of photogrammetry.
>
> The LiDAR instruments consist of a system controller and a transmitter and receiver. As the aircraft moves forward along the line-of –flight, a scanning mirror directs pulses of laser light across a track perpendicular to the line of flight. Most LiDAR systems used for topographic mapping use eye-safe near infrared laser light in the region from 1040 to 1060nm. LiDAR can be collected at night if necessary because it is an active system and is not dependent on passive solar illumination.
>
> As mentioned, a pulse of laser energy exiting the transmitter is directed toward the terrain at a certain angle by the rotation mirror. Depending upon the altitude of the LiDAR and the angle at which the pulse is sent, each pulse illuminates a near circular area on the ground called the instantaneous laser footprint. This single pulse can generate one return or multiple returns. If there are any materials whatsoever with local relief within the instantaneous laser foot print (e.g. grass, small rocks twigs) then there will be multiple returns. The first return will come from the materials that have local relief and the 2nd and perhaps last return will come from the bare-earth (Figure 23).
>
> Most LiDAR systems provide an intensity file in addition to the multiple return data. The recorded intensity is, in most cases, not the integration of the retuned echo from all the pulse returns but just the maximum.

Clearly the detailed LiDAR returns can be used to generate some amazing visualizations such as a recent image of Central Park (Figure 24). More importantly, by analyzing the different returns they can be used to penetrate even dense vegetation to reveal the bare earth (Figure 25). From an operational GIS viewpoint the LiDAR data can be incorporated into digital elevation data set and analyzed or visualized any number of ways (Figure 26)

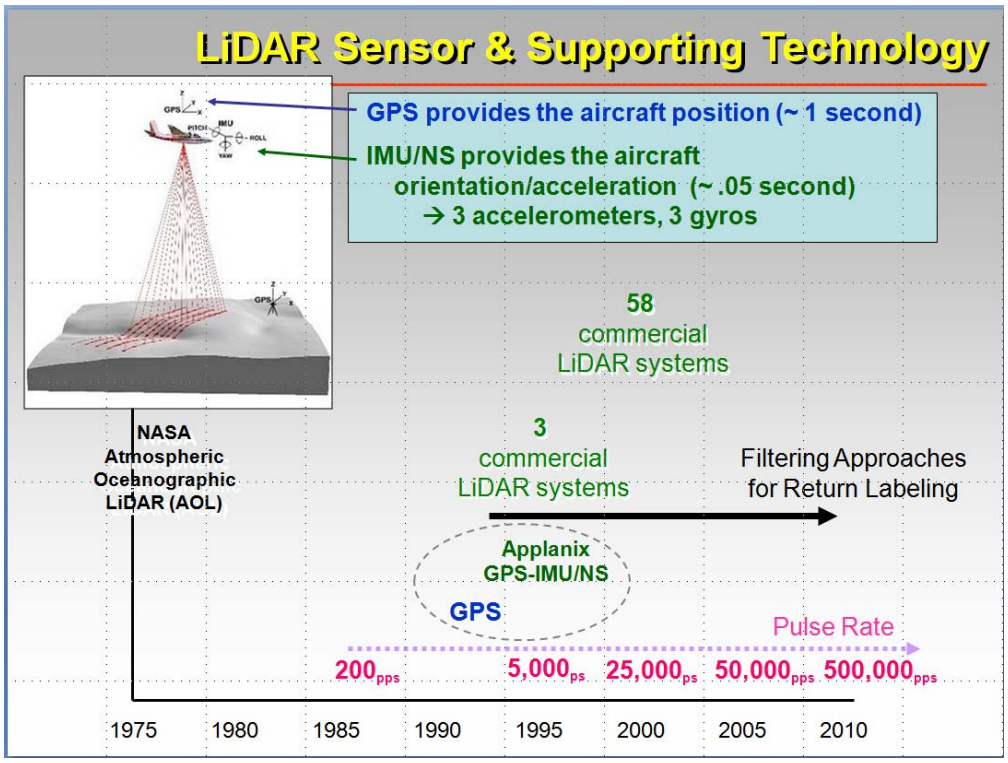


Figure 20 Evolution of LiDAR sensor technology (Hodgson and Jensen, 2010)

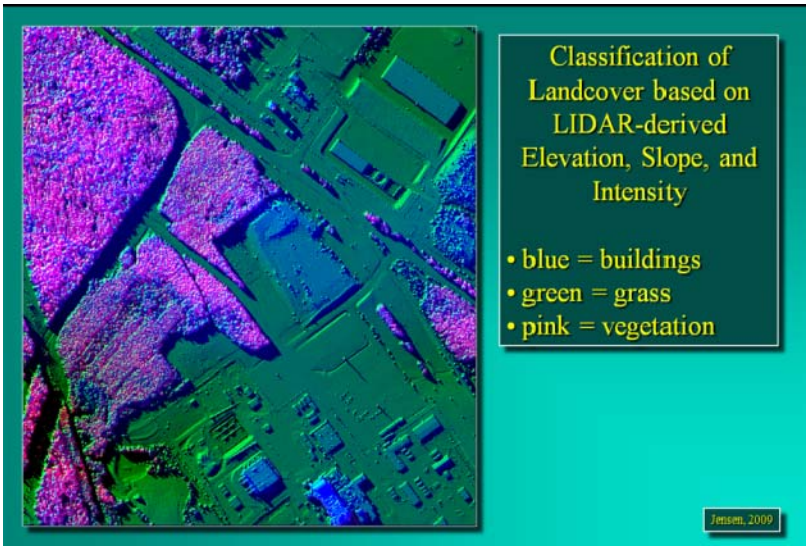


Figure 21 Landcover classes from LiDAR (Jensen, 2007)

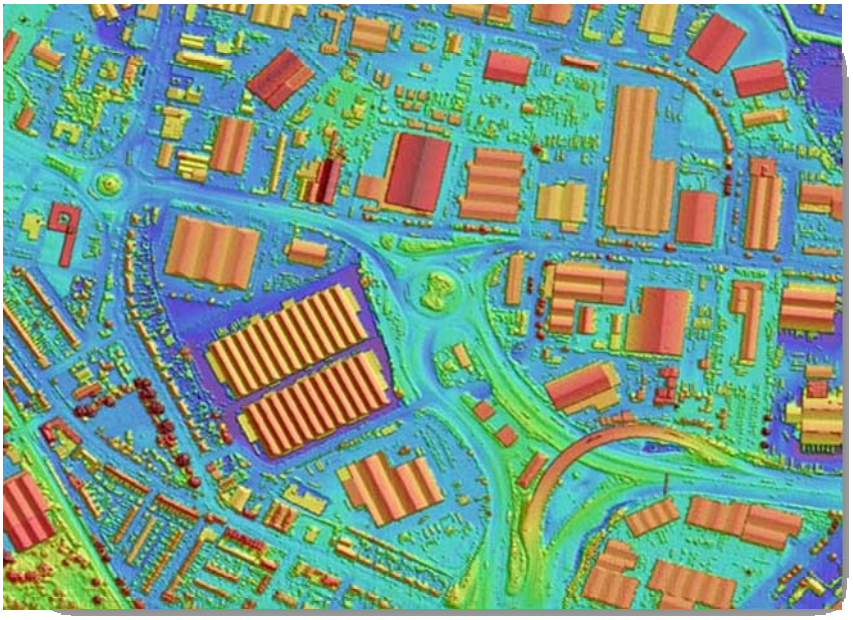


Figure 22 LiDAR display of urban features (Jensen, 2007)

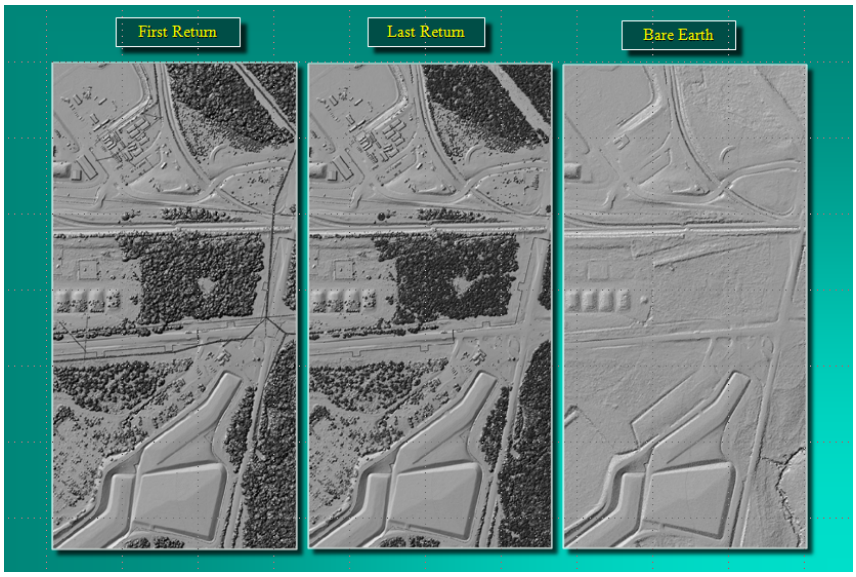


Figure 23 First return, last return and bare earth from LiDAR (Jensen, 2007)

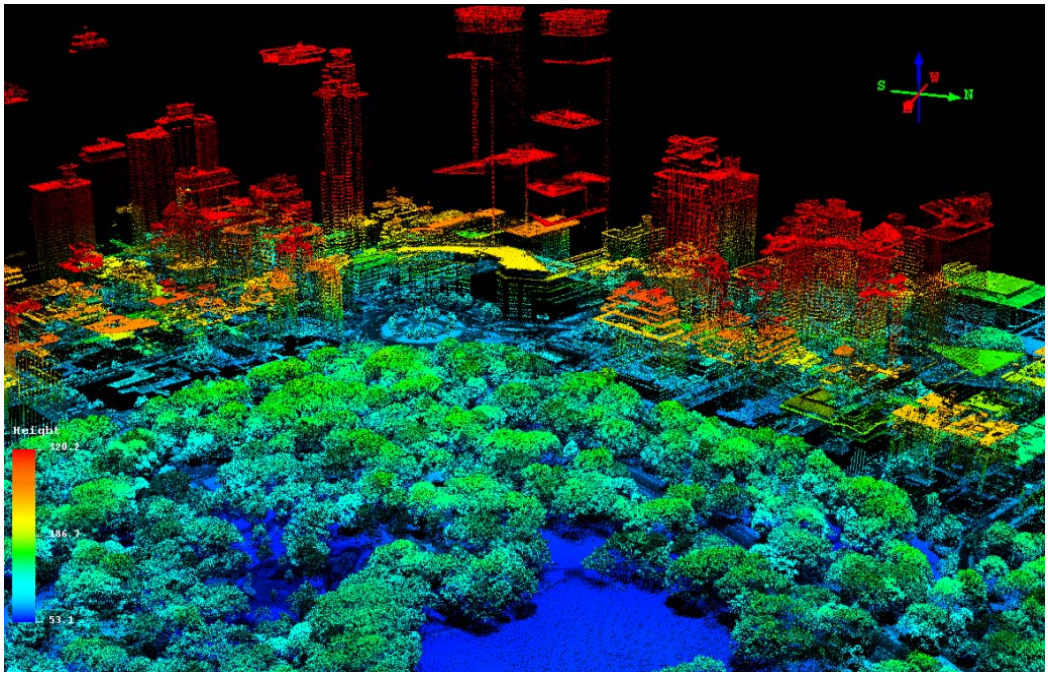


Figure 24 LiDAR elevation model and intensity measures for Central Park (Courtesy of Sean Ahearn)

FIGURE 3.7 These false-color geodetic images of a landslide near Flathead Lake, Montana, were made from airborne LiDAR data collected by the NSF National Center for Airborne Laser Mapping (NCALM). Laser shots passed through openings in the forest and reached the ground, allowing a filter to be used to remove the returns from the trees and reveal the landslide, which is not normally visible to the eye or a camera. The white scale bars are 500 meters in length. SOURCE: Ramesh Shrestha, National Center for Airborne Laser Mapping (NCALM), University of Houston; see also Carter et al., 2007.

Figure 25 Illustration of ability of LiDAR data to penetrate vegetation (NRC, 2010 b)

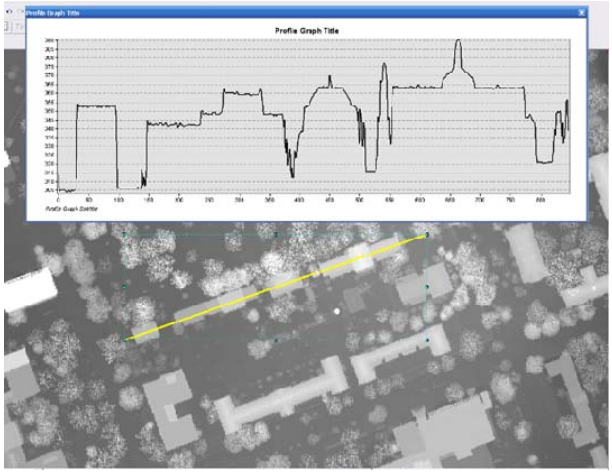


Figure 26 Cross section of University of South Carolina campus based on LiDAR derived elevation model (Cowen, unpublished)

3.1.5.2 Federal Organizational Issues Relating to LiDAR

Within the Federal government, FEMA has taken the lead in funding the acquisition of LiDAR data for use in creating Flood Insurance Rate Maps. At the same time, the USGS is the steward of National Elevation Data (NED) and operates the Center for LiDAR Information Coordination and Knowledge (CLICK). The National Digital Elevation Program (NDEP) is a consortium of agencies coordinating the collection and application of high-resolution, high accuracy elevation data. The NDEP consists of:

- Bureau of Land Management
- Federal Emergency Management Agency
- U.S. Fish and Wildlife Service
- National Aeronautic and Space Administration
- National Geospatial-Intelligence Agency
- National Oceanic and Atmospheric Administration
- National States Geographic Information Council
- USDA Natural Resources Conservation Service
- U.S. Army Corps of Engineers
- ***U.S. Census Bureau***
- USDA Forest Service
- U.S. Geological Survey

Since protocols and standards are still evolving, we believe the Bureau should take an active role in the NDEP. This includes the *Current Assessment of the Business Requirements and Benefits of enhanced National Elevation Data* that is being conducted by Dewberry for the USGS. For example, most of the federal and state users are concerned about floodplain mapping while the GSS Initiative is concerned with features on the surface of the earth. Therefore it is essential to retain the all returns and intensity measures from LiDAR missions that assist in the identification of buildings and roads.

The current status of LiDAR coverage is represented in Figure 27. The USGS has developed an idealized schedule for the Enhanced National Elevation Data Program:

- 2010 - Initiate assessment of business requirements for enhanced national elevation data.
- 2011 - Complete assessment in summer of 2011 and process new elevation data acquired through American Recovery and Reinvestment Act (ARRA) funding.
- 2012 - Establish management structure and implementation plan for an enhanced elevation program.
- 2013 - Begin national program implementation.
- 2014 – Achieve operational program status

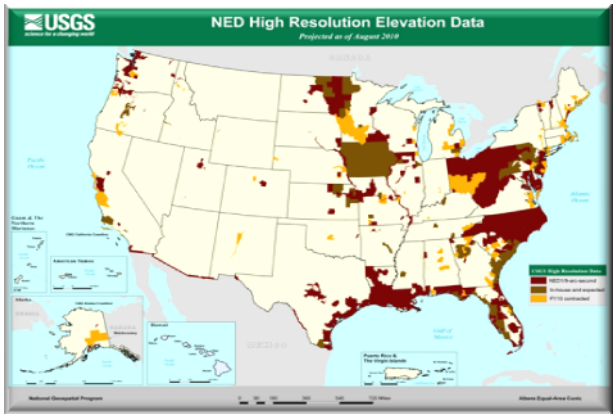


Figure 27 Current status of high resolution (LiDAR) (USGS)

3.1.5.2.1 NGA's LiDAR Data for Greater Washington DC

In 2008, NGA funded a LiDAR mission covering approximately 8,253 square kilometers of the Washington DC metropolitan region and a second more detailed coverage of the Mall area. According to the metadata there are no use constraints on the data and it can be obtained on DVD. From the perspective of evaluating LiDAR utility for the GSS Initiative, it is interesting to examine the 2D and 3D building data that were extracted from the LiDAR collection (extraction also funded by NGA). As the metadata describes:

> "This data was collected at a resolution of 1 meter and includes reflective surface, last return, bare earth model, and intensity data in separate data files. This raw LIDAR data was then exploited using SAIC proprietary Automated Feature Extraction software that automatically detects and delineates features present in the LIDAR data. These features include 3D building features, tree points and forest polygons. This file represents individual components of building structures as detected in the LIDAR data. Components are related to building "footprints" via the "Bldg_id" attribute. Extraction of individual elements makes the resulting visualization of data in a 3D viewing environment more realistic and accurate." (NGA, Metadata)

The building outlines were provided to the District of Columbia (District of Columbia Geographic Information System Steering Committee, 2010). The District has been assessing the value of these building outlines with those created by conventional stereo methods (Figure 27.1)

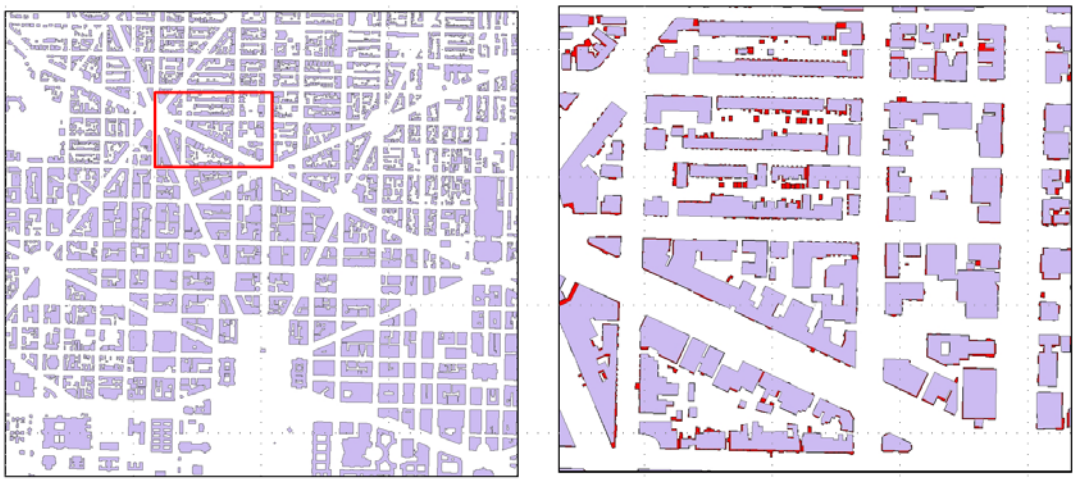


Figure 27.1 Comparison of LiDAR Derived and Orthophoto Stereo Derived buildings – Red represents differences (Courtesy of DC GIS)

The DC GIS office provided a tile with 21,929 3D building outlines for analysis in this report (Figure 27.2). The data were opened by ArcGIS and overlaid with data from a TIGER web service (Figure 27.3a). Since the data include several elevation attributes and roof shapes, it is possible to portray them as 3D features (Figures 27.3b 27.3c and 27.3d). The value of the 3D LiDAR data is highlighted by the ability to identify structures that may be surrounded by forests or other obstructions that might impede visibility from the road (Figure 27.4). While this particular project was funded by NGA, it is representative of the type of large scale LiDAR projects that can be expected to be conducted in this decade.

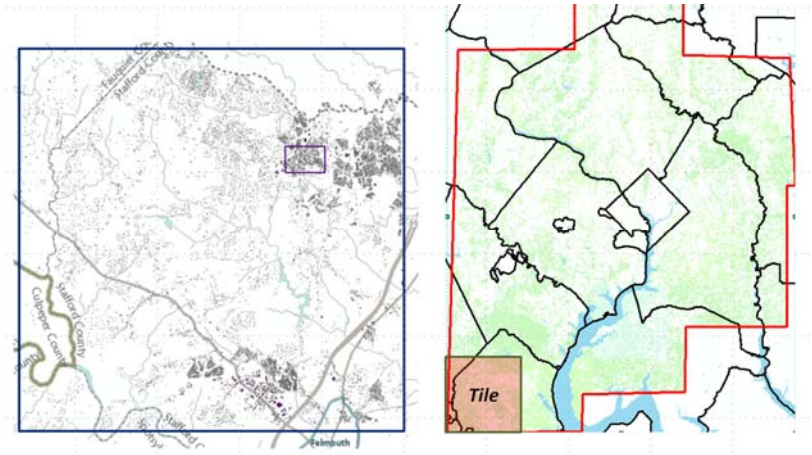


Figure 27.2 NGA LiDAR Coverage Area And Tile Used For Analysis (Cowen)

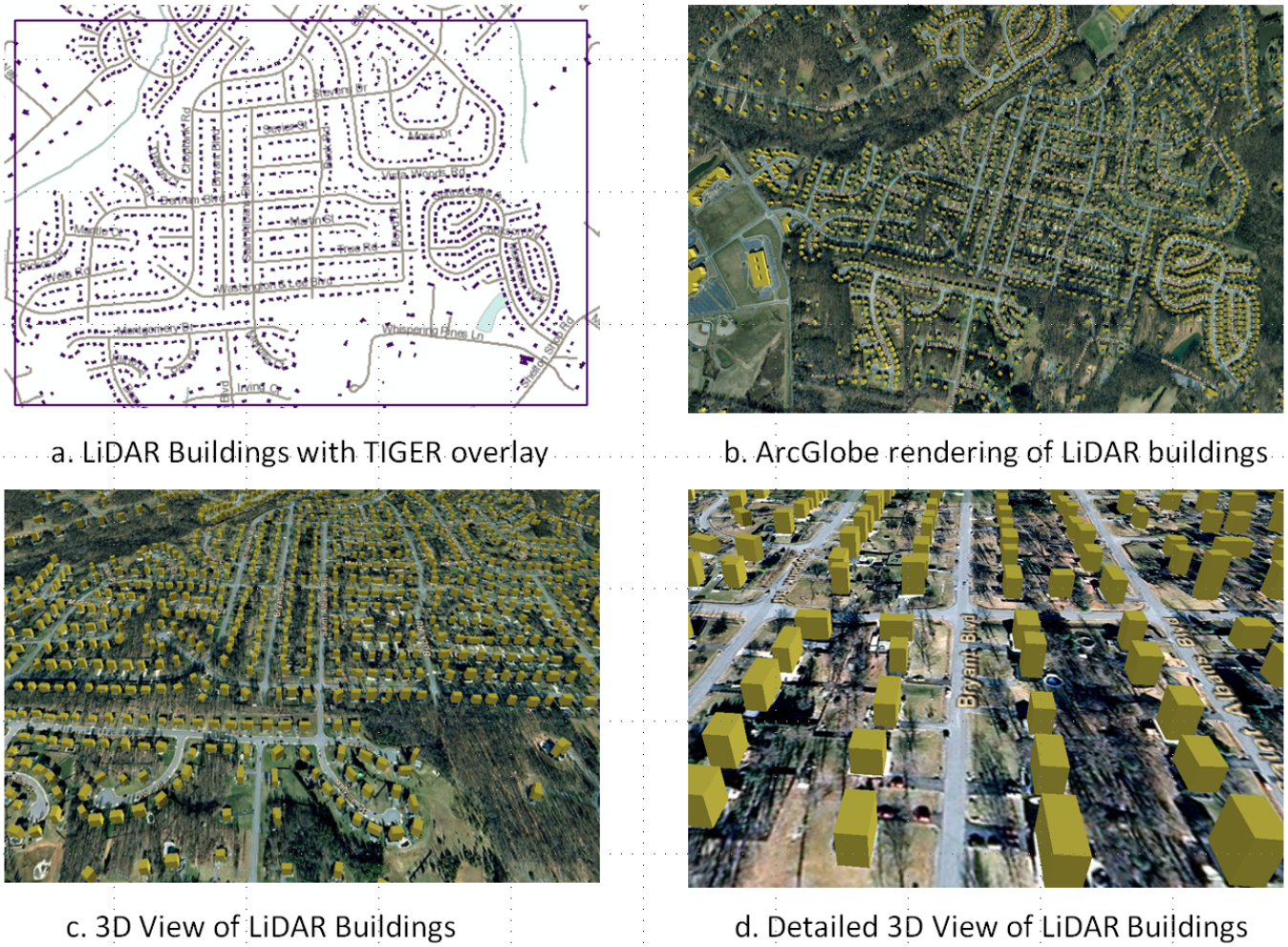


Figure 27.3a – d LiDAR 3D Buildings Portrayed with TIGER Web-Service and ArcGlobe 3D Visualizations (Cowen)

Figure 27.4 Example of 3D buildings surrounded by dense forest (Cowen)

3.1.5.3 Terrestrial and Mobile LiDAR Systems

Terrestrial and mobile LiDAR systems are quickly emerging as a viable alternative to conventional surveying procedures. These devices are able to collect coordinate data faster and more safely than traditional survey methods. In some engineering situations a terrestrial LiDAR device is mounted at a strategic location and used to create a 360 degree survey (Figure 28). In other cases the device is mounted on a moving vehicle (Figure 29). The mobile option is increasingly popular for collecting surveys in inaccessible or dangerous locations. Hodgson and Jensen (2010) suggest that the most recent mobile units are pushing the technology in terms of the ability to process the data volumes. They believe that these mobile units, which can generate .01 foot spacing, are capable of generating "1000X the data in 10% of the time compared to in situ surveying".

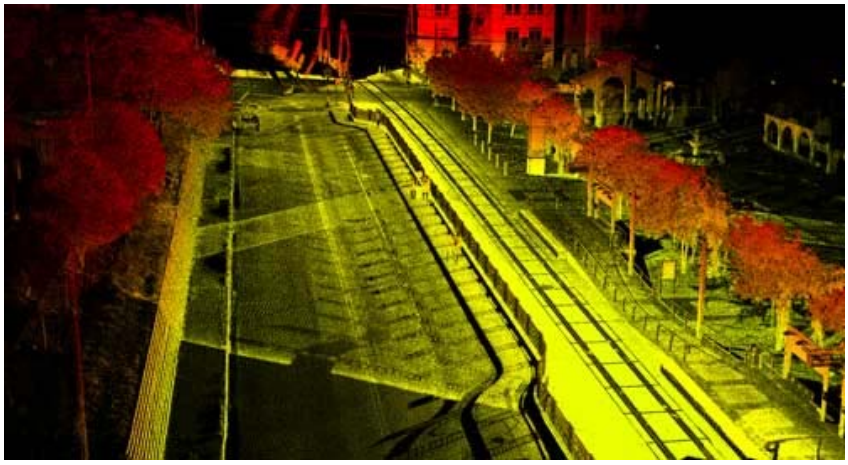


Figure 28 Terrestrial LiDAR example (Renslow 2010)

Figure 29 Mobile LiDAR vehicle (Lynx)

It is interesting to note that NAVTEQ is conducting a nation-wide mobile LiDAR project that could have an impact on the GSS Initiative. The JourneyView program that is scheduled to be launched in 2011 includes a combination of interactive 360 degree street-level imagery (Figure 30) linked to the NAVTEQ map and point of interest (POI) content. As part of this program NAVTEQ's Trueprogram includes a significant mobile LiDAR effort as described:

NAVTEQ's LIDAR collection uses a proprietary architecture that can capture over 1.5 million 3D data points every second, which is a significant advancement over other LIDAR systems. The rotating lasers can collect over a longer range, up to approximately 120 meters, and at higher speeds, allowing NAVTEQ to collect highway signs while driving posted speed limits while also still being able to capture the detail on a small mile marker. While other LIDAR systems use single line scans which collect to an angle of 90 degrees, NAVTEQ's 64 rotating lasers collect 360 degree LIDAR images. Other features of this system include the ability to automatically generate attributes such as bridge height and lane width down to the centimeter, and to look behind objects to better generate accurate dimensions in the execution of 3D representations versus those seen when using sources such as aerial imagery.
(http://corporate.navteq.com/webapps/NewsUserServlet?action=NewsDetail&newsId=836&lang=de&englishonly=true)

Figure 30 Example of NAVTEQ mobile LiDAR (navteq.com)

Renslow (2010) recently reported on a related mobile 3D geospatial video capture program the Japanese Iwane Laboratory (Figure 31). This system is also being incorporated into a linear feature extraction system (Figure 32). These systems may represent the next generation of mobile data collection systems.

Figure 31 Example of Iwane Laboratories 3D geospatial vdeo (Renslow, 2010)

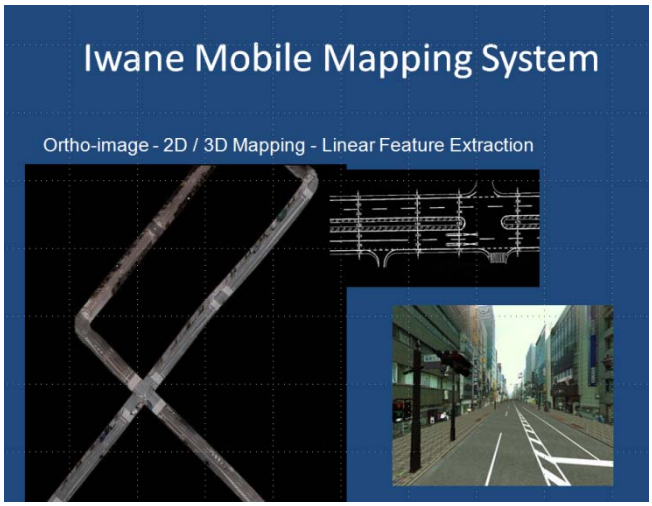


Figure 32 Iwane Laboratories mobile mapping system (Renslow 2010)

3.1.5.4 Research Findings regarding LiDAR

The following section reviews some recent studies that have utilized LiDAR for image classification.

Tullis, et al.,(2010) provide a useful approach to integrating LiDAR with high resolution imagery to implement a rule based classification process to extract impervious, forest, and herbaceous classes (Figures 33 and 34).

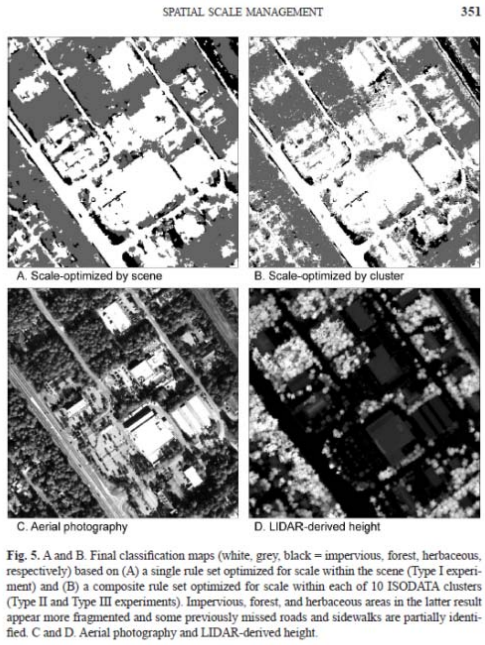


Figure 33 Data inputs to classification process (Tullis et al. 2010)

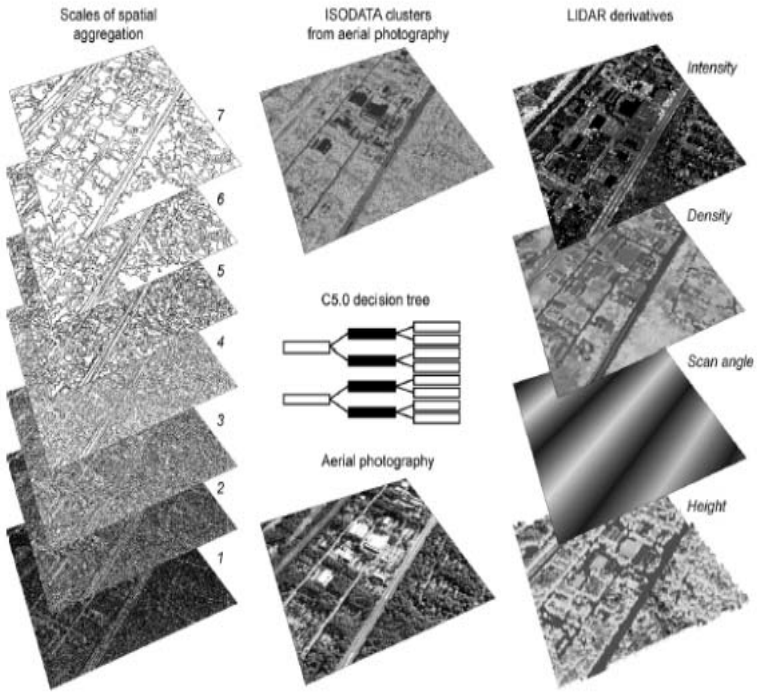


Fig. 4. Optical image– and LIDAR-derived input layers, scales of spatial aggregation (1–7), and unsupervised clusters covering the 0.6 × 0.6 km study area on Hilton Head Island, SC. Machine learning decision trees were used to create the rule sets that were subsequently evaluated in terms of their reliability and economy.

Figure 34 Decision tree process for classification based on LiDAR and imagery (Tullis et al. 2010)

Awrangjeb et al. (2010) describe a process for automatic detection of residential buildings using LiDAR data and multispectral imagery (Figure 35). It is experimentally shown that the proposed technique can successfully detect urban residential buildings, when assessed in terms of 15 indices including completeness, correctness, and quality.

M. Awrangjeb et al. / ISPRS Journal of Photogrammetry and Remote Sensing 65 (2010) 457–467

Fig. 6. Detected buildings on the orthoimages.

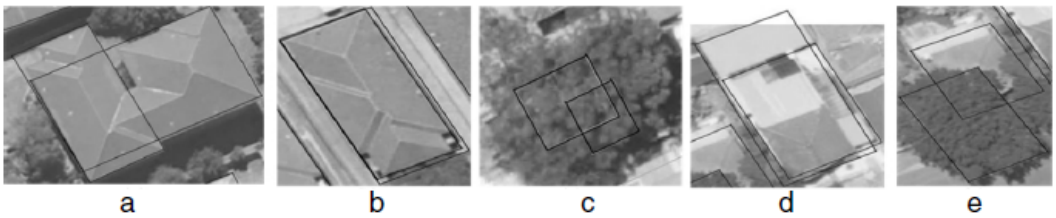


Fig. 7. Different types of detection overlaps: (a) natural, (b) multiple detection, (c) false–false, (d) true–true and (e) true–false.

Figure 35 Automated building extraction from LiDAR (Awrangjeb et al. (2010)

Mitishita, et al., (2008) describe an interesting method to automatically generate the centroid of a roof from LiDAR and use it as a control point for georeferencing other images (Figure 36).

FIG. 6. Profiles close to the building border.

FIG. 7. Roof corner points used to determine the 2D coordinates of the centroid of the roof in the image frame.

(a) (b) (c)

FIG. 14. Examples of building layouts used for the determination of the centroids: roofs connected to other roofs on one side (a, b) and a roof connected to other roofs on two sides (c).

Figure 36 Process for generating building centroids from LiDAR data (Mitshita et al. 2008)

Im et. al. (2008) demonstrate how LiDAR derived information such as height, intensity, and shape of features found in the scene can be used to classify land cover (Figures 37 and 38). They found that by using machine-learning decision trees these metrics yielded land cover classification accuracies > 90%.

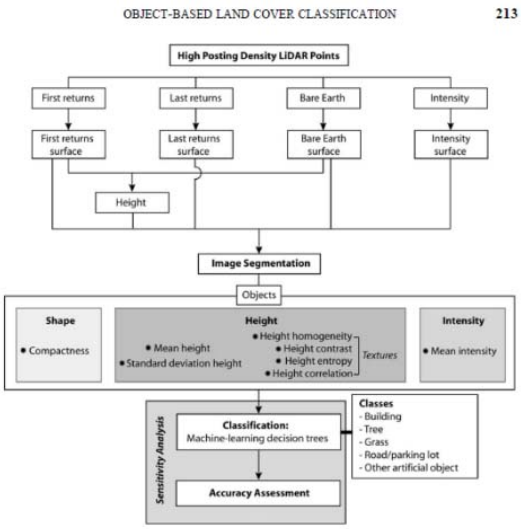


Fig. 2. Flow diagram of the methodology used in the study.

Figure 37 LiDAR based process for object based classification (Im et al. 2008)

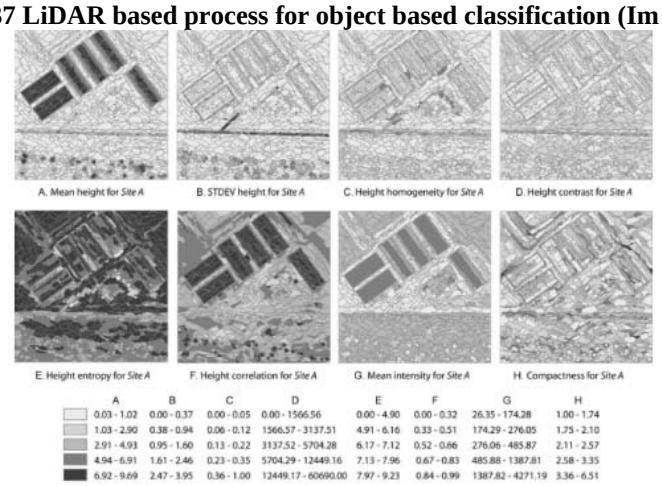


Fig. 4. Spatial distribution of the eight object-based metrics for Site A.

Figure 38 Example of outputs from the object based classification (Im et al. 2008)

3.1.6 Analysis of Existing and Planned Earth Observation Satellites

3.1.6.1 Overview of Existing Satellite Based Sensors

Renslow (2010) estimates that there are currently more than 40 earth observation satellites with more than 60 instruments. A complete list of orbiting sensor earth observation platforms can be found at the Committee on Earth Observation Satellites (CEOS).
http://directory.eoportal.org/missions_all_list.php?&view_all.
It is possible to track most of these earth observation satellites through the University of South Carolina's Remote Sensing Hazards Guidance System (RSHGS). Using the RSHGS web site (http://ww2.rshgs.sc.edu/Home.aspx) it is possible to query the current and future locations and collection abilities of any unclassified satellite-borne sensor. It is important to note that many of these sensors are designed to measure soil, vegetation, and land cover at spatial resolutions that would not assist the GSS Initiative. (The Landsat program is discussed elsewhere). Alternatively,

there are two commercial remote sensing platforms that are being utilized to detect and monitor buildings and transportation features. These are the GeoEye 1 (Figure 39) with a multispectral sensor with a spatial resolution of 1.65 m and the panchromatic sensor with a spatial resolution of 0.41 m and DigitalGlobe's Worldview 2 (Figure 40) with a panchromatic resolution of 0.46 m and multispectral resolution of 1.84 m. It is important to note that current US defense policy requires the data to be resampled to a resolution of .5 meters[1].

SATELLITE CHARACTERISTICS

SatelliteAbbr	LaunchDate	Country	LaunchSite	Period_min	Incl_deg	SatelliteOperational
GEOEYE 1	9/6/2008 12:00:00 AM	USA	VAFB	98.3	98.12	Y

SENSORS ON SATELLITE

SensorAbbr	Num_Bands	SpatialResolution_Min_m	SwathWidth_km	OffNadirMax_deg	AlongTrackMax_deg	SensorOperational
GeoEye-1	5	0.41	15.2	60	0	Y

BANDS ON SATELLITE

SatelliteAbbr	SensorAbbr	BandType	StartWavelength	StopWavelength	Wavelength_Unit	SpatialResolution_min_m	BandOperational
GEOEYE 1	GeoEye-1	Blue	0.45	0.51	um	1.65	Y
GEOEYE 1	GeoEye-1	Green	0.51	0.58	um	1.65	Y
GEOEYE 1	GeoEye-1	Red	0.655	0.69	um	1.65	Y
GEOEYE 1	GeoEye-1	NIR	0.78	0.92	um	1.65	Y
GEOEYE 1	GeoEye-1	PAN	0.45	0.8	um	0.41	Y

Figure 39 GeoEye - 1 technical specifications (ww2.rshgs.sc.edu)

SATELLITE CHARACTERISTICS

SatelliteAbbr	LaunchDate	Country	LaunchSite	Period_min	Incl_deg	SatelliteOperational
WORLDVIEW 2	10/8/2009 12:00:00 AM	US	AFWTR	100.19	98.55	Y

SENSORS ON SATELLITE

SensorAbbr	Num_Bands	SpatialResolution_Min_m	SwathWidth_km	OffNadirMax_deg	AlongTrackMax_deg	SensorOperational
Panchromatic	1	0.46	65.6	45	0	Y
Multispectral	8	1.84	65.6	45	0	Y

BANDS ON SATELLITE

SatelliteAbbr	SensorAbbr	BandType	StartWavelength	StopWavelength	Wavelength_Unit	SpatialResolution_min_m	BandOperational
WORLDVIEW 2	Panchromatic	PAN	0.45	0.8	um	0.46	Y
WORLDVIEW 2	Multispectral	Blue	0.4	0.45	um	1.84	Y
WORLDVIEW 2	Multispectral	Blue	0.45	0.51	um	1.84	Y
WORLDVIEW 2	Multispectral	Green	0.51	0.58	um	1.84	Y
WORLDVIEW 2	Multispectral	Red	0.585	0.625	um	1.84	Y
WORLDVIEW 2	Multispectral	Red	0.63	0.69	um	1.84	Y
WORLDVIEW 2	Multispectral	NIR	0.705	0.745	um	1.84	Y
WORLDVIEW 2	Multispectral	NIR	0.77	0.895	um	1.84	Y
WORLDVIEW 2	Multispectral	NIR	0.86	1.04	um	1.84	Y

Figure 40 Digital Globe Worldview 2 technical specifications (ww2.rshgs.sc.edu))

[1] Ikonos, launched in 1997 is expected to be retired in late 2011; and OrbView-2, launched in 1997 is now of marginal value

3.1.6.2 Planned Programs that will Impact GSS Initiative

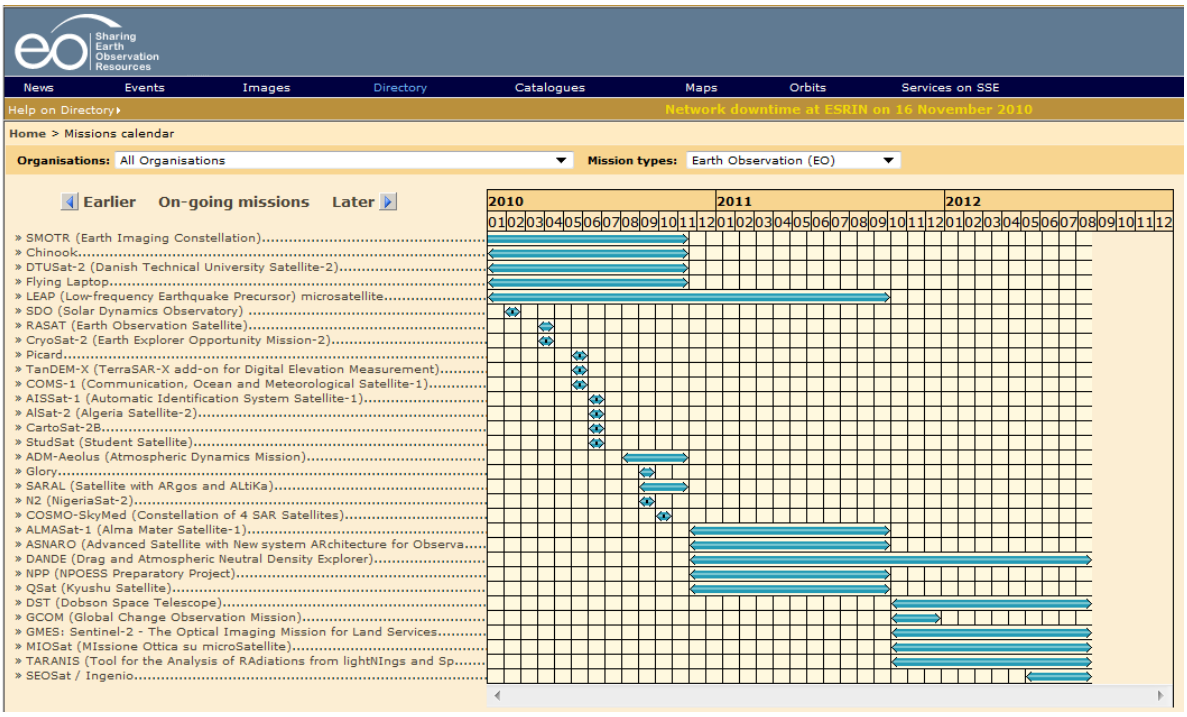
From the perspective of the GSS Initiative there are two major planned commercial earth observation satellite missions scheduled in the next four years that are of interest. Both of these are part of the National Geospatial-Intelligence Agency's Enhanced View program. Under this program NGA is providing more than $3.5 billion each to help them develop and launch two new satellites. DigitalGlobe was given a $3.55 billion 10-year contract to support the scheduled launch of Worldview 3 in 2014 and GeoEye, received a commitment of $3.8 billion to launch GeoEye-2 by 2013. The support from NGA is critical for the support of the US commercial satellite business. Hopefully these platforms will provide the type of current coverage that would support the GSS Initiative. It is strongly recommended that the Geography Division communicate with NGA regarding access to imagery from these platforms.

3.1.6.2.1 SPOT 6 and SPOT 7

It is noted that the French Spot Image Group, which pioneered commercial remote sensing in 1986, plans to launch SPOT 6 and SPOT 7 in the next few years. These satellites will have a panchromatic band with a spatial resolution of 1.5 meter that could provide a useful image for rural to urban land cover detection (see http://www.spot.com).

3.1.6.2.2 Schedules for other Earth Observation Satellites

A schedule of all future earth observation satellite missions is maintained by the Earth Observation Portal www.eoPortal.org **and is shown below:**



http://directory.eoportal.org/missions_calendar.php?page=2011&mtype=2&owner=0

3.1.7 Summary

Recently collected high resolution imagery will be an important resource for the GSS Initiative and fortunately there should be greatly expanded access to these data. Both aircraft and satellite based platforms can provide resolutions of less than a meter that are adequate to meet requirements for the update of structure points and road features. Many of these image sources are available through free web based mapping applications. It is also apparent that Google and Microsoft have invested heavily in resources that will provide them continuous update of their imagery. In fact, the Digital Globe /Microsoft Clear30 program is likely to become the de facto source of one foot Imagery for the Nation. It is also evident that local governments are making major investments in very high resolution (4 and 6 inch) orthrophotos and oblique imagery. Many counties are collecting these data on an annual basis to support their property assessment programs. There is an increasing trend for state level consolidation of these image resources. While the federal Imagery for the Nation program is stalled there is major support from NGA to support GeoEye and Digital Globe's continued operation which will include two new satellites in the next four years. Perhaps the major development in terms of digital imagery has been the explosion in the acquisition of airborne and mobile LiDAR. This technology is likely to change the way surveying is conducted and could be the best source for extracting structures and conducting change detection. With major efforts underway at the federal, state, local and commercial firms (e.g. NAVTEQ) LiDAR data will find a place in ever increasing application domains. While it is unlikely that the GSS Initiative will be able to access LiDAR data on a large scale, it should follow the developments very closely. Many states are likely to soon have complete coverage and the trends suggest that LiDAR could be deployed to generate complete inventories of structures and road features.

3.2 Investigate the Current State of the Art in Digital Imagery Processing Procedures for Image Classification, Automated Feature Extraction, and Change Detection to Identify, Represent, and Geographically Position New Roads and Structures.

The objective of this part of the report is to provide a review of the current state of the art in image processing tools that perform raster based classification, feature extraction, and change detection. These include advanced routines that incorporate a wide array of spectral and spatial information in object and rule based systems. It also reports on the advancements in the commercial off the shelf (COTS) software to extract and classify imagery and to detect changes between dates.

3.2.0 Harris Report

The 2008 report prepared by Harris Corporation provided some useful guidance regarding the implementation of image based analysis. Some of the issues they raised include:

- What source data types have been selected for use?
- Is the intent to support a visual review or feature extraction?
- What level of consulting support would the vendor provide?

- How flexible are the tool licensing requirements?
- How complex are the requirements for an imagery library that the tool supports?
- What level of trust will there be in the output (e.g., false positives, omissions) of an image analysis tool? (e.g. will a visual review always be required regardless of the process selected?)
- Can the tool ingest overlay information from the MAF to support the correlation of image features to addresses in the MAF?

3.2.1. Image Classification to meet GSS Initiative Requirements

In a white paper prepared for the 2010 workshop on *New Research Directions for the National Geospatial-Intelligence Agency,* Congalton (2010) provides a timely review of the current state of the art in digital remote sensing. The following discussion is based largely on his review. Digital image analysis is performed through a series of steps.

Pre-processing is defined as any technique performed on the image prior to the classification. There are many possible pre-processing techniques. However, the two most common techniques are geometric registration and radiometric/atmospheric correction. Geometric registration is performed for three possible purposes:

(1) to match imagery from one date to another,
(2) to match the imagery to another spatial data layer, or
(3) to match the imagery to the ground.

The overwhelming advances in GIS have made the use of geometric registration an integral part of digital remote sensing. The process of geometric registration requires resampling of the imagery to fit the image to the desired source (i.e. other image, spatial data layer, or ground).

The classification of digital data historically has been limited to spectral information (tone/color). While these methods attempted to build on the interpretation methods developed in analog remote sensing, the use of the other elements of photo interpretation beyond just tone/color has been problematic. The digital imagery is simply a collection of numbers that represent the spectral response for a given pixel for a given wavelength of electromagnetic energy (i.e. band of imagery). In other words, the numbers represent the “color” of the objects.

Unsupervised classification uses a statistical clustering algorithm to group the pixels in the imagery into spectral clusters. These clusters are spectrally unique, but may not be informationally unique. In other words, a single cluster may be a combination of a number of informational types (e.g., cluster 4 may be a combination of white pine and grass). The major advantage of the unsupervised classification technique is that all the spectral variation in the image is captured and used to group the imagery into clusters. The major disadvantage is that it is difficult to informationally label all the clusters to produce the thematic map.

Supervised classification is a process that mimics photo interpretation. The analyst “trains” the computer to recognize informational types such as land cover or vegetation in a similar way that the photo interpreter trains themselves to do the same thing.

Congalton also makes the following interesting assessment of image classification procedures

> "It should be noted that Anderson et al. (1976) in a publication for the USGS stated that thematic maps generated from digital remotely sensed imagery should be at least 85 percent accurate. This value has been used as a guideline ever since. However, the overwhelmingly vast majority of thematic maps generated from remotely sensed data created from 1976 to the present have failed to meet this accuracy level."

He also highlights three new methods of image classification that have demonstrated improvements over the conventional statistical routines. These include the following:

1. Classification and Regression Tree (CART) analysis. According to Congalton (2010):

> "CART analysis is a non-parametric algorithm that uses a set of training data to develop a hierarchical decision tree. This decision tree is created using a binary partitioning algorithm that selects the best variable to split the data into separate categories at each level of the hierarchy. Once the final tree is generated, it can then be used to label all the unknown pixels in the image. CART analysis has become widely used in the last few years both for pixel-based and object-based image classification. The method is extremely robust and provides significantly better map accuracies than have been achieved using the more basic approaches. "

A recent example of CART analysis is provided by Cheng et al., 2009. Using ArcGIS and Visual Basic they report an accuracy level of "near 90%" for soil Zn content in a study in China.

2. Artificial Neural Networks (ANN) According to Congalton (2010):

> "The power of the ANN is that it has the ability to "learn". While this approach seems to have great potential, the method is subject to over-fitting the data and is unnecessary complex. Therefore, while much work is still being done on this technique, it has not experienced the same widespread adoption as CART. "

Kavzoglu, and Reis (2008) provide a recent example of ANN classification (Figure 41) of Landsat imagery. They reported that ANN performed better than a maximum likelihood classifier.

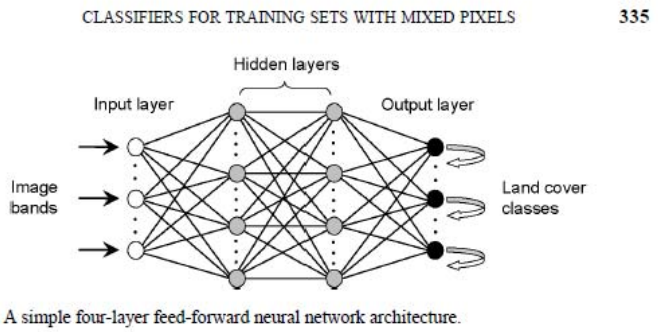


Figure 41 Example of the neural network architecture (Kavzoglu and Reis 2008)

3. Support Vector Machines (SVM) According to Congalton (2010):

> "SVM are derived from the field of statistical learning theory and have been used in the machine vision field for the last 10 years. More recently, these methods have been developed for use in creating thematic maps from remotely sensed imagery. SVM act by projecting the training data nonlinearly into feature space of a higher dimension than that of the input data. This projection is accomplished using a kernel function and results in a data set that now can be linearly separated. The ability to separate out the various informational classes in the imagery is a powerful advantage. Like ANN, SVM is subject to over-fitting. However, a technique that is part of the analysis works to minimize this issue. The use of SVM is relatively new and offers great potential for creating thematic maps from digital imagery."

Mirnalinee (2007) provides a particularly useful example of SVM based classification. In a study area in Chile road extraction was performed in two steps. In the first step, SVM was employed to classify the image into two categories: road and non-road. In the second step shape descriptor was used to extract roads from other spectrally similar objects. Using this approach completeness measure of 94% and a correctness measure of 92% was obtained with few false positives (Figure 42).

Figure 1. Experimental results: (a) 1m pan Ikonos image of Valparaiso, Chile; (b)Output of SVM classification; (c)Detected road segments; (d) Detected road segments overlaid on input image ; (e)Map of corresponding area; (f) Manual extraction

Figure 42 Output of the SVM procedures to identify roads (Mimalinee, 2007)

3.2.2 Feature Extraction

Once an image has been classified it is usually desirable to extract specific features (urban, impervious surfaces, roads, structures, etc.) into a GIS compatible format. In current computing environments, classified or labeled pixels are identical to a raster GIS data base that can be easily converted (extracted) to a polygon or vector GIS (Figure 43). A popular approach to feature extract is to utilize Overwatch's Feature Analyst that can be deployed as an ArcGIS extension (see detailed discussion later in this report).

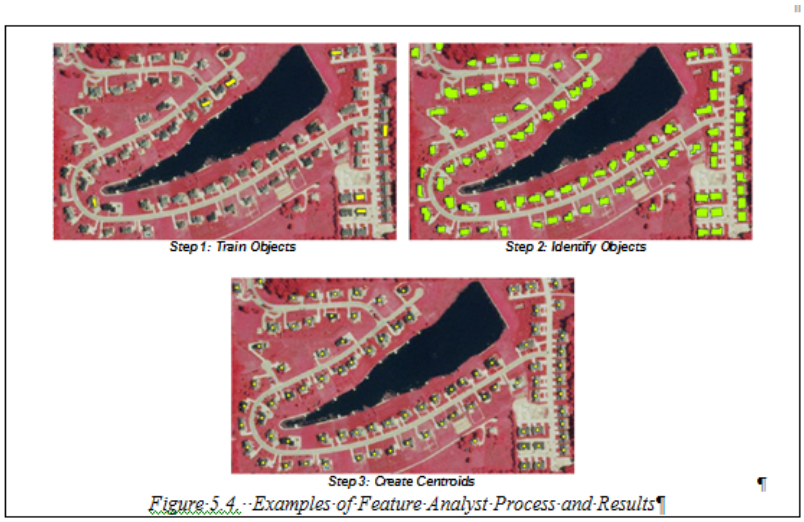


Figure 43 Example of Feature Analyst process (Harris, 2008)

A useful approach feature extraction is to incorporate rules about the geometry of the clusters of pixels with similar spectral characteristics (asphalt roads and roofs). These rules can be simple ones based on minimum size, eliminating gaps in linear features, or attempting to create rectangles. For example, it appears that the ratio of the square of the perimeter to the area can differentiate new buildings from new roads which have a similar spectral reflectance (Figure 44).

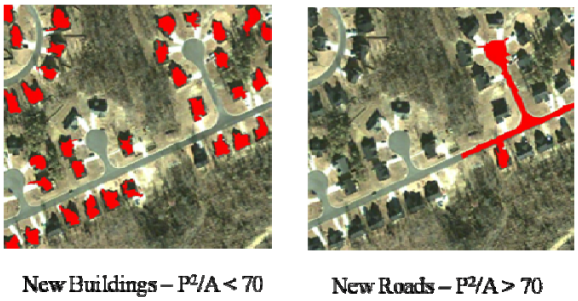


Figure 44 Example of rule based image classification (Fitzgerald, 2003)

3.2.2.1 Feature Extraction Research

The following section highlights a selection of research projects that have focused on issues relating to the extraction of features from imagery.

Hodgson et al., (2003) used an extensive set of rules to separate impervious land parcels (Figures 45 and 46).

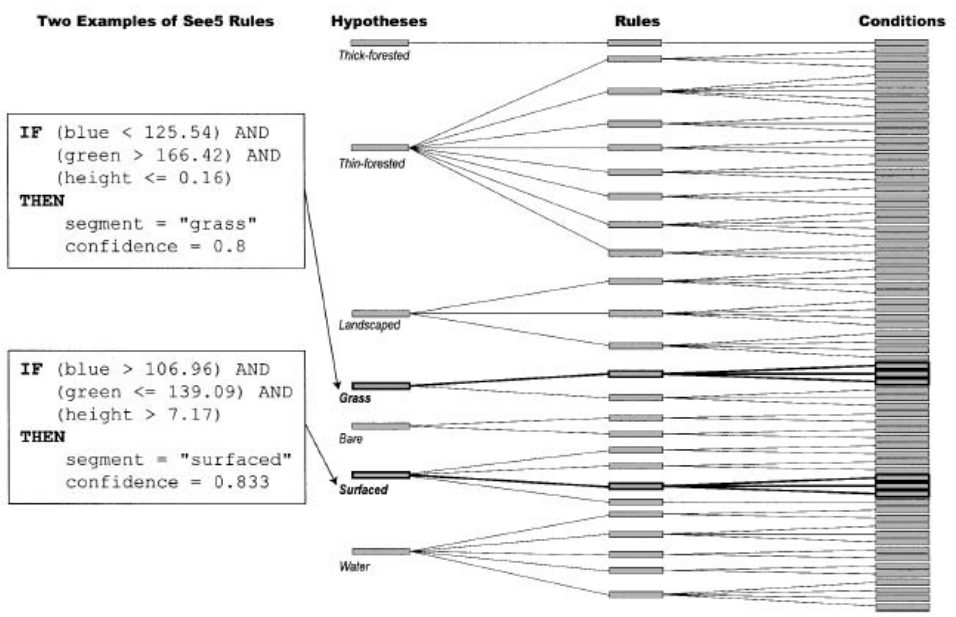


Figure 1. Two examples of the 25 See5 rules used to classify image *segments* into the seven imperviousness categories. Multiple conditions within a rule are separated by "AND" logic. Any rule evaluating to *true* causes the hypothesis to be *true*. If no hypotheses evaluate to *true*, the observation is classified as the category with the highest number of training samples.

Figure 45 Example of the See5 rule to extract impervious surfaces (Hodgson et al. 2003)

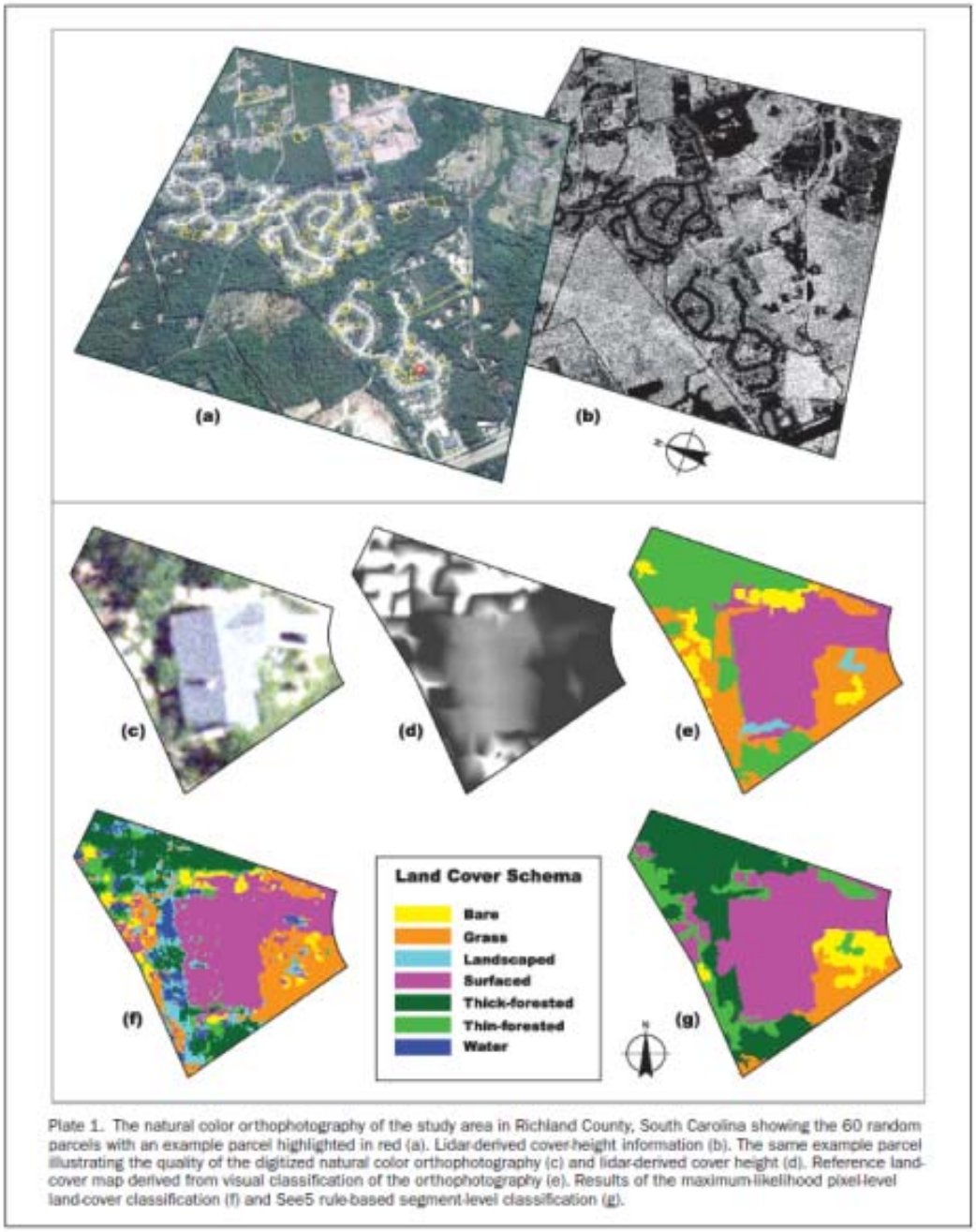


Plate 1. The natural color orthophotography of the study area in Richland County, South Carolina showing the 60 random parcels with an example parcel highlighted in red (a). Lidar-derived cover-height information (b). The same example parcel illustrating the quality of the digitized natural color orthophotography (c) and lidar-derived cover height (d). Reference land-cover map derived from visual classification of the orthophotography (e). Results of the maximum-likelihood pixel-level land-cover classification (f) and See5 rule-based segment-level classification (g).

Figure 46 Resultant land cover scheme from rule based classification (Hodgson et al. 2003)

Hodgson et al., (2004) also provide a general flow model for extraction of transportation features that has direct bearing on the GSS Initiative (Figures 47 and 48).

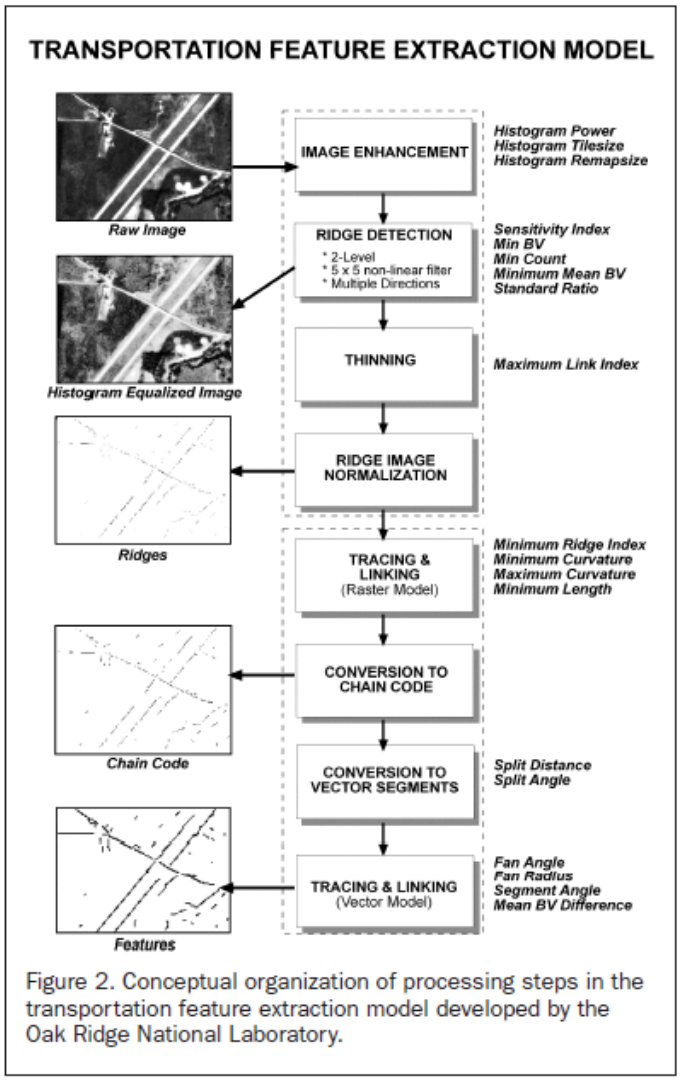


Figure 47 Transportation feature extraction model (Hodgson et al. 2004)

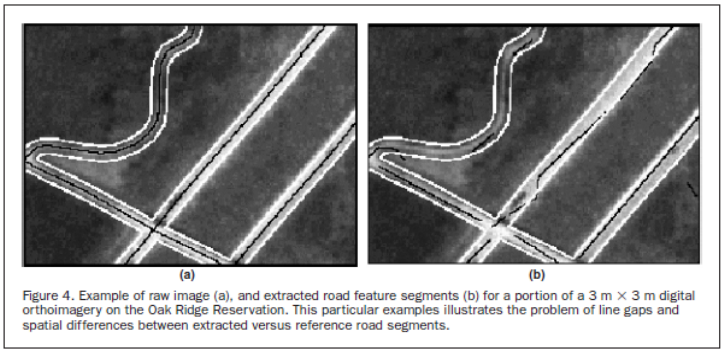


Figure 48Output of transportation feature extraction system (Hodgson et al 2004)

The University of Southern California Information Science Institutes' Information Integration Research Group under the direction of Craig Knoblock is probably at the forefront of extracting features from raster versions of maps. They have generated a series of publications and

presentations that document their developments. Their research has been supported by NSF and the Air Force. They have also collaborated with the USGS Center of Excellence for Geospatial Information Science (CEGIS). A good summary of this work is available in a recent dissertation Chiang (2010) *Harvesting Geographic Features From Heterogeneous Raster Maps.* His summary of the contribution of the research is as follows:

"The key contribution of this thesis is a method for separating and recognizing geographic features from raster maps and aligning the raster maps, separated layers, and recognized features to other geospatial data. This thesis offers several contributions:

- An automatic approach that separates road and text layers from raster maps
- An automatic approach that extracts road-intersection templates from road layers for the alignment of raster maps and other geospatial data
- A general approach for extracting and vectorizing road geometry from raster maps which includes:
    - A supervised approach that handles complex raster maps or maps with poor image quality for separating road layers from the maps
- An automatic approach for extracting road vector data from road layers
- A general approach for recognizing text labels in raster maps which includes:
    - A supervised approach that handles complex raster maps or maps with poor image quality for separating text layers from the maps
    - An automatic approach for recognizing text labels in text layers"

He also provides a flow chart of the steps in the process (Figure 49).

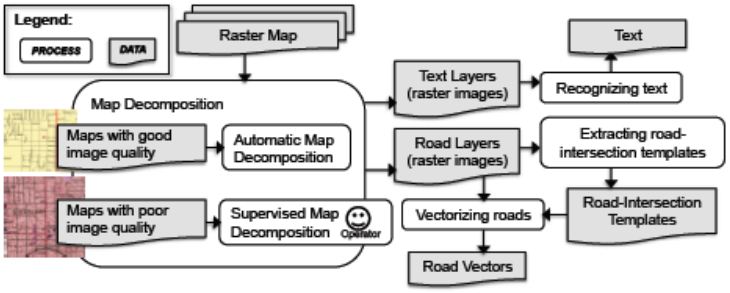


Figure 1.3: The overall approach to harvesting geographic features from heterogeneous raster maps

Figure 49 Flow chart of feature extraction from maps (Chiang 2010)

Over the past few years Li and Briggs have developed a number of significant tools for automated extraction and georeferencing of road features. Li and Briggs (2006) describe a procedure for automated georeferencing of a raster image to a vector road network. This utilizes a topological point pattern to link image and vector data and to identify control point pairs between an image and a vector road network (Figure 50).

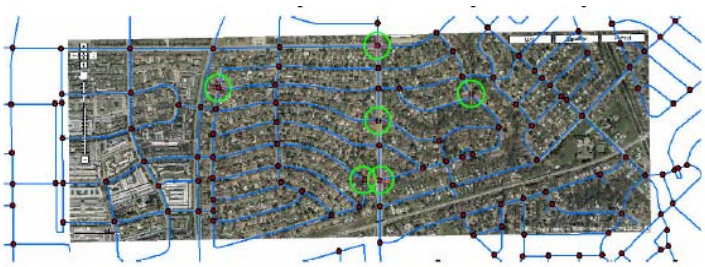


Fig. 7: Georeferencing result for 7 image point tests. Total RMS error is 7.064. 1 false point is filtered out.

Figure 50 Automated georeferencing of image to road network (Li and Briggs, 2006)

Li (2009) and Li and Briggs R. (2009) describe how they employ a *reference circle*, to properly identify the pixels that belong to the same road and use this information to recover the entire road network. They demonstrate how the system is able to extract roads reliably from images with complex content such as the high resolution aerial/satellite images available from Google maps (Figure 51).

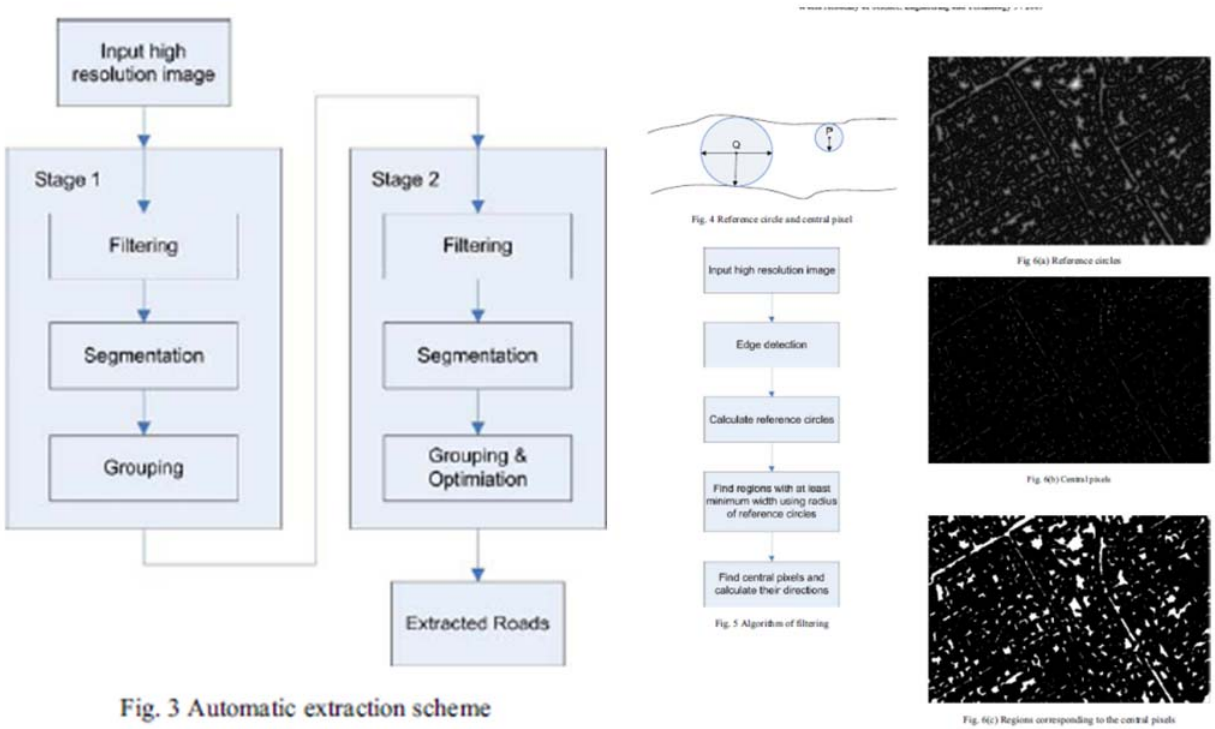


Figure 51Automated feature extraction system based on reference circles (Li, 2009)

Mayungaa, et al., (2005) describe a procedure for extracting buildings from QuickBird imagery (Figure 52). They report an extraction rate of 91 percent.

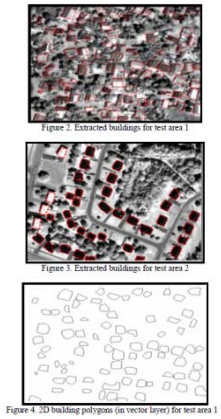


Figure 52 Building polygons extracted from QuickBird imagery (Mayungaa 2005)

3.2.2.1 Object Based Image Analysis

Object based image analysis represents the current state of the art in digital image processing. According to Congalton (2010):

> "By far the greatest advance in classifying digital remotely sensed data in this century has been the widespread development and adoption of object-based image analysis (OBIA). In the OBIA approach, unlabeled pixels are grouped into meaningful polygons that are then classified as polygons rather than individual pixels. This method increases the number of attributes such as polygon shape, texture, perimeter to area ratio and many others that can be used to more accurately classify that grouping of pixels. Polygons are created from pixels in OBIA using a method called segmentation. There are a number of current image analysis software packages that provide the means of performing object-based image analysis. In all these algorithms, the analyst must select of series of parameters that dictate how the segments or polygons are generated. Depending on the parameters selected, it is possible to create large polygons that may incorporate very general vegetation/land cover types or very small polygons that may divide even a specific cover type into multiple polygons. The power of the segmentation process is two-fold. First, the imagery is now divided into polygons that can, in many ways, mimic the polygons that may have been drawn by an analyst that was manually interpreting this same image. In this way, some of the additional elements of manual interpretation mentioned earlier in this paper become relevant for digital image analysis. Secondly, as previously mentioned, the creation of polygons results in a powerful addition of attributes about the polygons that can be used by the classification algorithm to label the polygons. Both these factors significantly add to our ability to create accurate thematic maps".

Im, et al. (2008) provide an example of object based image analysis that incorporated eight LiDAR derived metrics (mean height, standard deviation (STDEV) of height, height homogeneity, height contrast, height entropy, height correlation, mean intensity, and compactness) to separate grass, roads/parking lots, and buildings with greater than 90% accuracy (Figures 53 and 54) .

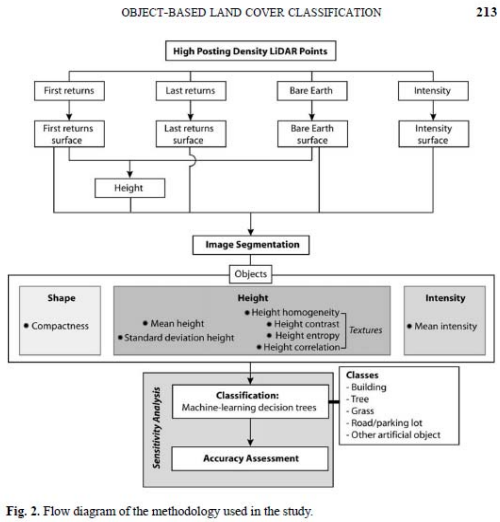


Fig. 2. Flow diagram of the methodology used in the study.

Figure 53 LiDAR based object classification (Im el at. 2008)

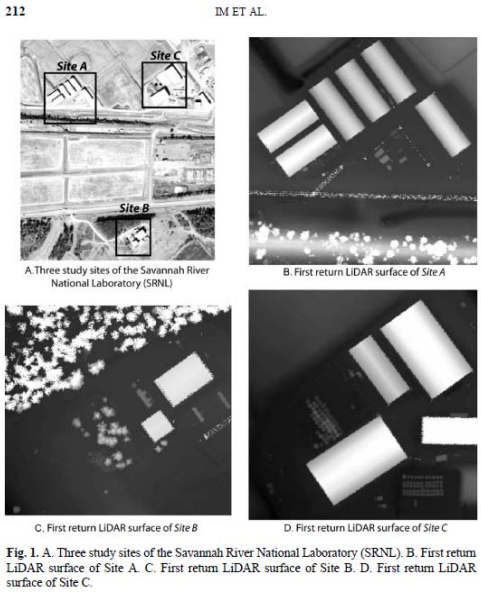


Fig. 1. A. Three study sites of the Savannah River National Laboratory (SRNL). B. First return LiDAR surface of Site A. C. First return LiDAR surface of Site B. D. First return LiDAR surface of Site C.

Figure 54 Output of LiDAR based object classification (Im et al. 2008)

Further discussions of OBIA is included as part of the discussion of eCognition software later in the report.

3.2.2.2 Image Processing Trends

Congalton (2010) provides a valuable assessment of current trends in digital image processing. These include:

> "(1) Hyperspectral imagery has changed the way we perform digital image analysis. Given this imagery collected over narrow bandwidths across a large portion of the electromagnetic spectrum, it is possible to create spectral libraries of various information

types and compare these for identification on the imagery. These libraries exist for a variety of rock and mineral types and have even been created for some simple land cover/vegetation classifications. Currently, significant research is occurring in this field. As the costs associated with this technology continue to decline, more and more uses of hyperspectral imagery will be developed.

(2) Digital cameras are revolutionizing the collection of digital imagery from airborne platforms. These cameras sense electromagnetic energy using either a charged-coupled device (CCD) or a Complementary Metal Oxide Semiconductor (CMOS) computer chip and record the spectral reflectance from an object with greater sensitivity than the old analog film cameras. While more expensive than analog film cameras, new digital camera systems, especially large format systems, are rapidly replacing traditional analog systems.

(3) Large Federal Programs, such as the USDA National Agriculture Imagery Program (NAIP), are providing a tremendous source of digital imagery that the analyst can readily digitally analyze. Most digital camera imagery is collected as a natural color image (blue, green, and red) or as a color infrared image (green, red, and near infrared). Recently, more projects are acquiring four wavelengths of imagery (blue, green, red, and near infrared). The spatial resolution of digital camera imagery is very high with 1 – 2 meter pixels being very common and some imagery having pixels as small as 15 cm.

(4) Since the turn of the century, the general public has become increasingly aware of the use of remotely sensed imagery and geospatial technologies. Events such as the World Trade Center attack on September 11, 2001 and various natural disasters in New Orleans and elsewhere have made these technologies commonplace on the nightly news and the Internet. Applications such as Google Earth, Bing Maps, and others have given everyone the opportunity to quickly and easily use remotely sensed imagery anywhere in the world. This increased awareness and widespread understanding of these technologies is very positive".

A paper by Heipke, et al. (2008) also provides some interesting observations about trends in a paper entitled *Updating geospatial databases from images.* They draw the following conclusion*:*

"The first and perhaps most important observation with respect to system architectures is a major trend to database-centric architectures. This centralizing tendency, together with the implications of more complex data models, has profound implications for update (Woodsford, 2004). Update processes need to be 'data model-aware' and to be tightly coupled with validation services, to avoid costly and lengthy error detection-correction cycles. Whilst it is possible to replicate such validation services in each update client (field, photogrammetric systems, imagery change detection systems) as models become more complex, and business rules come to the fore, such an approach becomes more costly and difficult to sustain." (p. 357)

They highlight the use of Radius Studio from 1Spatial as a useful rule based software solution that can also be deployed using a SOAP (Simple Object Access Protocol) Web Service that provides an interface for the purposes of validating features remotely (Figure 55).

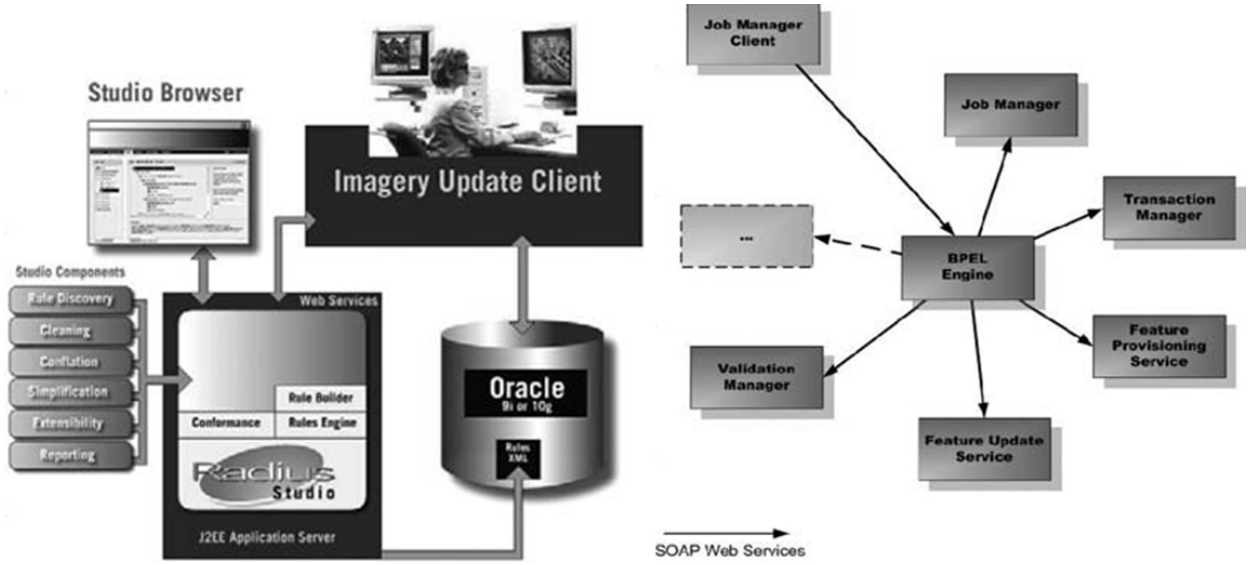


Figure 55 Radius Studio architecture (Heipke et al. 2008)

3.2.3 A Review of Change Detection Procedures

It can be argued that the focus of GSS Initiative is the detection of changes in roads and structure points. Of course, change detection is a broad topic with important applications in medical imaging, military surveillance and image compaction. In fact Rindfuss et al., (2004) suggest that a science of land change is evolving. They suggest (p. 13976) that:

> "Land-change science has emerged as a foundational element of global environment change and sustainability science. It seeks to understand the human and environment dynamics that give rise to changed land uses and covers, not only in terms of their type and magnitude but their location as well. This focus requires the integration of social, natural, and geographical information sciences. Each of these broad research communities has developed different ways to enter the land-change problem, each with different means of treating the locational specificity of the critical variables, such as linking the land manager to the parcel being managed. The resulting integration encounters various data, methodological, and analytical problems, especially those concerning aggregation and inference, land-use pixel links, data and measurement, and remote sensing analysis."

They identify the following issue (p.13980):

> "Accuracy Assessment of Land-Change Models. A major thrust of most household–pixel linkages is to generate data that inform and improve models of land change, which range from explanatory to integrated assessment models. Such efforts confront the problem that the pixel of satellite imagery is neither a landscape nor a social unit. It is convenient,

however, to treat the pixel as a unit of observation because data are often organized around it and software packages are designed to perform pixel analyses relatively easily. These image processing packages tend to employ statistical methods that

(i) treat each observational unit as something meaningful to the phenomenon,
(ii) assume each pixel is independent of other units in the image, and
(iii) assume the contents of each pixel are largely independent across space and time; put differently, the stationary or Markovian assumption is often used."

3.2.3.1 Review of European National Mapping Organizations Criteria and Findings Regarding Change Detection

In March 2010, several national mapping organizations that are part of the European Spatial Data Research Network (EuroSDR) met at the Ordnance Survey (OS) for a workshop on change detection. Holland of the OS presented a useful overview of characteristics of change detection procedures that are important to national mapping organizations. His analysis included the following issues:

Things we (national mapping organizations) have in common

- Detecting building change
- Augmenting, rather than replacing, manual processes
- Working with imagery and height data
- Large (-ish) scale map data (but different specifications)
- Mainly research projects up until now

Things we may want

- Production-quality systems
- Proven cost/time savings
- More automation
- Support from systems vendors (Leica, Zeiss, BAE Systems, Inpho, etc…apologies to those in the audience I haven't listed!)
- Not just detection – feature extraction…?

When comparing images, we need to account for:

- Differences in time of day
- Differences in time of year
- Radiometric differences
- Clouds
- Shadows
- Cloud shadows
- Climatic differences (floods one year, drought the next)
- Differences in position, flying height, camera angle
- Transient objects (vehicles, etc.)

They are also often not the changes we want:

- Parked vehicles or containers
- Vegetation differences (leaf-on/leaf-off)
- Transient features
- Artifacts of automatic DSM creation can be a problem

- Saturated pixels may cause false DSM values
- Better matching algorithms are helping to solve this

Simple pixel classification

- Classifying pixels doesn't always help us classify objects
- There is always a lot of variation in any scene
- Mixed pixels often occur
- Comparing pixel by pixel will lead to a huge over-estimate of the number of changes

Comparing vectors with vectors

- Searching for features in the image, such as:
  - Lines
  - Right angles
  - Parallel lines
  - Closed polygons
- Techniques, among many, include:
  - Edge detection (Canny, etc)
  - Line detection (Hough transform, etc.)
  - Line following (snakes)
- The detected features are compared with vectors in the database

Comparing vectors with vectors

- Matching features identified in the image with features in the database can be difficult due to:
  - Occlusions
  - Breaks
  - Displacements
- It can help if the matching is done using context information
  - two parallel lines may indicate road/railway
  - multiple right angles may indicate a building

Vectors vs. Images

- Image to image comparisons tend to highlight many changes which are not of interest for mapping
- Comparing an image to a vector map would, ideally, indicate changes to features in the map data
- Variations of this approach have been taken by Ordnance Survey, IGN France and BKG, among others.
- The approach taken generally uses some form of image classification, followed by a comparison

Commercial services and products

- Erdas Imagine® DeltaCue
  Image to image comparison. Our brief evaluation found it to be unsuccessful for building change. Probably more useful for land cover change.
- Feature Analyst®
  Can be used to find building changes. Something of a "black box" approach, requiring trial and error to find the best combination of parameters. Works best on well-defined shapes.

Examples from mapping organizations

- Ordnance Survey GB – building & rural feature detection
- BKG Germany – road verification
- IGN France – building detection
- FGI Finland – building detection

(http://www.eurosdr.net/workshops/cd_2010/p-2.pdf)

At the workshop, Butcher (2010) from the Defence Science and Technology Laboratory of the UK Ministry of Defence provided a particularly interesting discussion of its Multi-source Automated Toolkit for Improved Spatio-Spectral Exploitation (MATISSE) (Figure 56). This represents one of the first public change detection discussions from a military perspective. The system includes pixel matching, feature matching and object matching procedures based on Fourier space comparisons using spatial and spectral information, edge based matching, and entropy comparisons.

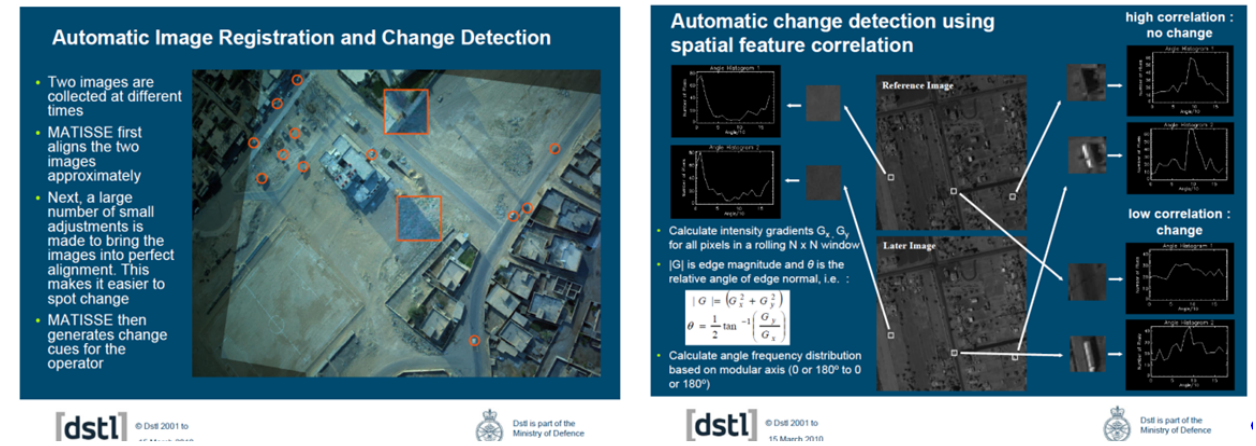


Figure 56 Description of the MATISE system (Butcher 2010)

Domenech, et al., (2010) provided a detailed description of their approach to change detection in Spain. This system uses eCognition to incorporate high resolution imagery, cartographic objects such as parcels as well as LiDAR. Some of the steps in their process are included in Figure 57.

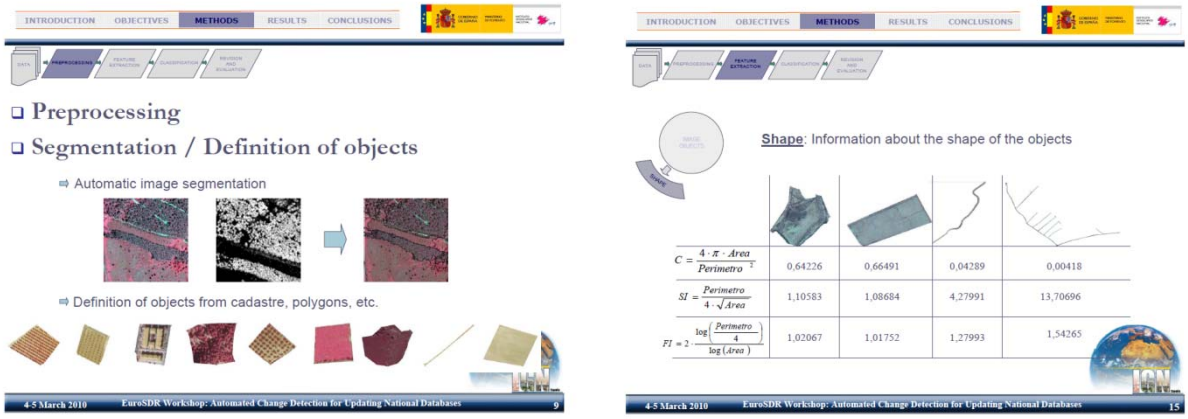


Figure 57 Steps in Spanish change detection process Domenech, et al., (2010)

3.2.3.1.1 Change Detection Research at the Ordnance Survey

The efforts of the OS to conduct change analysis have been presented in a research paper by Holland, et al., (2008). They have discovered that:

> "This object-based method offers some advantages over the traditional pixel-based approach, in that, in addition to the spectral data, it can incorporate shape, texture and local context into the classification process. Also, by segmenting the image into homogeneous objects prior to classification, this approach helps to reduce the spectral variability within each class."

They also conclude that:

> "the tool was well received and was shown to significantly reduce the amount of time taken to identify the changes over a 48 km2 area. At the time of writing, a large production trial is planned, which will use recently-captured imagery in an area which has changed significantly since the previous update to the topographic data. This will allow us to test the method in another area, and to determine its effectiveness and efficiency in an area with a large number of changes."

They also provide a useful table of error rates (Figure 58).

	Demol	New	Total
Actual Changes	34	17	51
Objects classified as changes	79	96	175
Actual changes correctly classified (True Positives)	33	16***	49
Non-changed objects classified as change (False Positives)	46	84	130
Actual changes not classified as changes (False Negatives)	1	1	2
% Classified as changes that were actual changes	42%	17%	28%
% of actual changes classified as changes	97%	94%	96%

Table 3. Results of change detection from object-based classification for the 2 km^2 Heathrow test site.

Figure 58 Results in change detection process (Holland et al. 2008)

3.2.3.1.2 European Experiments with Building Level Change Detection

In addition to the group of researchers who attended the EuroSDR meeting in March a group of researchers in France, Germany, Finland, and Denmark have been conducting a series of experiments that evaluate change detection methodologies and data sources to analyze individual buildings. In a series of papers (Champion 2007, Heipke et al., 2008, and Champion et al., 2009, Alobeid et al., 2010) these groups have discussed and compared the results of a series of

experiments. The results of these experiments have direct bearing on the GSS Initiative and should be followed closely. The summary article (Champion et al., 2009) provides an evaluation of different methods. It also provides a valuable discussion of error assessment based on completeness and correctness (Figures 59 and 60) and draws some important conclusions:

- Satellite only context appears to be the most challenging for the current state-of-the-art
- The aerial context and the LIDAR context appear to be a viable basis for building an operative system in the future.
- Thus, the high completeness rates for *demolished* buildings and the high correctness for *unchanged* buildings that could be achieved in these contexts highlight the effectiveness of the presented approaches in verifying the existing objects in the databases.
- The main limitation in terms of qualitative efficiency concerns the relatively high number of *FN new* buildings – up to 12.1% in the Marseille test area with (Rottensteiner, 2008) – that are mostly related to the object change size.
- The economical efficiency of the presented approaches seems to be promising, with 80-90% of the existing buildings requiring no further attention by the operator.
- These buildings are reported to be *unchanged*, which saves a considerable amount of manual work.
- In terms of the economical efficiency, the main limitation is a high number of *FP demolished* buildings that have to be inspected unnecessarily.
- Again, this is mainly caused by problems in detecting small changes. Areas of improvement should concern input data and methodologies.
- Thus, the resolution of LIDAR data (1 point / m^2) used in this test appeared to be critical for the change detection performance: using higher density LIDAR data (e.g. 5-10 points / m^2) should improve the situation.
- As far as methodology is concerned, new primitives should be used in the algorithms, in particular 3D primitives (representing e.g. the 3D roof planes or building outlines) that can now be reliably reconstructed with the 3D acquisition capabilities, offered by recent airborne/spaceborne sensors. Another concern should be the improvement of the scene models used in object detection such that they can deal with different object classes and their mutual interactions.
- By incorporating different object classes and considering context in the extraction process, several object classes could be detected simultaneously, and the extraction accuracy of all interacting objects could be improved.

Approach	Completeness	Correctness
Marseille (Imagery – Aerial context)		
(Champion, 2007)	94.1%	45.1%
(Matikainen et al., 2007)	**98.8%**	54.3%
(Rottensteiner, 2008)	95.1%	**59.1%**
Toulouse (Imagery – Satellite context)		
(Champion, 2007)	78.9%	**54.5%**
(Rottensteiner, 2008)	**84.2%**	47.1%
Lyngby (LIDAR context)		
(Matikainen et al., 2007)	94.3%	48.8%
(Olsen and Knudsen, 2005)	**95.7%**	53.6%
(Rottensteiner, 2008)	91.4%	**76.1%**

Table 1. Completeness and Correctness achieved by the four algorithms for the three datasets.

Figure 59 Completeness and correctness assessment (Champion et al. 2009)

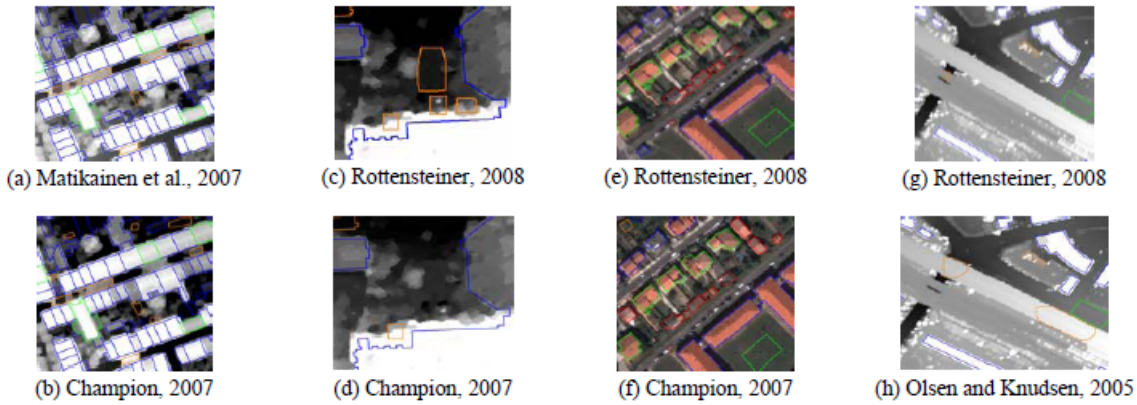


Figure 3. Evaluation Details (same colour code as Figure 1). *FP new* cases related to DSM errors (shadow areas), in Marseille streets (a)-(b) and Toulouse (c)-(d); (e)-(f) *FN new* cases (small changes); (g)-(h) *FP* new buildings related to bridges.

Figure 60 Sample data sets used in the evaluation of change detection (Champion et al. 2009)

3.2.3.2 Binary Change Detection Research

The first step in any change detection processes is to discover where change has occurred. In effect, a comparison of two data sets triggers an alarm that signals that something is different. This type of change may be considered to be a binary (change / no-change) operation. Given the proper resources, the location of these changes may be further investigated. The type of response depends on whether one is in a battlefield, a hospital, or updating map features. The work of a group of researchers at the University of South Carolina provides a useful approach to binary change detection (Im, et al., 2007 and 2008). They use a threshold-based calibration approach that was implemented as a dynamic linked library in ESRI ArcMap 9.1 using Visual Basic

(Figures 61 and 62). In 2008, they described an algorithm based on “systematic searching” that reduced the processing time required to identify the optimum binary discriminate function without decreasing accuracy (Figures 63 and 64).

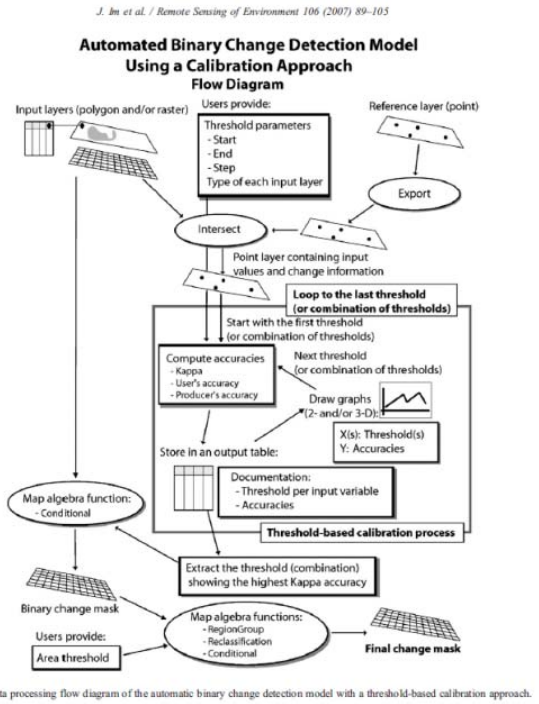


Fig. 2. Data processing flow diagram of the automatic binary change detection model with a threshold-based calibration approach.

Figure 61Automated binary change detection (Im et al. 2007)

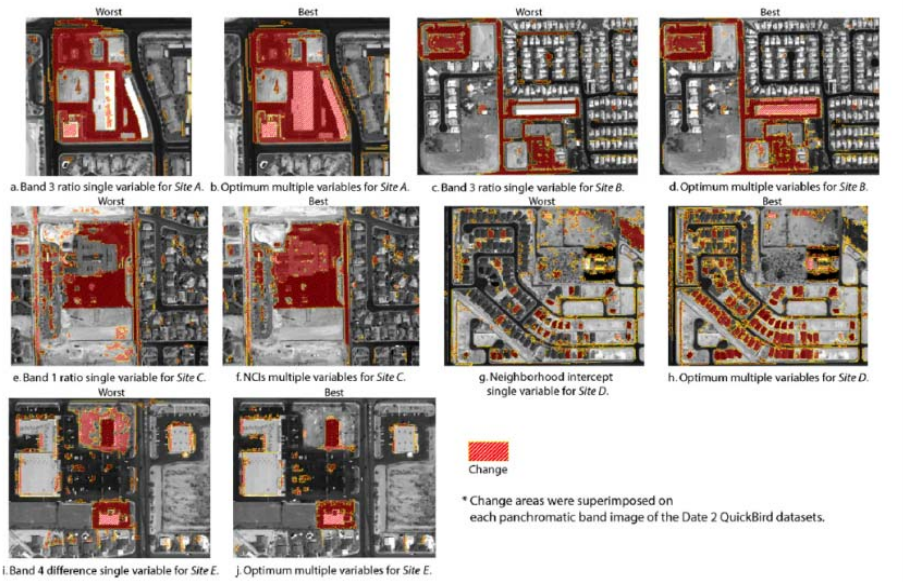


Fig. 9. Ten selected binary change results after applying the area threshold of 100 cells to remove "salt-and-pepper" noise. Change areas in a), c), e), g), and i) were extracted from the variable(s) yielding the lowest (worst) Kappa accuracy for each site, respectively. Change areas in b), d), f), h), and j) were extracted from the variable(s) yielding the highest (best) Kappa accuracy for each site, respectively.

Figure 62 Results of binary change detection process (Im et al. 2007)

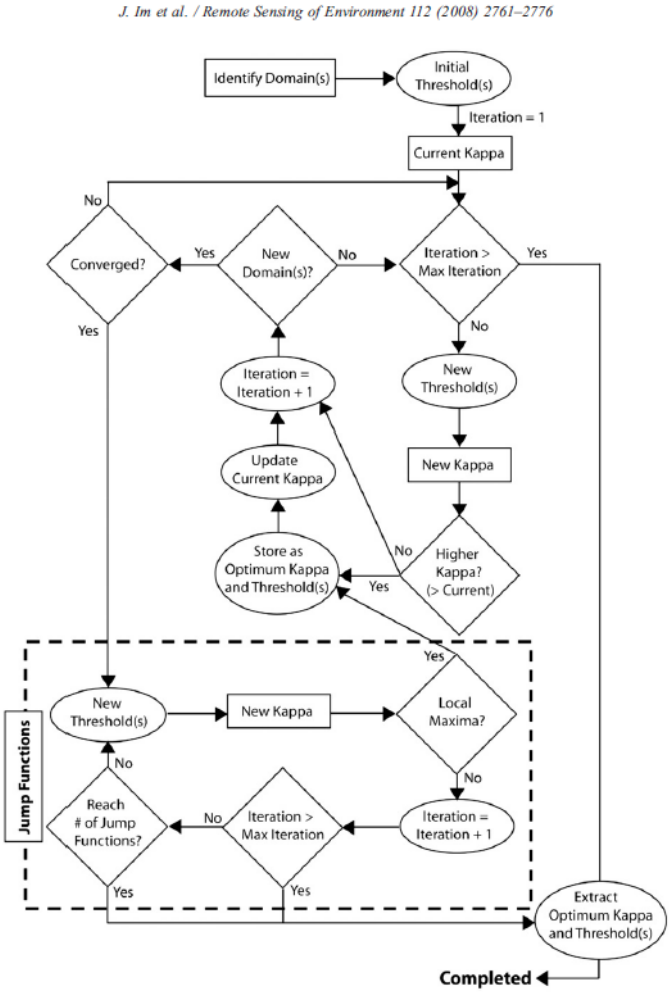


Fig. 5. Process flow diagram of the SS algorithm proposed in the study.

Figure 63 Flow chart for calibrating parameters for binary change detection (Im et al. 2008)

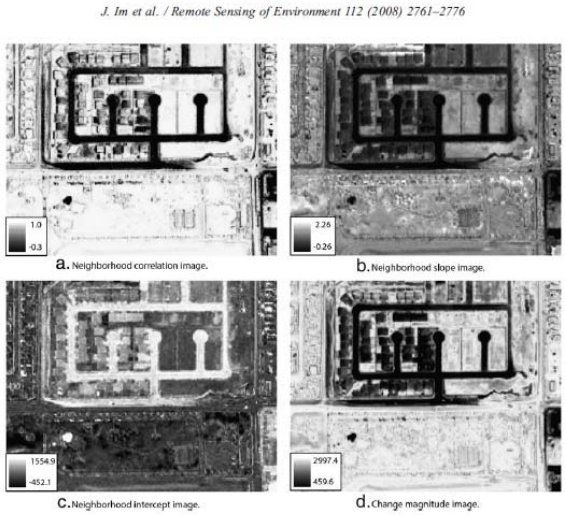


Fig. 7. Four change-enhanced images used to evaluate the optimizing algorithms (i.e., NC, NS, NI, and CVM) for *Site A*.

Figure 64 Results of binary change detection process (Im et al. 2008)

3.2.3.3 Object Based Change Detection

Almost all of the current research, including that being conducted by the Europeans, falls into the category of object based change detection. Some of the cutting edge research in this field is being developed by Jensen and his students at the University of South Carolina. Im, et al., (2008) describe an image analysis process that uses the correlation of brightness values from the same geographic area to identify change. The process is outlined in Figure 65 with example results in Figure 66.

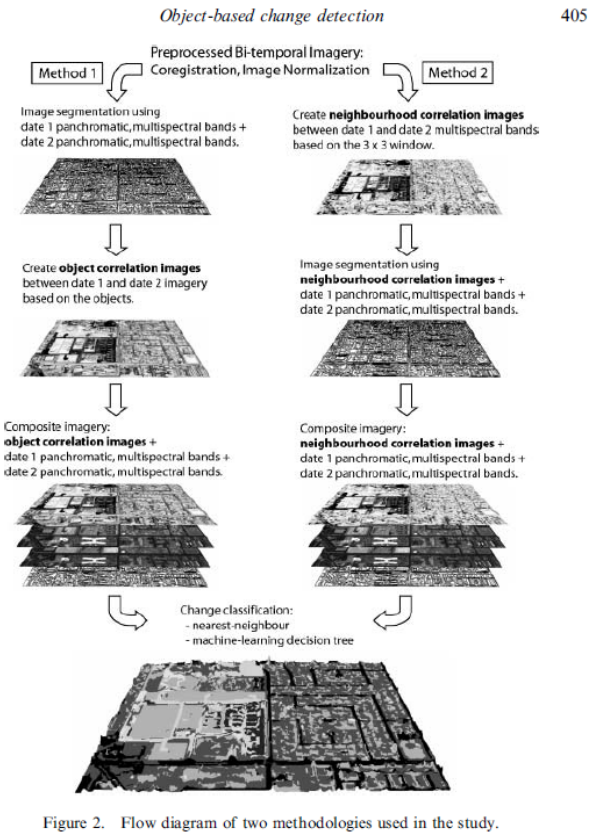


Figure 65 Object based change detection (Im et al. 2008)

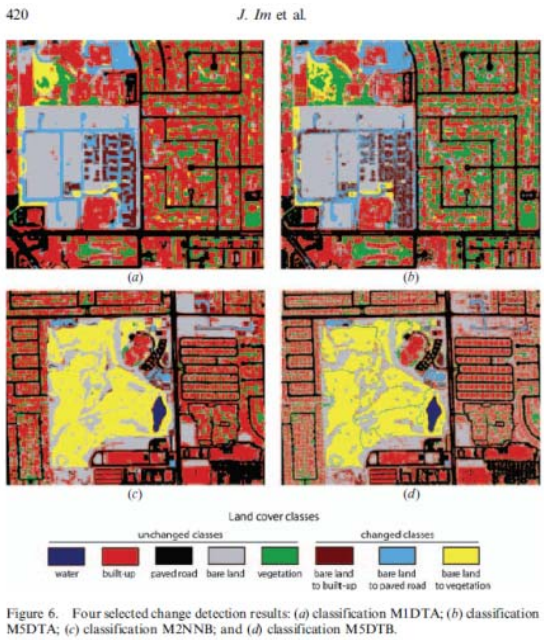


Figure 66 Results of object based change detection (Im et al. 2008)

Butkiewicz, et al., (2008) have utilized raw, irregularly spaced LIDAR points, to extract relevant changes as individual 3D models. A three-tiered level-of-detail system maintains a scale-appropriate, legible visual representation across the entire range of viewing scales, from individual changes such as buildings and trees, to groups of changes, such as new residential developments (Figures 67 and 68).

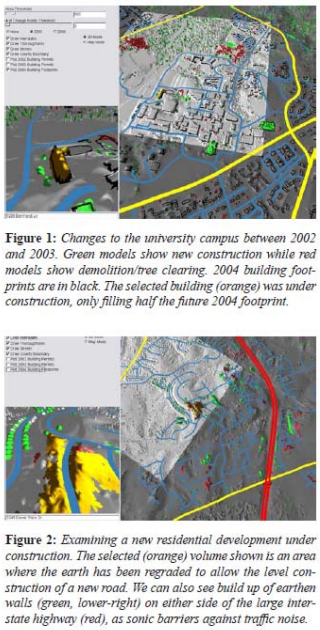


Figure 67 Illustration of LiDAR based change detection (Butkiewicz et al. 2008)

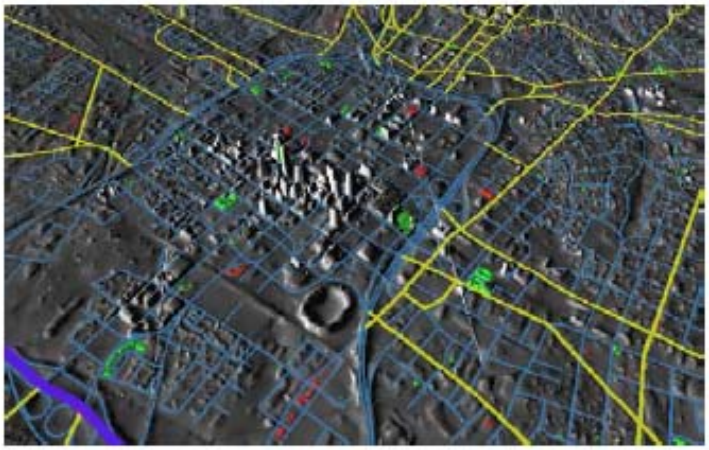


Figure 3: *A view of the downtown area in hybrid mode. Change models of new construction are rendered in green and change models of buildings that have been demolished are rendered in red.*

Figure 68 Map of changes in downtown Charlotte NC (Butkiewicz et al 2008)

3.2.4 A Review of Commercial Software Solutions

This section provides a review of commercial image and LiDAR processing software that has direct applicability to the GSS Initiative

3.2.4.1 Image Classification within ArcGIS

Over the past two decades there has been a major trend toward the integration of GIS and digital image processing. What originally were separate domains have now evolved into software tool kits with considerable overlap. From a data base viewpoint there is little difference between imagery, elevation models and any GIS grid theme. They are all forms of a georeferenced matrix of numbers. This trend is best exemplified by the incorporation of a large number of image processing tools in Version 10 of ArcGIS listed under the Spatial Analyst extension simply as "Multivariate Analysis". Within the GIS environment the raster data from digital image processing systems can be converted to vector data formats to enhance geographical analysis of these data.

ArcGIS 10 Image processing Tools (Esri)

- Band Collection Statistics - Calculates the statistics for a set of raster bands.
- Class Probability - Creates a multiband raster of probability bands, with one band being created for each class represented in the input signature file.
- Create Signatures - Creates an ASCII signature file of classes defined by input sample data and a set of raster bands.

- Dendrogram - Constructs a tree diagram (dendrogram) showing attribute distances between sequentially merged classes in a signature file.
- Edit Signatures - Edits and updates a signature file by merging, renumbering, and deleting class signatures.
- Iso Cluster - Uses an isodata clustering algorithm to determine the characteristics of the natural groupings of cells in multidimensional attribute space and stores the results in an output ASCII signature file.
- Iso Cluster Unsupervised Classification - Performs unsupervised classification on a series of input raster bands using the Iso Cluster and Maximum Likelihood Classification tools.
- Maximum Likelihood Classification - Performs a maximum likelihood classification on a set of raster bands and creates a classified raster as output.
- Principal Components - Performs Principal Component Analysis (PCA) on a set of raster bands and generates a single multiband raster as output.

ArcGIS also has a fairly robust set of tools for handling huge LiDAR datasets

3.2.4.2 Commercial Image Processing Systems

There are a number of specialized commercial software tools that facilitate automated change detection. It should be noted that in a GIS environment change is calculated as either a difference in two raster data bases or the intersection of polygons. The challenge is to develop two comparable data sets for different time periods. While the mechanics of change detection can be performed in any number of GIS environments there are a number of specialized image processing tools designed to assist the analyst. According to Congalton (2010) eCognition, ENVI, and Erdas Imagine are important commercial software solutions for object based classification. Butcher (2010) from the UK Ministry of Defence provided a useful hierarchy of the commercial technology (Figure 69). This graphic places several software solutions on continuums based on sensitivity and specitivity. For example, the ERDAS Delta Cue is high on sensitivity while eCognition is high on specitivity.

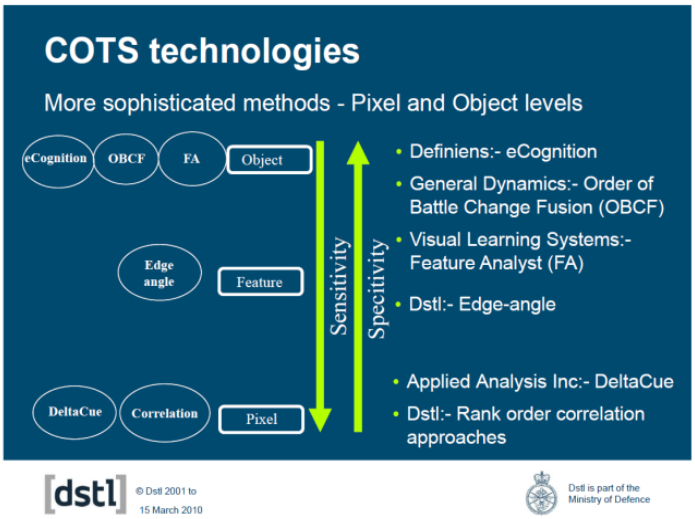


Figure 69 Classification of COTS change detection software (Butcher, 2010)

3.2.4.2.1 Overwatch Feature Analyst, LiDAR Analyst and Building Collection Toolkit
www.overwatch.com

Feature Analyst is a popular extension to ArcGIS that extracts and attributes features from satellite and aerial imagery. That is used by several federal agencies. It is based on machine learning, hierarchical learning, post processing, and automated feature extractions. It is used to automatically classify and extract features (polygons) based on color, size, shape, and texture. It can handle an unlimited number of hyperspectral bands. The software supports multiclass feature extraction and change detection (Figures 70 and 71).

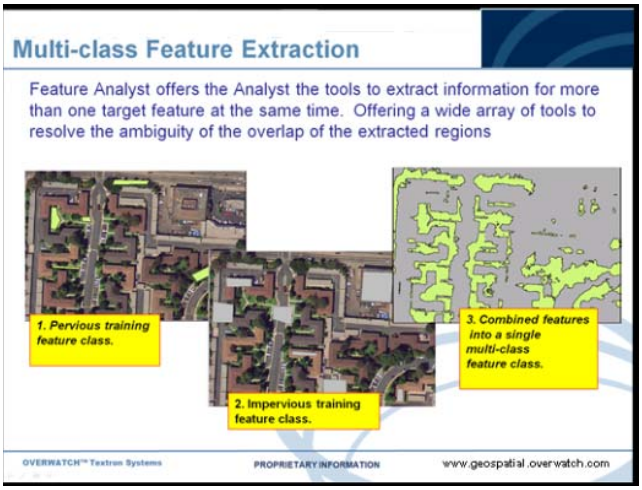


Figure 70 Feature Analyst multiclass feature extraction (Overwatch.com)

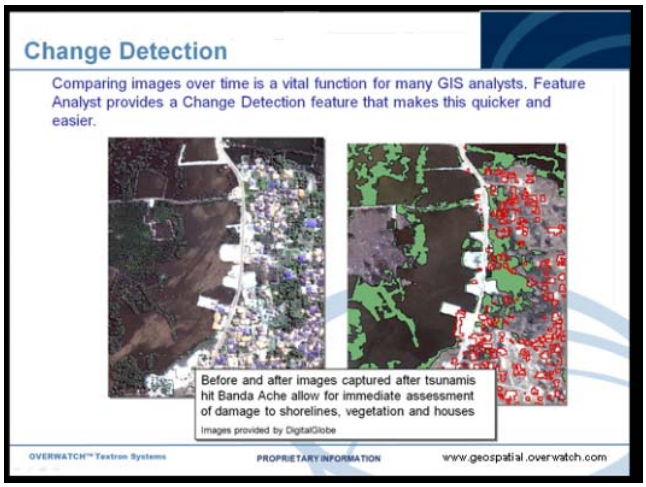


Figure 71 Feature analyst change detection example 9Overwatch .com)

LiDAR Analyst:
www.featureanalyst.com/lidar_analyst.htm

LiDAR Analyst is another product from Overwatch. It is designed to deal with huge LiDAR point clouds and LAS (public binary file format for the interchange of LIDAR) data. It has a suite of tools to convert LAS to raster GIS data formats, extract bare earth, extract buildings (Figure 72) and generate contours.

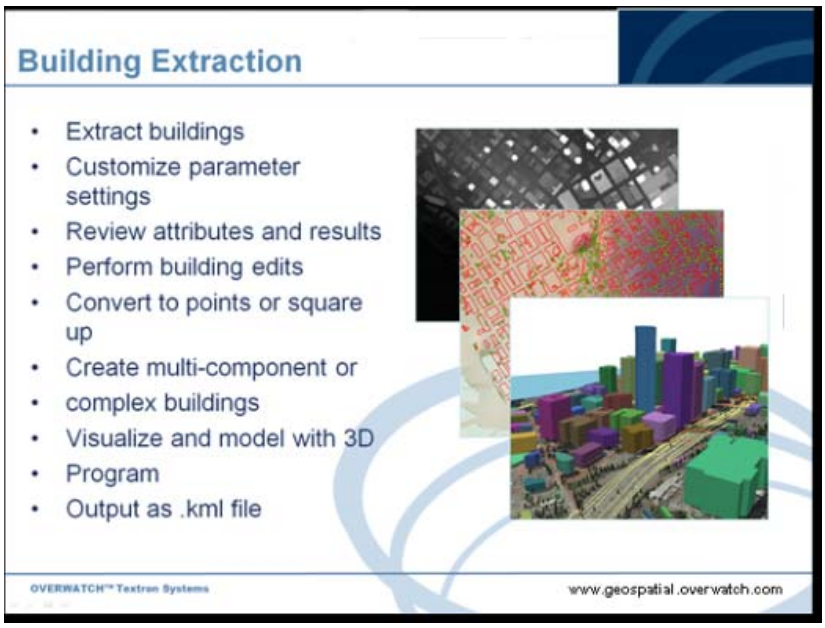


Figure 72 Building extraction from LiDAR analyst (Overwatch .com)

Overwatch Building Collection Toolkit: A special set of tools provide functions to capture buildings from imagery (Figure 73).

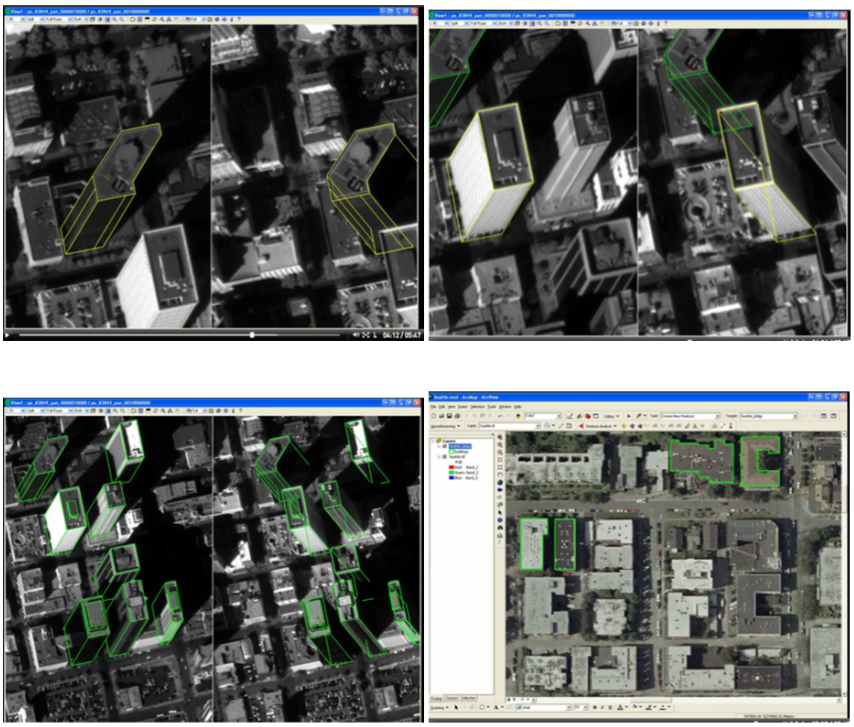


Figure 73 Example of Overwatch building collection toolkit (Overwatch.com)

3.2.4.2.2 Geoeye RoadTracker (Currently Part Of Overwatch Software)

RoadTracker® 2.0, an extension to Feature Analyst is a particularly interesting set of tools that have been developed by GeoEye to assist NGA with the extraction of roads "outside the United States". According to GeoEye (Grodecki, 2010) the software will be available as a plug-in for ArcGIS next year. The interactive software is designed for semi-automated extraction of road features. It has functions to control width and smoothness of the lines, support topological snapping, automatic topology cleaning, and make adjustments based on 1 and N node detours (Figure 74). It is also interesting to note that their experiments include TIGER updates (Figure 75). GeoEye and NGA have been evaluating the performance of RoadTracker based on user-clock-time and spatial consistency (Doucette et al. 2009). The research employed metrics such as the number of mouse clicks when "performing a variety of editing tasks such as adjusting vector vertices, smoothing and straightening vectors, moving terminals (e.g., road intersections), and adding/deleting vectors" They also include useful graph of error assessment (Figure 76)

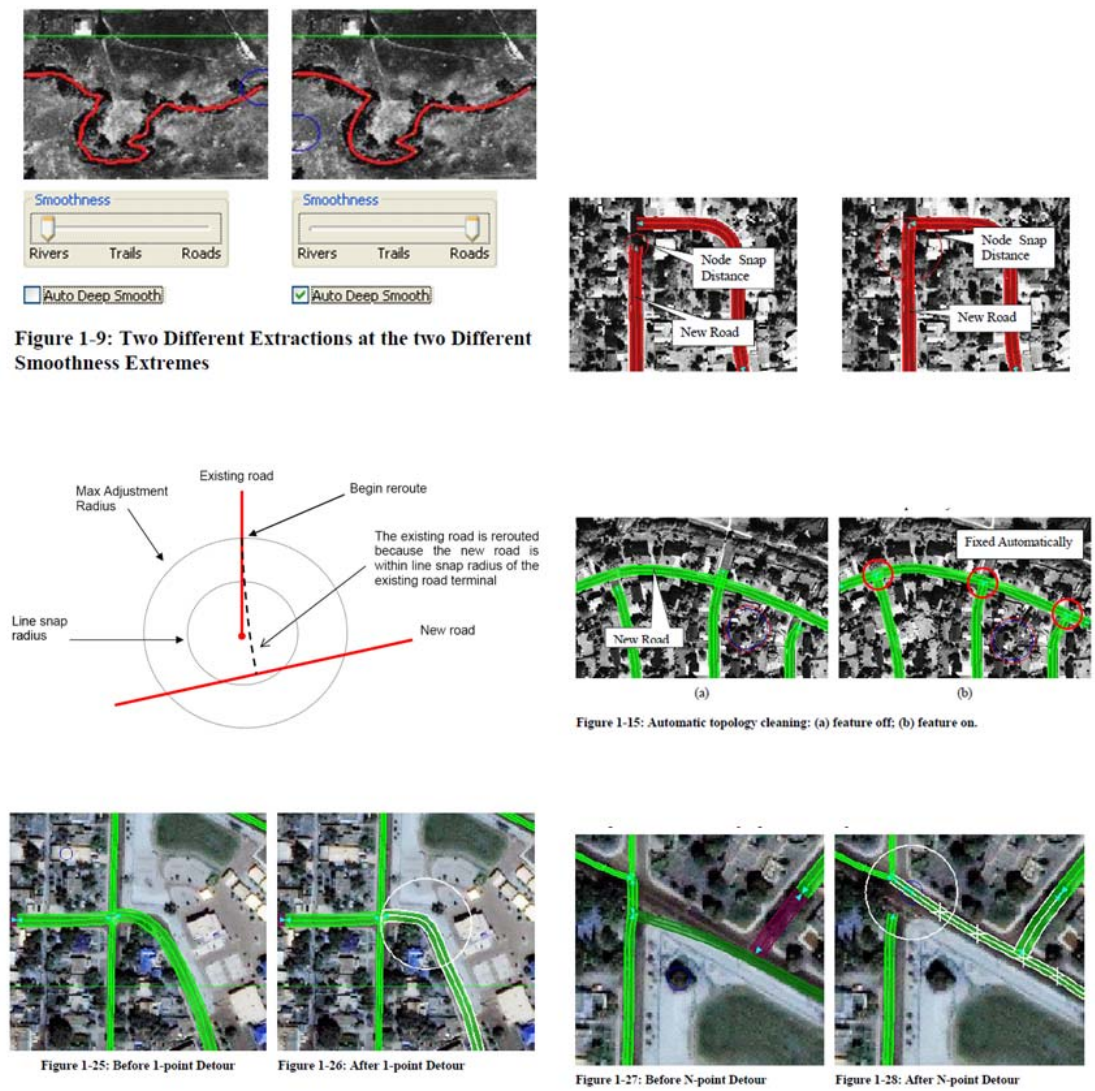


Figure 74 Various functions for interactive road tracking (GeoEye)

Figure 75 Example of RoadTracker comparison with TIGER (GeoEye)

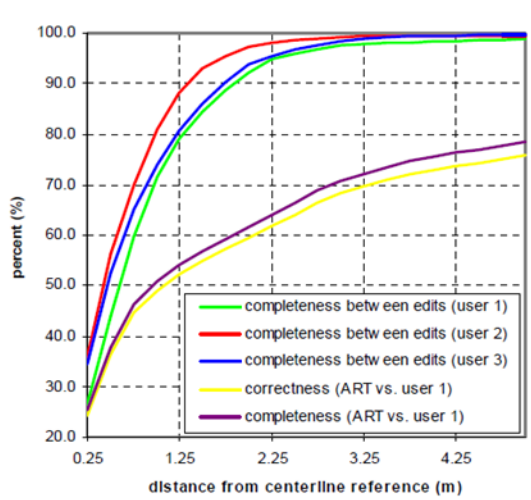


Fig. 6a. Spatial consistency results for test scene 1 (0.7m pan).

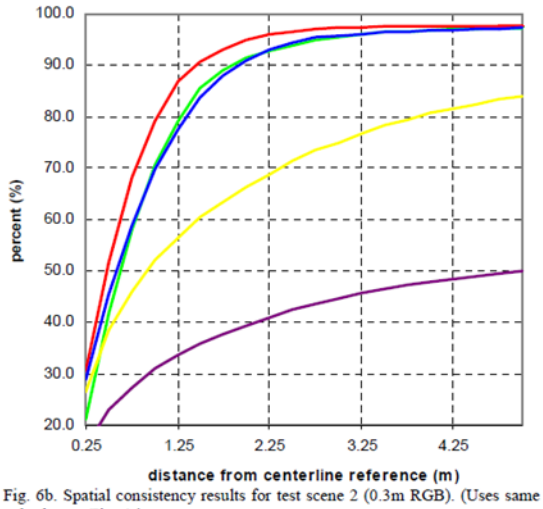


Fig. 6b. Spatial consistency results for test scene 2 (0.3m RGB). (Uses same color key as Fig. 6a)

Figure 76 Spatial consistency graphs for users of RoadTracker (Doucette et al 2009).

3.2.4.2.3 ERDAS

www.erdas.com

ERDAS, along with Intergraph, is now part of Hexagon a Swedish company. ERDAS Imagine Objective is an object oriented toolset that supports:

- Discrete, single feature object detection and multi-class, wall-to-wall object-based classifications (Figure 77).
- Handles the ingestion and fusing of different types of geospatial data including remote sensing imagery, GIS data in both raster and vector formats, scanned paper maps, terrain

data including LiDAR and LiDAR derived data, and polygon-based templates for use when appropriate.

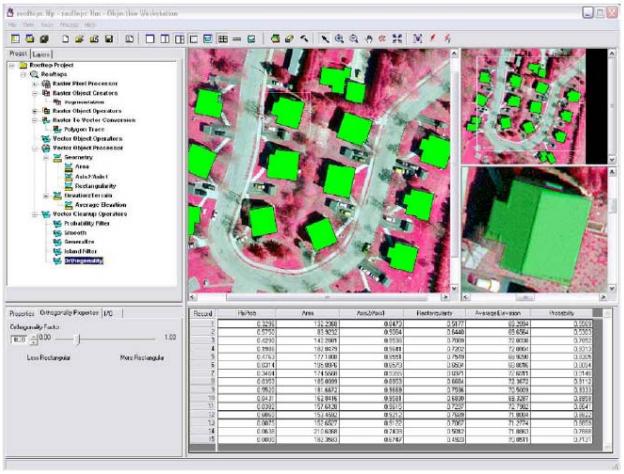


Figure 77 ERDAS Imagine object oriented classification (ERDAS)

IMAGINE DeltaCue is specialized change detection software that provides the following tools

- Change detection algorithms
    - o Magnitude
    - o Tasseled Cap
    - o Primary Color (RGB)
    - o Single Band
    - o Band Slope
- Interactive Change Threshold
- Automatic percent change threshold
- Change Filters
    - o Spectral class
    - o Material type
    - o Area of change
    - o Shape of change
- Provides misregistration filter to minimize the effect of false alarms due to poorly co-registered image pairs
- Automatic Preprocessing
- Automatically crops analysis down to just the area in common to the image pair
- Performs image to image radiometric normalization with accommodation for the effect of clouds and cloud shadow
- Customized Change Viewer
- Geo-linked windows that pan and zoom together with automatic synchronization
- Overlay of change results on top of original imagery
- Change magnifier that allows rapid investigation of before and after areas
- Flicker and swipe tools available off the toolbar
- Contrast and Brightness enhancements easily applied to all layers

• Interactive spectral segmentation tool allows for removal of spectral classes not of interest across entire image
• Access to ERDAS IMAGINE's AOI, annotation and vector tools for output development
• Rapid export of change detection results to ESRI shapefile format

3.2.4.2.4 ENVI
www.idl-envi.com

ENVI SPEAR provides a set of change detection tools that perform the following functions:

- Perform Change Detection: Changed areas can be highlighted and compared to images from two datasets using band ratio techniques
- Detect Anomalies: Search an image for unique objects that are not native to the environment
- Browse Metadata: Browse NITF metadata for multiple images to assess exploitation suitability
- Perform Terrain-based Orthorectification: Compensate for image distortion due to terrain differences
- Register Images to Maps: Improve the georeferencing of an image by tying it to an accurate base map
- Bridge to Google Earth™: Export and open data in Google Earth™
- Pan-sharpen Images: Panchromatic images can be fused with multispectral data to create colorized high resolution images
- Detect Lines of Communication: Roads and waterways can be highlighted and emphasized to aid in map generation
- Find Watercraft: Images can be searched for moving and non-moving watercraft in open water environments
- Measure Relative Water Depth: Images of relative water depths can be generated and exported to graphic files for use in reports or briefings

3.2.4.2.5 eCognition
www.ecognition.com

As noted in the discussion of research efforts in Europe, eCognition has gained widespread support for object oriented classification and change detection. eCognition was developed by Gerd Binnig of Definiens and acquired by Trimble in 2010. It utilizes the similar Cognition Network Technology® that is deployed in medical image processing. Based on spectral and spatial homogeneity, data are segmented into similar regions. By aggregating pixels into image objects, additional information such as texture, shape and context can be leveraged within the analysis process. Object-based image analysis can be done either using samples as guidelines for the computer or using rules in which features are described by their specific characteristics. Once a feature has been identified it can then be extracted in standard GIS data formats.

eCognition can handle shape, texture and data fusion, raster, vector and LiDAR data. Based on these inputs the study area is segmented into polygons which can then be used to classify the entire area (Figure 78). The change detection process and error assessment is outlined in Figures

79 and 80. The company literature cites the following examples of large scale land cover classification:

England 130,000 km2 land cover
Ohio 105,000 KM2 impervious surface polygons and change detection
Fairfield City Australia – land cover
Lower Austria 20,000 km2 LiDAR and CIR Buildings and vegetation
(Accuracy of 94.3 for buildings and 96.1 for vegetation in Austria 60,000 sheets lines)

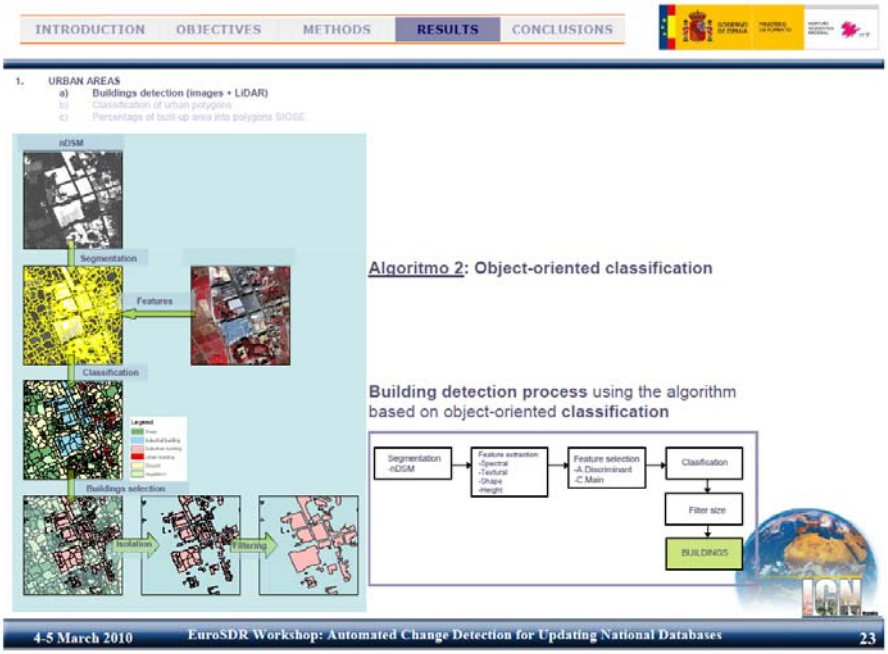


Figure 78 eCognition segmentation process (Domenech, et al., (2010)

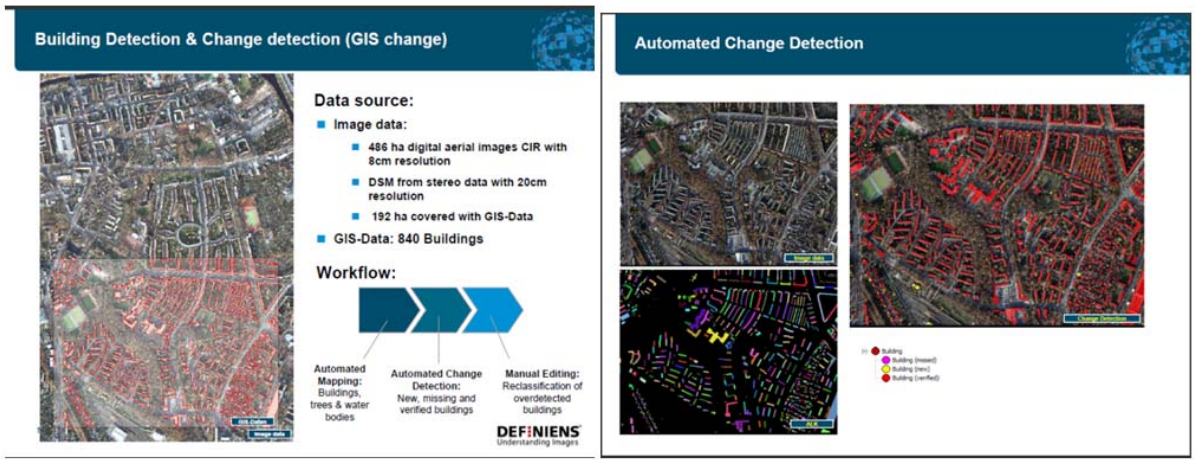


Figure 79 eCognition change detection process (Kühnen and Sohlbach 2010)

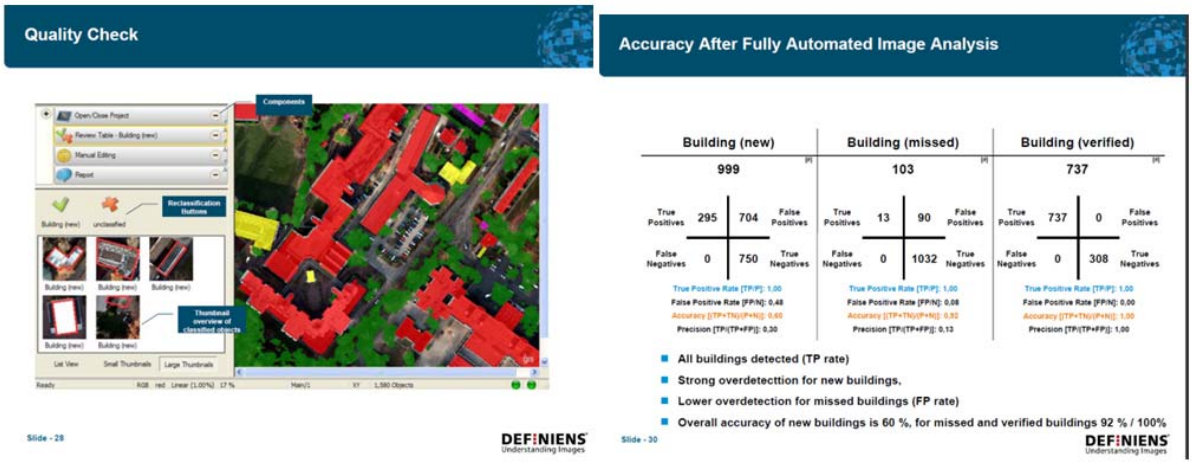


Figure 80 eCognition quality assessment (Kühnen and Sohlbach 2010)

3.2.4.2.6 LP360 (QCoherent Software)

www.qcoherent.com

LP360 software for ArcGIS is a specialized set of LiDAR tools. It includes an interface for generating point cloud statistics and previewing basic classification filters, height from surface filters, height filters, building point classification, building outlines, and macro filter stacks.

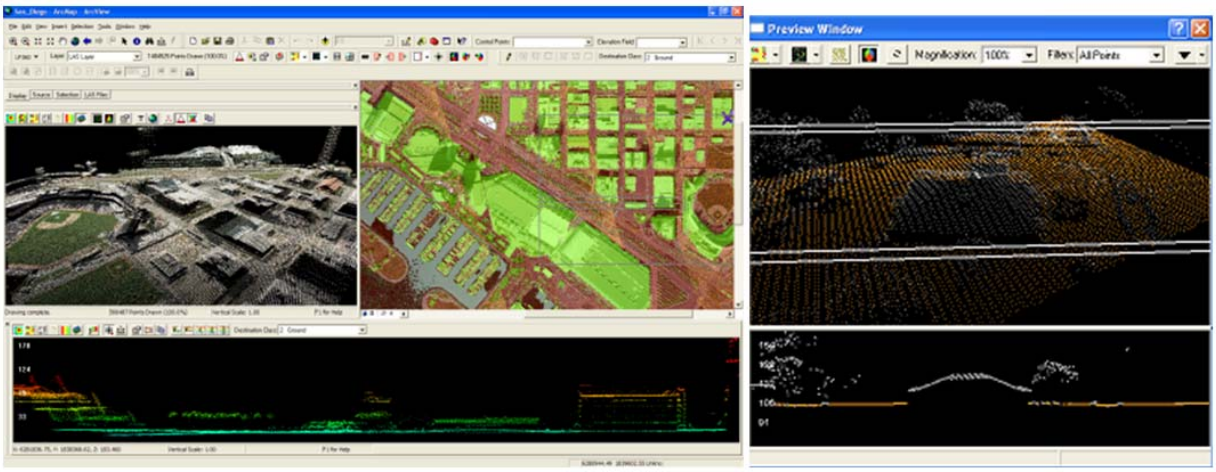


Figure 81 LP 360 ArcGIS interface (Qcoherent)

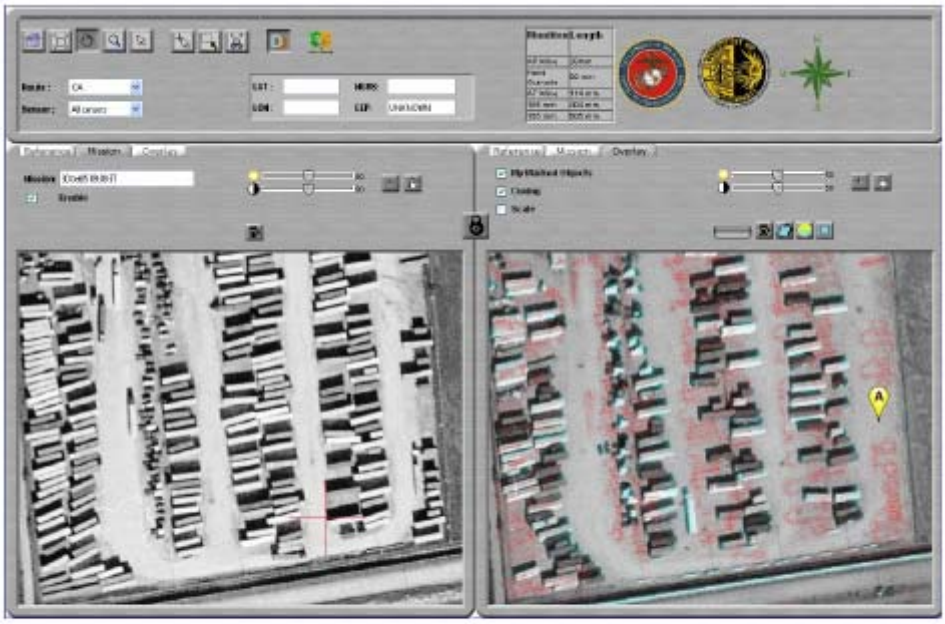
3.2.4.2.7 Observa Change Detection Server (CDS) for Aerial & Satellite Imagery

www.observera.com

According to its promotional literature Observa offers a Change Detection Server (CDS)

> The CDS automatically captures, mosaics, georeferences, projects, compresses, stores imagery and presents it for change detection analysis over a Service Oriented Architecture

(SOA) Web-enabled application called the Change Detection Application (CDA) (Figure 82). The CDA provides side-by-side, geolinked displays that make it easy to compare images from different days. To further facilitate rapid analysis with the CDA, automated cueing has also been integrated into the system. This cueing capability uses a colorized overlay and helps focus the analyst exploitation effort on those highlighted areas of significant change. To date, a total of 20 different sensor/platforms have been successfully tested on the system. Further customization to support additional imagery and metadata formats is easily accomplished.



The CDA Web Application is accessible from any PC connected to the same network as the CDS, offering a powerful imagery exploitation SOA.

Figure 82 User interface for Observa Change Detection Server (CDS)

3.2.4.2.8 IDRISI

www.clarklabs.org

IDRISI from Clark University provides a set of Land Change Modeling tools as either a standalone product or an ArcGIS Extension. The Land Change Modeler provides tools for land cover change analysis and tools for modeling and predicting future landscape scenarios. Specific tools include:

Land Change Modeler
- pairwise image comparison
- multiple image comparison
- predictive modeling and assessment

Image Comparison Tools
- image differencing
- image ratioing

regression differencing
change vector analysis
and qualitative data analysis

Multiple image comparison techniques for use in examining trends and anomalies across multiple images (time series) using:

Principal Components analysis
Fourier analysis
spatial/temporal correlation
image profiling over time

Predictive Land Cover Change Modeling

Markov Chain Analysis
Cellular Automata
Logistical Regression and Multinomial Logistical Regression
GEOMOD and
Artificial Neural Networks

3.2.4.2.9 Hitachi/ Pictometry ChangeFindr™

Many local governments are utilizing the combination of image processing software and visualization tools from Pictometry and Hitachi to identify and report changes in structures (residence, commercial buildings, etc) within a scene over time. These changes are used to modify property assessments and even to prepare input to the LUCA process. This section of our report provides an overview of the components of the process and discusses two case studies that highlight its utility.

The ChangeFindr™procedures include two tools:

(1) The Hitachi HouseDiff™ software creates building outlines (not footprints) from the imagery and detects changes from one set of images to the next. It uses ortho images (not necessarily Pictometry), with a maximum pixel size of 12”. These polygons (Figure 83) form the basis for vector based change detection which can be handled with standard polygon overlay procedures. (Figure 84)

(2) The Pictometry Change Analysis module is a split screen function (available on both thick and thin client applications) that provides views of an old and new image at the same time (Figure 85). The inspection of any two images can start with a tax parcel, lat/long or address. When looking at any new image (ortho or oblique), the software automatically finds the old image for the same point and displays it in side by side format. When changing views on one image, the interface automatically changes the other view to match. Examples of new, changed and demolished properties are portrayed in Figure 86.

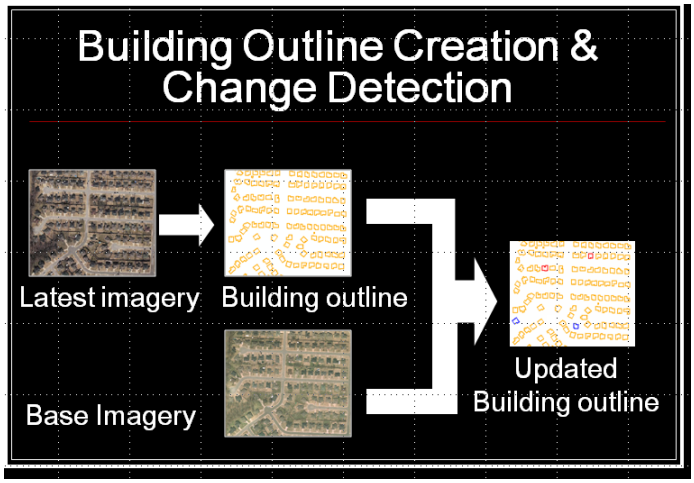


Figure 83 ChangeFinder Building outline creation and change detection (Pictometry)

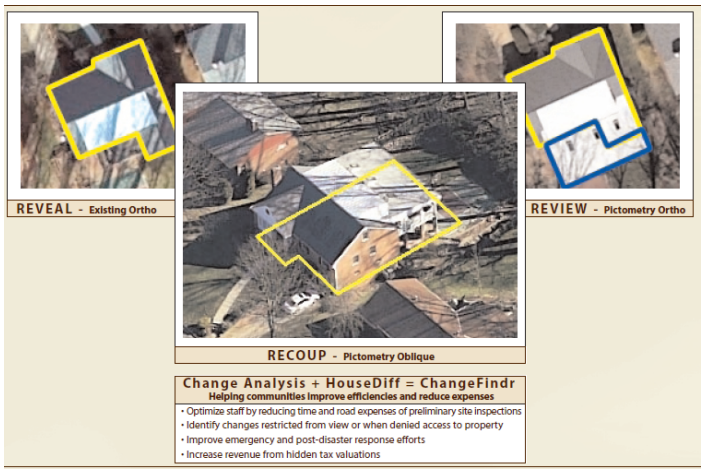


Figure 84 ChnageFinder polygons (Pictometry)

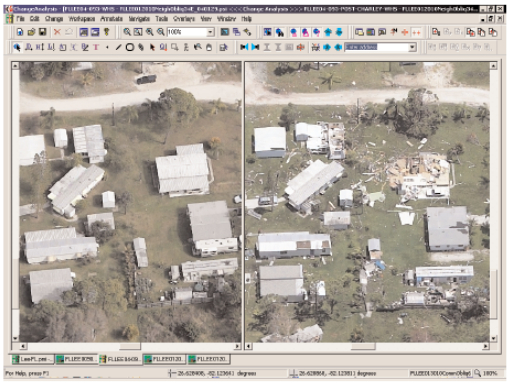


Figure 85 ChangeFinder dual screen visualization (Pictometry)

Figure 86 Examples of existing, new, changed and demolished houses (Pictometry)

Case Study: Waxahachie, Texas

Rodrigue (2008) presented a useful assessment how ChangeFinder was utilized in Waxahachie, Texas, including useful comparisons of field versus office based inspection. In addition, the study documented the ability of the system to successfully identify existing, new, changed or demolished buildings (Figures 87, 88, and 89).

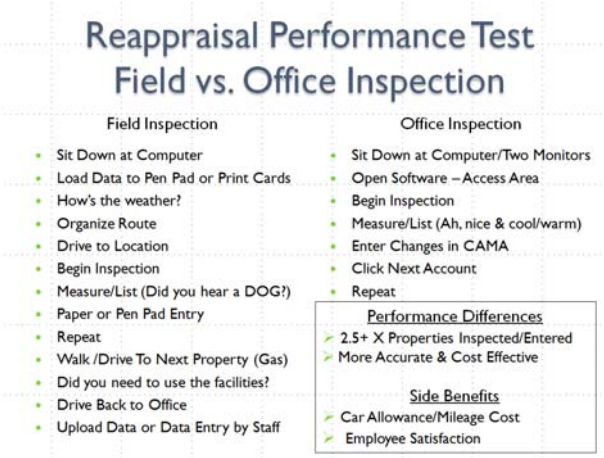


Figure 87 Comparison of field and office inspection (Rodrique 2008)

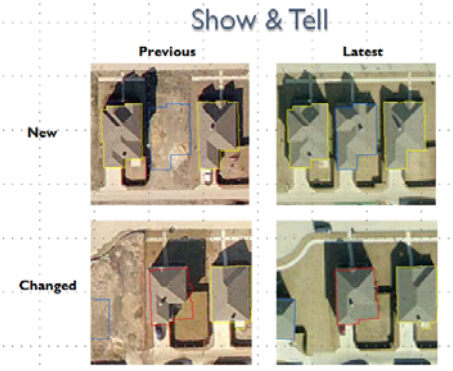


Figure 88 Example of new and changed houses (Rodrique, 2008)

2008 Change Detection Results

Candidate Category	Count	Percentage
Final Number of Buildings	96,390	
Existing	87,819	91.1%
New	3,474	3.6%
Changed	1,485	1.5%
Possible Changed	1,172	1.2%
Demolished	1,133	1.2%
Unknown	1,307	1.4%

These results are representative of 77,914 parcels.
The change count for all candidates other than existing is 8,571.
We identified 2,588 permits.
Potentially 5,983 changes identified through change detection software!
½ the price because we compared images from like warehouses.

Figure 89 Change detection results (Rodrique,2008)

Case Study – Delaware County OH

In 2008 Delaware County Ohio conducted a change analysis for the previous two years with a 95% percent level of accuracy. These data were an important part of the LUCA process and enabled the county to add several thousand additional living quarters to the 2010 address list. She documented that the total cost of using ChangeFinder was $25,000 and noted that the application generated $230,000 in new assessed value for a ROI of $205,000. The following is a summary of the results and cost of the study (Elhami, 2010). During her analysis she found:

Over 76,000 polygons identified as “Existing” (No Change)
Over 3,600 polygons identified as “New”
Over 3,441 polygons identified as “Changed”, “Possibly Changed” & “Unknown”
Over 860 polygons identified as “Demolished”

3.2.4.3 Summary

There is strong evidence that the requirements for detecting changes in building and transportation features require an object based type of classification system that can ingest high resolution imagery, as well as LiDAR derived high resolution elevation data. While pixel based classification and differencing may be useful for broad (Level 1) land cover classifications and changes (such as the MDS CLC product) the ability to find and extract individual features must be based on extracting discrete vector based features. The Pictometry/Hitachi Change Finder has been demonstrated to be an accurate and cost effective solution in rapidly changing counties (suburban “bedroom” communities) and has been adopted by numerous local governments. There is clear evidence that object oriented change detection is gaining favor with many European nations that have high resolution raster and vector source materials. In this environment, change detection is easily performed as a vector based polygon overlay process.

3.2.5 Geoprocessing Procedures

The following section of our report highlights selected vector based geoprocessing procedures that were specifically evaluated for this report, combined with other techniques discovered in the relevant literature.

3.2.5.1 General Procedure For Discovering New Roads

During the course of the research for this report the authors conducted several geoprocessing experiments. The first was designed to determine whether it was possible to efficiently compare TIGER road files with a different set of roads and extract only the new road segments.

As an experiment, the E911 road file from Charleston County, SC was obtained via FTP from the SC GIS Council repository and compared to the most recent TIGER version. Using standard desktop GIS tools all roads in the E911 road file that were positioned beyond a specified search distance from TIGER were identified, segmented and extracted as a separate GIS data layer (Figure 90).

As part of the "address canvass" system these new road segments could be tabulated by tracts. While the execution time was not accurately recorded it took less than ten minutes to find all the new roads in Charleston County. Therefore it is likely that the entire state could be completed in a single day. The results of this experiment suggest that the Geography Division could accept a complete road file from a county without any preprocessing and quickly extract candidate roads for updates to TIGER. This would minimize the burden on local or state partners to identify new roads. Furthermore the files could be harvested from FPT sites with minimal impact on the provider.

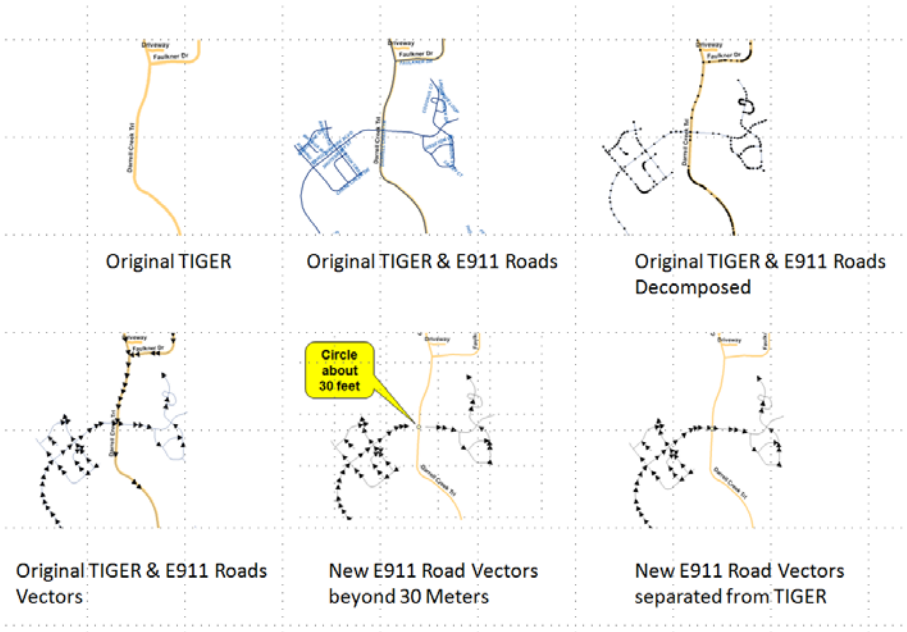


Figure 90 Automated geoprocessing procedure for discovery and separation of new roads (Cowen unpublished)

3.2.5.2 Procedure To Generate Street Centerlines From Parcel Data

A second examination was focused on the use of parcel polygons to create blocks and extract roads. The assumption behind this project is that the local parcel fabric may be the most spatially accurate and current data base maintained by a local government. In fact, for a planned subdivision the new parcels probably exist in digital form before they are constructed. Since tax parcels are the spatial representation of land ownership they differentiate taxable from public land. In effect, the non-parcel areas should correspond with the right of ways where roads should exist. Conversely the parcel data represent where buildings, lawns, forest etc. should be found. Therefore, by dissolving contiguous parcels one should be able to replicate census blocks. Street centerlines should be located in the median of the non parcel areas. This concept was tested in Charleston County SC with a set of core logic parcels (Figure 91). We understand that NAVTEQ employs a similar process to ensure that its street network is properly aligned with the non parcel areas. The parcel based blocks provide a potentially useful way to examine the classification of TIGER street segments. For example, it would be simple to flag any segments that intersect the parcel-based blocks and determine whether they should be classified as a "driveway"

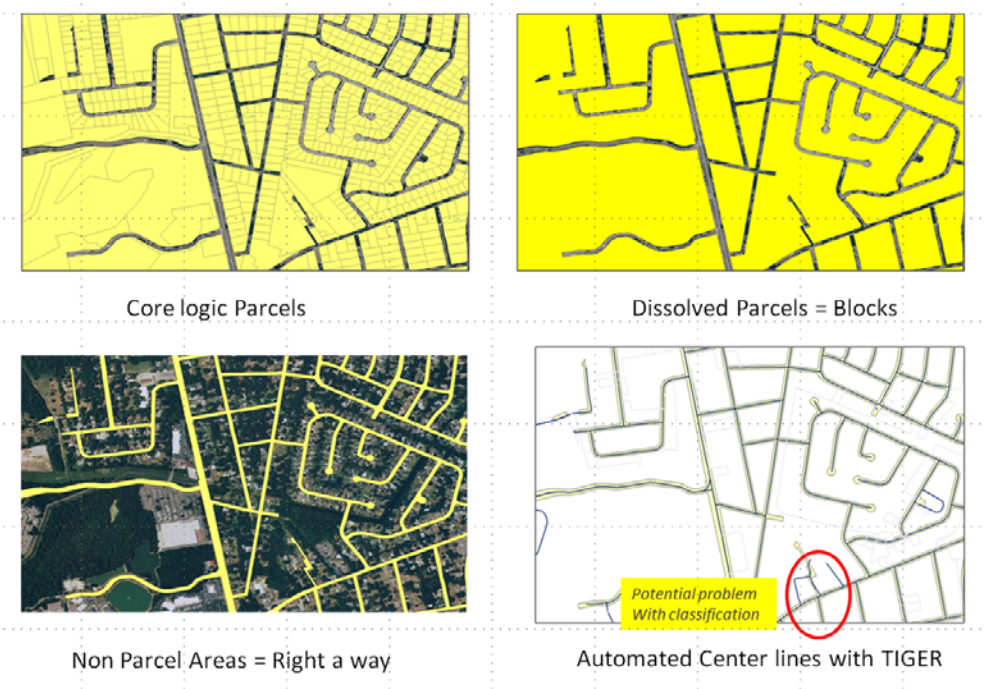


Figure 91 Automated geoprocessing procedure to generate centerlines from parcel data (Cowen, unpublished)

3.2.5.3 Other Recent Geoprocessing Research

Wang H. et al. (2010) provide a "comprehensive approach to support dynamic, large-scale spatiotemporal vector database development" for the National Fundamental Geospatial Information Dynamic Database (NFGIDD) in China. The approach is based on a version-difference spatiotemporal data model that targets common problems in developing large-scale spatiotemporal vector databases (Figure 92).

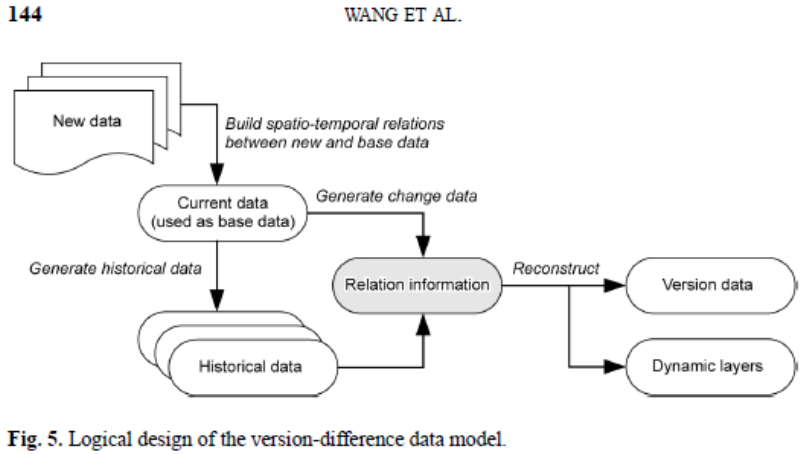


Fig. 5. Logical design of the version-difference data model.

Figure 92 Logical design for version differencing (Wang et al., 2010)

Praing and Schneider (2009) describe process for exploring topological relationships. The goal of the research was to develop efficient implementation techniques of topological predicates for all combinations of the complex spatial data types point2D, line2D, and region2D, as they have been specified by different authors and in different commercial and public domain software packages. Their approach (Figure 93) discovers topological events like intersection and meeting situations and then evaluates them to determine the Boolean result of a topological predicates.

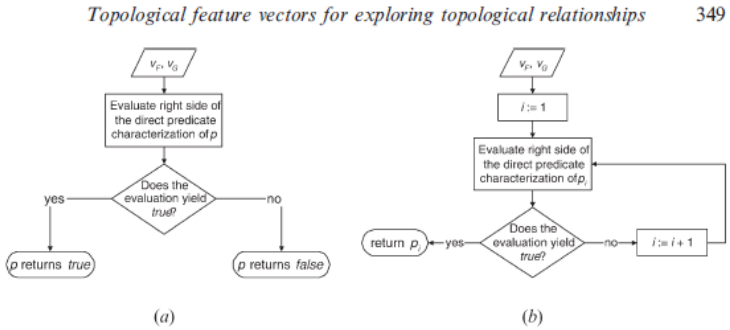


Figure 18. Predicate verification (a) and predicate determination (b) based on the direct predicate characterization method

Figure 93 Predicate verification process (Praing and Schneider, 2009)

Delafontaine et al (2009) describe a general, semi automated method for the assessment of slivers in vector polygon layers (Figures 94 and 95).

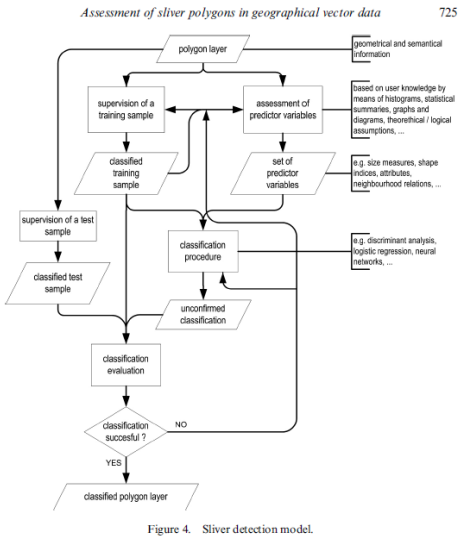


Figure 94 Flow chart for assessing sliver polygons (Delafontaine, 2009)

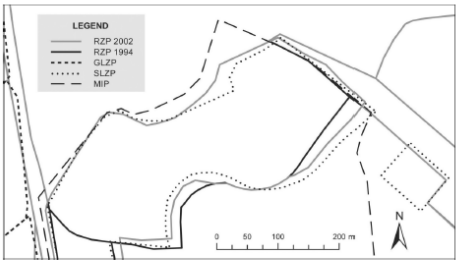


Figure 6. Example extract of slivers in the spatial bookkeeping layer with indication of the layer.

Figure 95 Example of extraction of sliver polygons (Delafontraine, 2009)

Frontiera et al (2008) focus on geographic information retrieval (GIR) performance. All methods are applied to a test collection of queries and documents indexed spatially by two convex conservative geometric approximations: the minimum bounding box (MBB) and the convex hull. Their paper presents a conceptual framework for a modern GIR system as well as its related key technologies.

3.2.6 Summary

After several decades of research and development, the field of digital image processing has become a critical component of production level spatial data update. When these mature tools are coupled with very high resolution image and LiDAR derived elevation data it is possible to extract and update a wide range of features. Based on a review of current practices in local government as well as national mapping organizations it is clear that object oriented processes have superseded pixel based classification and extraction operations. Reports by researchers at national mapping organization in Europe indicate object oriented processes incorporating imagery, LiDAR and vector based GIS files are providing solid data for change detection and data base maintenance even at the level of individual structures. Many local governments in the United States are now routinely deploying automated procedures such as the Hitachi/ Pictometry ChangeFinder system to automatically find and inspect buildings on an annual basis. It is clear that they prefer to conduct virtual building inspections rather sending personnel into the field. A

review of commercial image processing software reveals the importance of object oriented solutions and powerful tools to handle huge amounts LiDAR data. There is also a continuing trend for GIS software (ArcGIS) to integrate image processing and LiDAR functionality either as new tool sets or extensions. Our research suggests that the GeoEye RoadTracker solution is particularly relevant to the GSS Initiative. It may provide the best set of procedures for extracting and editing road features from imagery.

3.3 Investigate the Current and Future Technology to Support Precise Location Measurements (Advanced GPS Positioning).

3.3.0 Introduction

The objective of this section of our report is to provide details on the status of advanced positioning technology and an assessment of potential future developments. Some of the major points of this review were included in our report *Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data*. The important conclusions are that within this decade we will see an expanded system of satellite based technology enhanced further by technology that utilizes the location of cell towers, Wi-Fi routers and even other cell phones. This will be driven by the need for E-911 support for mobile phone users and the nominal cost (about $2) of adding location capability to a cell phone. As the current US-based GPS network expands to a Global Navigation Satellite Systems (GNSS), it will be possible to use 25 or more satellites to calculate a position. The increased number of satellites will provide approximately one-meter accuracy from a standalone receiver with almost no dead spots. Concurrently, the location based services sector is pushing for mobile devices that continuously know their location even within buildings. Why not send a coupon to a customer as they are debating whether to buy that 3D television?

While the latest breed of smart phones provide tools for combining GPS and cellular systems it is useful to categorize them as augmented versus assisted GNSS. This merger was recently enunciated by Higgins (2010) in *The Role of Positioning Infrastructure in the Technological Future of our Profession* https://www.fig.net/pub/fig2010/ppt/ps04/ps04_higgins_ppt_4653.pdf. He highlights the following trends:

- From GPS to GNSS ~ sub meter accuracy is coming
- New possibilities from new GNSS capabilities
- Evolution to a true, integrated global infrastructure
- Increasingly demanding users
- Real time as the next wave of enablement
- From Machine Guidance to Full Automation
- Making positioning truly ubiquitous
- The changing relationship between Positioning Infrastructure and Spatial Data Infrastructure

Higgins also suggests that the field of precise positioning is moving from surveying to machine guidance which will have its greatest impact on agriculture, construction and mining (Figure 96).

In effect there are huge cost savings by bringing the level of precision we expect on a factory floor to exterior activities.

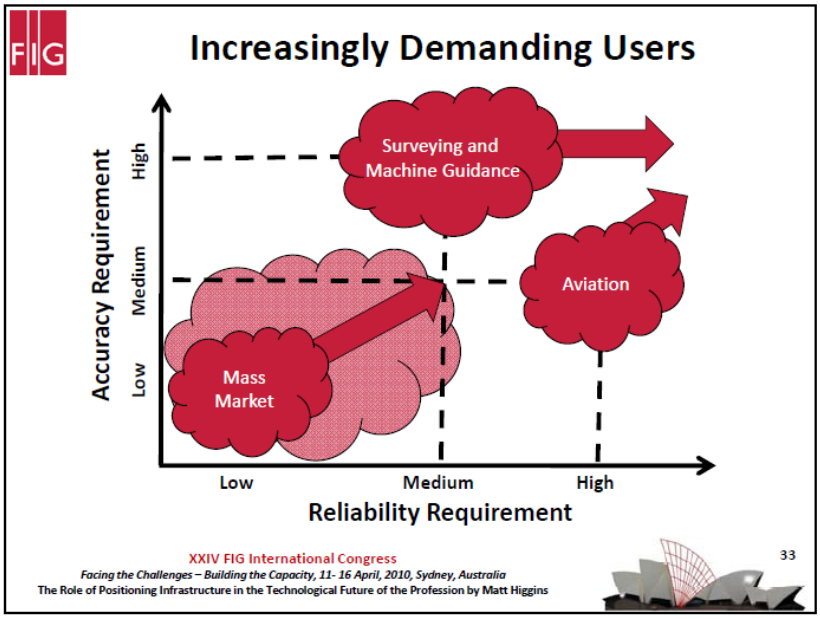


Figure 96 Demand for positioning accuracy from different markets (Higgins 2010)

Another interesting perspective on positioning technology is provided by a 2010 National Research Council Report *Precise Geodetic Infrastructure: National Requirements for a Shared Resource.* The report provides a useful diagram that relates positioning precision and time scale to different application domains (Figure 97). For example, it highlights the need for centimeter level accuracies to monitor changes in sea level. It also highlights some of the technical advancements that are expected to emerge. The report offers the following trends:

- As part of the modernization of the GPS system, three new civilian navigation signals will be introduced besides the C/A L1 signal broadcasted at 1575.42 MHz. L2C will be added to the existing military L2 carrier at 1227.60 MHz.
- The new signal will improve positioning performance in weak signal environments through longer codes that reduce problems caused by cross-correlation between strong and weak GPS signals and lower data demodulation threshold that allows reading of the Navigation Message even when signals are very weak.
- A third civil code – L5 – located at 1176.45 MHz will be introduced later. The L5 code is intended to make GPS a more robust service for aviation and other applications. The first Block IIR-M satellite to add the L2C code was launched in September 2005. Beginning in 2010, GPS Block IIF satellites will introduce the L5 code.
- Finally, the L1C signal will use BOC modulation and consists of a pilot signal spread by a ranging code and a carrier that is spread by a ranging code and modulated by a data message. The launch of the first Block III satellites with L1C has been delayed to 2014. All new civil signals will become initially operational in 2014 and fully operational at the end of the decade.

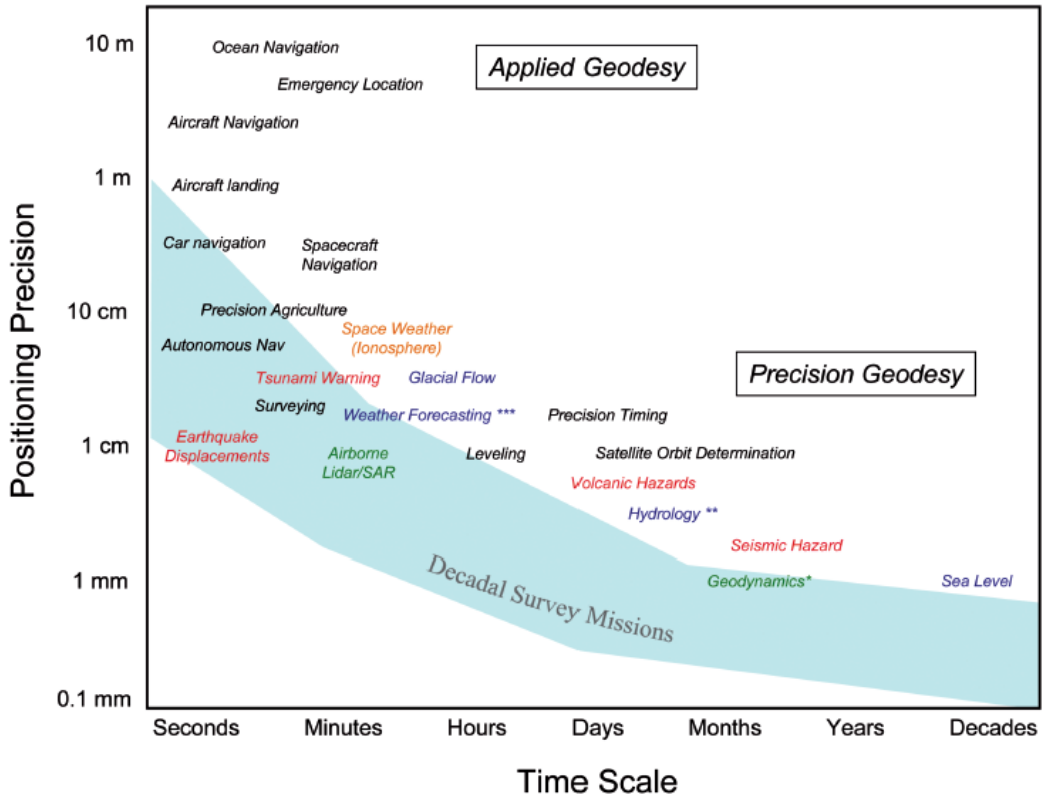


FIGURE 3.1 This schematic plots the precision of current geodetic applications as a function of the required time interval. The most demanding applications at the shortest time intervals include GNSS/GPS seismology and tsunami warning systems. At the longest time intervals, the most demanding applications include sea level change and geodynamics. Note that the positioning scale is in powers of 10 and that range of geodetic applications spans approximately nine orders of magnitude in the time scale. Consistency in connecting the longest to the shortest time scales requires an accurate and stable global terrestrial reference frame, which drives the most stringent requirements on the geodetic infrastructure.

Figure 97 Positioning requirements for different applications (NRC 2010b)

LaMarca and de Lara (2008) provide a useful summary of positioning technology. Their report includes three tables (Figs 98, 99 and 100) that provide an analysis of accuracy, coverage, cost, privacy issues and use cases. Figure 98 covers the standard augmented GPS technology. It demonstrates that the GPS with WAAS augmentation is a good fit for the field based aspects of the GSS Initiative. Figure 99 highlights the General System for Mobile Communications (GSM) and Code Division Multiple Access (CDMA) capabilities. This table indicates that the GSM single strength fingerprinting technology can achieve four meter accuracy when there are 3three or more cell towers visible. The third table summarizes the capabilities of 802.11 [2]technologies to achieve 1 – 3 meter accuracy even in some indoor environments.

[2] A set of standards carrying out wireless local area network (WLAN) computer communication in the 2.4, 3.6 and 5 GHz frequency bands.

★★★★ ★★★☆ ★★☆☆ ★☆☆☆

best ⟷ worst

Technology	Basic GPS	GPS + WAAS	Real-time Kinematic GPS
Accuracy	★★☆☆ 3D coordinates with 10 m median accuracy	★★★☆ 3D coordinates with 2 m median accuracy	★★★★ 3D coordinates with 10 cm accuracy
Coverage	★★★☆ Outdoors with clear view of 4+ GPS satellites	★★☆☆ Outdoors in USA — Requires clear view of 4+ GPS satellites and a WAAS satellite	★☆☆☆ Outdoors with 4+ GPS satellites and requires line-sight between mobile unit and surveyed unit
Infrastructure cost	★★★★ $14 B US initial cost + $500 M US yearly for global coverage	★★★★ WAAS satellites needed in addition to GPS constellation	★☆☆☆ Beyond GPS constellation, requires calibrated, surveyed ground unit
Per-client cost	★★★☆ GPS antenna and chipset required	★★★☆ GPS antenna and WAAS-capable chipset required	★☆☆☆ Special RTK unit required
Privacy	★★★★ Location is estimated passively on the GPS unit	★★★★ Location is estimated passively on the GPS unit	★★★★ Location is estimated passively on the GPS unit
Well-matched use cases	Outdoor navigation for land, sea and air, emergency response, turn-by-turn driving directions, outdoor mapping/information/tour guide services, personell/pet tracking, fitness/activity tracking, gaming	Outdoor navigation for land, sea and air, emergency response, turn-by-turn driving directions, outdoor mapping/information/tour guide services, personnel/pet tracking, fitness/activity tracking, gaming	Outdoor navigation requiring extreme accuracy, surveying, aircraft landing, maritime construction

Figure 98 Summary of the performance characteristics of some common GPS variants (LaMarca and de Lara 2008)

Technology	GSM signal-strength fingerprinting	GSM TOF and signal-strength modeling	GSM/CDMA proximity	Assisted GPS (A-GPS)
Accuracy	★★★☆ 2D coordinates with 4 m median accuracy in dense cell environment	★★☆☆ 3D coordinates with 100 - 200 m accuracy	★☆☆☆ Accuracy dependant on cell tower density (150 m – 30 km)	★★☆☆ 3D coordinates with 10 -150 m accuracy depending on number of GPS satellites visible
Coverage	★★☆☆ Building to campus scale. Requires cell network coverage and radio map. Best accuracy when 3+ cells are visible	★★★☆ Areas with GSM coverage and radio map. Best accuracy when 3+ cells are visible	★★★★ Anywhere with cell coverage and cell-to-location map	★★★☆ Outdoors with 4+ GPS satellites or indoors with cell network support + view of 1+ GPS satellite
Infrastructure cost	★★★☆ No additional infrastructure is needed beyond cell network. Creating radio map is time intensive	★★★★ No additional infrastructure is needed beyond cell network and map of tower locations	★★★★ No additional infrastructure is needed beyond cell network and map of tower locations	★★★☆ Beyond GPS constellation, requires deployment of fixed GPS receivers
Per-client cost	★★★★ Software only solution	★★★★ Software only solution	★★★★ Software only solution	★★★☆ GPS antenna and chipset required for handset
Privacy	★★☆☆ Even if location is computed on client device, the network still tracks a handset's associated cell	★★☆☆ Even if location is computed on client device, the network still tracks a handset's associated cell	★★☆☆ Even if location is computed on client device, the network still tracks a handset's associated cell	★★☆☆ Even if location is computed on client device, the network still tracks a handset's associated cell
Well-matched use cases	Asset and personnel tracking in indoor environments, indoor mapping/navigation/tour guides	Social networking, emergency response, neighorhood-scale information access, fitness tracking, outdoor mapping / navigation	Regional information access (weather, traffic, etc.)	Emergency response, indoor/outdoor information/tour guide services, personnel/pet tracking, activity tracking, gaming

Figure 99 A summary of the characteristics of common GSM/CDMA location estimation (LaMarca and de Lara 2008)

Technology	802.11 signal-strength fingerprinting (e.g., RADAR)	802.11 signal-strength modeling (e.g., Place Lab)	802.11 proximity (e.g., GUIDE)
Accuracy	★★★☆ 2D coordinates with 1-3 m median accuracy	★★☆☆ 2D coordinates with 10-20m median accuracy	★☆☆☆ Location accuracy dependant on AP density
Coverage	★★☆☆ Building to campus scale. Requires 802.11 coverage and radio map. Best accuracy achieved when 3+ APs are visible.	★★★☆ Areas with 802.11 coverage and radio map. Best accuracy achieved when 3+ APs are visible.	★★★☆ Anywhere with 802.11 coverage and an AP, location map.
Infrastructure cost	★★☆☆ No additional infrastructure is needed beyond 802.11 APs. Creating radio map is time intensive and new/moved APs require remap.	★★★☆ No additional infrastructure is needed beyond 802.11 APs. Creating radio maps is less work than for fingerprinting.	★★★★ No additional infrastructure is needed beyond 802.11 APs.
Per-client cost	★★★★ Software-only solution for devices with 802.11 NICs.	★★★★ Software-only solution for devices with 802.11 NICs.	★★★★ Software-only solution for devices with 802.11 NICs.
Privacy	★★★★ when localization is performed on the client. ★☆☆☆ when localization is performed in the infrastructure.	★★★★ when localization is performed on the client. ★☆☆☆ when localization is performed in the infrastructure.	★★★★ when localization is performed on the client. ★☆☆☆ when localization is performed in the infrastructure.
Well-matched use cases	Asset and personnel tracking in indoor environments, indoor mapping/navigation/tour guides	Social networking, tour guides, indoor/outdoor navigation/tour guides, fitness/activity tracking	Outdoor tour guides, nearby resource advertisement, activity tracking

Figure 100 A summary of the performance characteristics of 802.11 location estimations techniques (LaMarca and de Lara 2008)

3.3.1 Overview of Positioning Technology

As a result of some enlightened policy decisions, the Department of Defense's Navigation Signal Timing and Ranging Global Positioning System NAVSTAR GPS, system became a freely accessible global positioning technology. (See box below for an overview.) This not only changed navigation and surveying but has fostered an unprecedented expansion of location based services.

Berg Insights (Malm, 2010) provides a useful summary of the existing Global Positioning System (GPS)

GPS is operated and maintained by the US Department of Defense. The official name of the system is NAVSTAR GPS, Navigation Signal Timing and Ranging Global Positioning System. The space segment consists of 24 operational and 3 spare satellites, called space vehicles (SVs). The satellites are placed in intermediate circular orbits about 20,200 kilometres above the ground, with satellites grouped into 6 orbital planes of 4 satellites each. The orbital planes are spaced 60 degrees apart, inclined 55 degrees from the equatorial plane.

A GPS receiver calculates its position (latitude, longitude and elevation) in four basic steps.

First, the receiver measures the distance to the GPS satellites,
second it calculates the exact position of the satellites,
third, it performs trilateration calculations and
fourth, it adjusts its internal clock.

The distance to visible satellites is simply the travel time of the signal multiplied by the velocity of the signal, i.e. the speed of light. The receiver utilises the C/A code to obtain the travel time of the signal; the receiver replicates the signal sent by the satellite by generating the same PRN code as the satellite and measures the delay between the locally generated signal and that received from the satellite. The exact position of the satellites can be calculated from the ephemeris parameters embedded in the Navigation Message. The receiver can then calculate the distance to each visible satellite.

The receiver is located at the intersection of three imaginary spheres, one around each satellite, with a radius equal to the time delay of the signals travelling from the satellite to the receiver multiplied by the speed of the radio signals. Using trilateration, the receiver obtains its pseudo location. The intersection of three spheres yields two intersection points of which one can usually be discarded at once (a position in outer space). However, the internal clock of the receiver is not very accurate and if the internal clock errs by only one second every month the position would be off by over 100 meters due to the timing error. Thus, a signal from a fourth satellite is needed. If the signal from a fourth satellite gives an intersection with the first three signals, the internal clock is correct and the pseudo position is then the correct position. Otherwise, there is a timing error and the receiver clock is adjusted until the imaginary spheres around all four satellites intersect at one point. The position is then known and the internal clock is adjusted accordingly.

In a real world setting performance is greatly enhanced when the GPS device quickly determines a position (Time-to-first-fix) Modern receivers have very high signal processing performance and are therefore able to calculate positions fast. However, because the navigation message is transmitted at only 50 bps, the time needed for collection of the Navigation Message can have a significant impact on the total time to calculate a position. If the receiver has been switched off for less than 20 minutes, it can perform a quick-start where the receiver estimates the correct time and does not have to wait for time stamps in the satellite signals. A quick-start can be performed in a few seconds. A quick-start needs a valid word, requiring up to 1.18 seconds for collection. If the ephemeris data in the memory is valid, the receiver can perform a hot-start, where only the internal clock has to be synchronized using the time stamp from a subframe. Collecting a valid subframe can take up to 6.6 seconds. In a warm-start, the ephemeris data is over 4 hours old and thus invalid. To collect the ephemeris data a frame is needed, which takes up to 36 seconds. From a cold-start, when the receiver does not hold valid ephemeris data and the clock is not synchronized, the time to the first fix is over 36 seconds. Besides collection of a valid frame, taking up to 36 seconds, the receiver must search the whole sky for the signals.

Lapine (2010) also provides a recent summary of positioning technology in another NRC report *A Brief History of Satellite Geodesy - October 4, 1957 to Present*. He provides the following review:

- A user equipped with a GPS receiver that can receive the correctors in real time can achieve a positional accuracy of one meter or less. Accuracy of decimeters can be achieved if the DOP is sufficiently low. This type of relative positioning is frequently used for mapping applications
- Mapping grade accuracy can also be obtained by post-processing using data from the National Geodetic Surveys CORS. Relative positioning can also be performed by occupying known and unknown points simultaneously and collecting up to one hour of simultaneous data. The data are processed using a method known as double differencing, which mathematically removes the systematic errors related to the satellite orbit, clock and atmosphere. Accuracies of a few centimeters are possible with this technique

- Incremental positioning uses a network of known and unknown points connected by a series of relative observations. The relative vectors between points are mathematically fit by least squares computations to improve the overall accuracy. Horizontal accuracy of 1 cm and vertical (ellipsoid) accuracy of 2 cm is possible when following the NOAA National Geodetic Survey's guidelines for National Height Modernization.
- Accuracies of centimeters require the use of GNSS receivers that can take advantage of both L1 and L2 frequencies to remove the effects of ionospheric refraction.
- More recently, surveyors and engineers have demanded real time centimeter accuracy for applications ranging from precision agriculture to machine-controlled road construction. The initial solution was to use Real-Time Kinematic (RTK) surveying, in which one GNSS receiver was positioned over a known geodetic control point (the base station) and transmitted its raw satellite data via a VHF radio modem to a second receiver located over an unknown point (the rover).
- The accuracy of rover coordinates was entirely dependent on the accuracy of the base station coordinates. RTK has since matured into the technology of Real Time Networks (RTN). Like the CORS only with closer spacing of reference stations, centimeter level positions can be obtained anywhere within the RTN where cellular coverage is available.
    - A single dual frequency receiver cell phone with CDMA capability is all the equipment that is necessary. RTNs are usually established by state governments, although some private networks exist
    - Nearly all RTNs use both GPS and GLONASS to increase the number of satellites available at any given moment

The impact of the improvements in positioning technology has been profound. To celebrate NOAA's 200th anniversary Smith (2007) speculated about additional impacts in the next twenty years *The Future of Global Navigation Satellite Systems: Anyone, Anywhere, Anytime, Any Accuracy* (See Box below). For example, he predicts that we will have autonomous (standalone) positional accuracies of about 3.3 feet. In addition, with access to more than 20 satellites at all times, users would always have access to at least six satellites eliminating the problem of multipathing in urban canyons, as well as providing four dimensional accuracy that will take into account subsidences of as small as five mm per year.

Dru Smith with NOAA's National Ocean Service has speculated about the future of positioning technology based on the expansion of the GPS to GNSS. In this presentation The Future of Global Navigation Satellite Systems: Anyone, Anywhere, Anytime, Any Accuracy by 2027 (Smith, 2007)

“ A GNSS allows electronic receivers to determine longitude, latitude, and elevation of the receiver on the Earth's surface. Future users of GNSS will see many more satellites than today with more signals per satellite. Positioning errors will drop to below one meter! “

Smith provides interesting scenarios about the impact of the GNSS

Portable, Hand-held Devices

Using tracking signals from GPS, Galileo, and GLONASS, her positional error is about 3.3 feet (1 meter), an improvement over the 33-foot (10-meter) accuracy provided by hand-held devices in 2007. Nodding with approval, she glances at her friend Jean, who is using her cell phone/GNSS receiver to check their position. Marie is amused to see Jean using one of those "old fashioned" single-frequency units, but Jean's positional accuracy is not much worse than Marie's, as Jean is also receiving ionosphere corrections from a wireless Internet connection. These corrections, provided by NGS and NOAA's Space Environment Center, remove about 23 feet (7 meters) of positioning error normally caused by radio waves passing through the ionosphere. Still, Marie can't help but point out that her own receiver is receiving 10 different positioning signals from three constellations. In 2027 strong GNSS signals will penetrate the tree canopy and a position above sea level, which has been calculated using a NOAA-generated, highly accurate model of the Earth's gravity field. Using tracking signals from GPS, Galileo, and GLONASS, positional error will be about 3.3 feet (1 meter), an improvement over the 33-foot (10-meter) accuracy provided by hand-held devices in 2007

Automobile Navigation

She flicks on the GNSS navigator (a "navaid") built into her car's dashboard and hears the command "tracking 30 satellites." Startled at this high number, she checks the unit, now remembering that she set it to also track Compass satellites, the Chinese GNSS that was only announced in 2006. The navigational unit is also tracking the High-accuracy Differential GPS (HA-NDGPS) correction signals. These signals tell Marie what other nearby GNSS receivers are experiencing as far as atmospheric errors and GNSS satellite orbit uncertainties, and help her own GNSS receiver correct for them. With more satellites in the sky, GNSS based navigation aids will still work in many cities. With more satellites in the sky, GNSS based navigation aids will still work in many cities. The navaid informs her that her positional accuracy is four inches (10 centimeters). As she drives along, the navigational unit tracks what lane she is in, occasionally giving her commands such as "move one lane over to prepare for left turn." Marie recalls that back in 2007, early navigational devices in cars could only determine what road a driver was on—not in what lane. She smiles thinking about how far technology has come in the last 20 years! As she drives through the city, the view of the sky is significantly blocked by tall buildings. Her navaid tells her that it is only tracking six satellites. Luckily, six satellites are enough to fix her position, at least to the level of the correct road, if not the correct lane.

Surveying Work

Marie works in a state that runs its own Real-Time Kinematic (RTK) GNSS corrector network. This network is where data from a collection of permanently set GNSS receivers with known positions, called "base stations," are accessed by users with a "roving" GNSS receiver to determine a differential position of the rover relative to the fixed "base stations." Because of this network, Marie's surveys will be relative to the state-owned base stations. Although today's work will cross over into another state, this doesn't worry her as she knows that all RTK networks in the United States have been accredited by NOAA as being "NSRS consistent," meaning they will work together seamlessly. With the widespread use of independent RTK networks, it is the duty of NGS to verify that each of them is consistent with the others in providing coordinates that are reported in the National Spatial Reference System (NSRS). Marie loads up her equipment and heads to the field where she will be performing a topographic survey for the installation of a new tri-state power plant. She flicks on her geodetic-quality RTK equipment, sees that it is tracking four different GNSS systems, as well as the statewide RTK correctors from the network. She gets to work. She moves from point to point, pausing for only a second to get her one-centimeter position before moving on to the next point. As she collects points, a map appears on the computer screen in her truck, indicating the topographic layout being collected.

After a few hours of data collection, Marie is getting ready to leave the field. Before doing so, however, she hits the "submit and check" button on her receiver. She receives a "subsidence warning" from NOAA. Unlike the three-dimensional geodetic control of 2007, in 2027, geodetic control is four-dimensional. Four-dimensional control means that areas of subsidence (land sinking) are well known and warnings as to changing topography are built into surveys. Marie's warning reads: "Note, the area you surveyed has an average subsidence rate of 5 mm / year. Any data collected today will be relative to the new North American Datum (NAVD), with today's date stamped. Future surveys in this region will account for this subsidence rate." Since 2007, the aging NAVD 88 and NAD 83 official datums of the United States were replaced by more accurate, dynamic datums which reflect the variety of changing elements in the world, such as subsidence, tectonics, and sea level rise.

As a result of the increase in the increased availability of satellites it is feasible that the Accuracy of an unassisted Single Point Position using 25+ Satellites with dual frequency signals will be at the several decimetre level. This is essentially without augmentation from stations on the ground or geostationary satellites sending additional information to the receiver.

3.3.2 Sources of Error in GPS

In order to utilize satellite based positioning technology it is important to understand the sources of error that degrade the process. Generally the sources of error and the associated errors are: (http://www.kowoma.de/en/gps/errors.htm)

- Ionospheric effects ± 5 meters
- Shifts in the satellite orbits ± 2.5 meter
- Clock errors of the satellites' clocks ± 2 meter
- Multipath effect ± 1 meter
- Tropospheric effects ± 0.5 meter
- Calculation- and rounding errors ± 1 meter

Other factors

- Multipath
- Earth Tides
- Receiver Clock
- Receiver Biases

Since the 24 GPS satellites are in different orbits, the ability to calculate a coordinate is a function of the number and position of satellites that can be "seen" from the specific location. The term Dilution of precision (DOP) or geometric dilution of precision (GDOP) is used to measure the effect of GPS satellite geometry on GPS precision. In fact, there are respectively horizontal (HDOP), vertical (VDOP), positional (3D) (PDOP), and temporal (TDOP) dilution of precision. When visible GPS satellites are close together in the sky, the geometry is said to be weak and the DOP value is high; when far apart, the geometry is strong and the DOP value is low. Basically, one should not utilize a GPS device if the DOP value is higher than 6 (Figure 101). Since the path of the satellites is well known and predicable it is possible to coordinate the field data collection activity to correspond with the best configuration of the satellites.

DOP Value	Rating	Description
1	Ideal	This is the highest possible confidence level to be used for applications demanding the highest possible precision at all times.
1-2	Excellent	At this confidence level, positional measurements are considered accurate enough to meet all but the most sensitive applications.
2-5	Good	Represents a level that marks the minimum appropriate for making business decisions. Positional measurements could be used to make reliable in-route navigation suggestions to the user.
5-10	Moderate	Positional measurements could be used for calculations, but the fix quality could still be improved. A more open view of the sky is recommended.
10-20	Fair	Represents a low confidence level. Positional measurements should be discarded or used only to indicate a very rough estimate of the current location.
>20	Poor	At this level, measurements are inaccurate by as much as 300 meters with a 6 meter accurate device (50 DOP × 6 meters) and should be discarded.

Figure 101 Description of PDOP values and data capture quality (Esri)

3.3.3 Augmentation

In order to meet the specific user requirements for positioning, navigation, and timing (PNT), a number of augmentations to the Global Positioning System (GPS) are available. An augmentation is any system that aids GPS by providing accuracy, integrity, reliability, availability, or any other improvement to positioning, navigation, and timing that is not

inherently part of GPS itself. It can utilize satellite or ground base signals to improve the positional accuracy. Simply stated, there are a set of precisely located stations that are simultaneously receiving the same signals at the roving receiver. Given that these stations are fixed and well known then the position of the rover is augmented through trigonometrically calculations. At the highest end applications (survey grade) an operator in the field would need to be in constant contact with a ground based station that is transmitting highly accurate data. The ability to perform augmentation is a function of the bells and whistles on the device. Specially if one has dual channel device that can receive data from more than 6 satellites and an antenna to receive radio transmissions – or now even a cell phone that can connect to a network - it is possible to capture highly accurate coordinates. A chart provided by American Surveyor provides an excellent review of the level of accuracy that is achievable with different forms of augmentation (Figure 102). In this chart the RTN refers to real-time networks, SBAS refers to Satellite based Augmentation Systems.

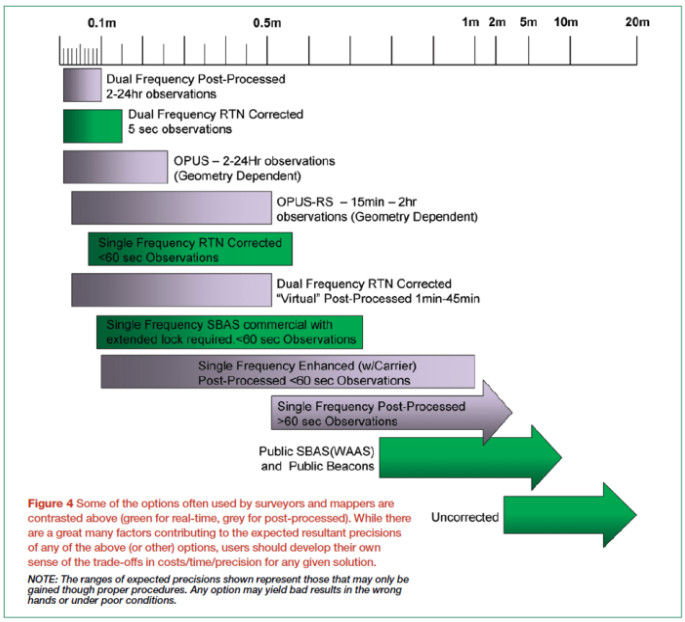


Figure 102 Chart of accuracy capabilities from different augmentation systems (Schrock 2008)

3.3.3.1 Real Time Versus Post Processing Augmentation

Augmentation can occur either in real time or through post-processing following the data collection. Real time processing is critical for a host of location based services that require continuous location. It is also of great assistance in field data collection efforts that benefit from immediate visualization of the coordinates as they are collected. However, real time augmentation requires the receipt and processing of additional correction information. Traditionally, this has meant additional antennas and communication devices. Therefore, precise differential correction or augmentation would often take place through post processing. Through satellite based augmentation systems such as WAAS, the planned significant increase in the number of satellites in the next few years and the incorporation of communication technology (Cell Towers, Wi-Fi and even other phones), it will be possible to capture real-time coordinates that will meet the needs of GSS for accuracy and precision.

3.3.3.1.1 SBAS Satellite (or Space) -Based Augmentation System.

SBAS technology utilizes a (geostationary) satellite system to send augmentation correction data to a roaming receiver. These systems retransmit data from ground stations that significantly reduce ionospheric effects (5 m) on positional accuracy. There is a great benefit from not being constrained by the location of ground based stations.

3.3.3.1.1.2 Global Satellite Based Augmentation System (GSBAS)

GSBAS systems provide absolute positioning, meaning they do not determine position relative to some fixed point of the Earth's surface. Instead, they determine a position within a space-based reference frame, thus they may also be considered to be within the Space Domain group. Augmentation for Standalone GPS consists of satellite orbit and clock updates which are generated by the GPS Ground Control Segment and distributed to the user with an ionospheric model in the GPS signal structure**.** It is this basic principle of determining each error source instead of amalgamating them which distinguishes GSBAS from dGPS.

- **Government Supported SBAS**
  - United States Wide Area Augmentation System (WAAS)
  - European Geostationary Navigation Overlay Service
  - Japan's Multi-Functional Satellite Augmentation System
  - Canada's CDGPS
- **Commercial (Access to these system require a subscription and are often used for precision agriculture, mining or construction)**
  VERIPOS, The VERIPOS Ultra service is a precise positioning service designed to deliver decimeter level position accuracy, globally. The service is based around Precise Point Positioning (PPP) – a technique where all GPS system errors are removed or minimized by direct calculation, precise modeling or estimation.
  StarFire NavCom a John Deere Company.
  OmniSTAR, Inc. is part of the Fugro Group of Companies.

3.3.3.1.1.3 Wide Area Augmentation System (WAAS)

For the GSS Initiative efforts the WAAS system of space based augmentation appears to meet the required level of accuracy. A technical discuss of how this can be incorporated into mobile devices is provided in the Report *Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data*. WAAS was established and is maintained by the FAA to support aircraft navigation for all phases of flight.

> "Unlike traditional ground-based navigation aids, the WAAS covers nearly all of the National Airspace System (NAS). The signals from GPS satellites are received across the NAS at many widely-spaced Wide Area Reference Stations (WRS) sites. The WRS locations are precisely surveyed so that any errors in the received GPS signals can be detected. The GPS information collected by the WRS sites is forwarded to the WAAS Master Station (WMS) via a terrestrial communications network. At the WMS, the WAAS augmentation messages are generated. These messages contain information that allows GPS receivers to remove errors in the GPS signal, allowing for a significant

increase in location accuracy and reliability. The augmentation messages are sent from the WMS to uplink stations to be transmitted to navigation payloads on Geostationary communications satellites. The navigation payloads broadcast the augmentation messages on a GPS-like signal. The GPS/WAAS receiver processes the WAAS augmentation message as part of estimating position. The GPS-like signal from the navigation transponder can also be used by the receiver as an additional source for calculation of the user's position. WAAS also provides indications to GPS/WAAS receivers of where the GPS system is unusable due to system errors or other effects. Further, the WAAS system was designed to the strictest of safety standards – users are notified within six seconds of any issuance of hazardously misleading information that would cause an error in the GPS position estimate." (www.faa.gov)

3.3.3.1.2 Ground Based Augmentation System (GBAS).

Ground based augmentation systems rely on additional correction information to be received by a roaming GPS receiver. They work by comparing satellite signals received by the rover with those received by a base station which is fixed over a highly accurate surveyed point. Base station correction values are calculated and then applied to the rover data to increase their accuracy to 5 meters or less depending on the GPS receiver grade. These systems are also referred to as Real-Time Networks, Network Corrected Real-Time, or Network RTK. Real-Time Kinematic (RTK) RTK is a highly precise technique that results in one inch pass-to-pass and year-to-year accuracy. RTK GPS requires two specialized GPS receivers and two radios. One GPS receiver is set up as a base station within a 6 mile (9.6 km) radius of the field you are working so it can send the correction message to the roving receiver. Both receivers collect extra data from the GPS satellites known as L2 Band that enables this better precision.

3.3.3.1.2.1 Differential GPS provides relative positioning; these positioning options determine corrections at the Earth's surface, either at individual sites or averaged over a region. These corrections are an amalgam of primary reference error sources and are transmitted to the user via radio or satellite link. It removes most types of error from GPS data and can occur back in the office – post processing or in the field. User coordinates are determined relative to the reference site(s). Accuracy degrades with distance from the individual reference site or the regional boundary, primarily because of differential ionospheric and tropospheric effects, satellite orbits, and clocks. Examples of correction services include:

- Beacon DGPS
- Commercial DGPS
- RTK / Network RTK
- Widelane RTK

The following terms are defined at (http://www.gps.gov/systems/augmentations/)

3.3.3.1.2.1.1 The Network RTK correction information provided to a rover can be considered as interpolated corrections between the reference stations in the RTK network. Depending on the actual conditions of the atmosphere, this interpolation is not perfect. A residual interpolation error has to be expected. With sufficient redundancy in the RTK network, the network server process can provide an estimate for residual interpolation errors. Such quality estimates may be

used by the rover to optimize the performance of RTK solutions. The values may be considered by the rover as a priori estimates only, with sufficient tracking data available the rover might be able to judge residual geometric and ionospheric errors itself.

3.3.3.1.2.1.2 Global Differential GPS (GDGPS) is a high accuracy GPS augmentation system, developed by the Jet Propulsion Laboratory (JPL) to support the real-time positioning, timing, and orbit determination requirements of the U.S. National Aeronautics and Space Administration (NASA) science missions. Future NASA plans include using the Tracking and Data Relay Satellite System (TDRSS) to disseminate via satellite a real-time differential correction message. This system is referred to as the TDRSS Augmentation Service Satellites (TASS). Ground based augmentation systems utilize use information gathered at specialized fixed well defined instruments to formulate correction messages that are transmitted to a roving receiver – via a communication link (VHF or cell phone) data.

3.3.3.1.2.1.3 The Local Area Augmentation System (LAAS) established by the FAA provides an all-weather aircraft landing system based that real-time, rapid-response monitoring system. The signal received from the GPS constellation is used to calculate the position of the LAAS ground station, which is then compared to its precisely surveyed position. This data is used to formulate a correction message which is transmitted to users via a VHF data link. A receiver on the aircraft uses this information to correct the GPS signals it receives.

3.3.3.1.2.1.4 Nationwide Differential GPS System (NDGPS) The NDGPS is a ground-based augmentation system operated and maintained by the Federal Railroad Administration, U.S. Coast Guard, and Federal Highway Administration, that provides increased accuracy and integrity of the GPS to users on land and water. Modernization efforts include the High Accuracy NDGPS (HA-NDGPS) system, currently under development, to enhance the performance and provide 10 to 15 centimeter accuracy throughout the coverage area. NDGPS is built to international standards, and over 50 countries around the world have implemented similar systems.

3.3.3.1.2.1.5 International GNSS Service (IGS) IGS is a network of over 350 GPS monitoring stations from 200 contributing organizations in 80 countries. Its mission is to provide the highest quality data and products as the standard for Global Navigation Satellite Systems (GNSS) in support of Earth science research, multidisciplinary applications, and education, as well as to facilitate other applications benefiting society. Approximately 100 IGS stations transmit their tracking data within one hour of collection. The focus is on products, available in real-time and in post-process flavors that rely on very tight clock and orbit data. These are available from such sources as Jet Propulsion Laboratory (JPL) and Natural Resources Canada (NR Can), and as services such as Precise Point Positioning (PPP).

3.3.4 Federal Programs for Ground Based Augmentation

The Federal government along with state partners has taken the lead in implementing and supporting a series of services to support the surveying and navigation communities. The following provides a brief description of these services.

3.3.4.1 National Spatial Reference System (NSRS)

The National Spatial Reference System (NSRS) is operated by NOAA's National Geodetic Survey (NGS). It provides a consistent coordinate system that defines latitude, longitude, height, scale, gravity, and orientation throughout the United States. This collection of points (over 1,500,000 of them) forms a network that is used to accurately position other points of interest. Surveyors and others use the NSRS throughout the country to ensure that their positional coordinates are compatible with those determined by others. In this way, when they create maps; mark off property boundaries; and plan, design, and build roads, bridges, and other structures, everything matches up. (http://celebrating200years.noaa.gov/transformations/spatial/side.html)

3.3.4.2 Continuously Operating Reference Stations (CORS)
(http://www.ngs.noaa.gov/CORS/)

Continuously Operating Reference Stations (CORS) provides Global Navigation Satellite System (GNSS) data consisting of carrier phase and code range measurements in support of three dimensional positioning, meteorology, space weather, and geophysical applications throughout the United States, its territories, and a few foreign countries. Surveyors, GIS users, engineers, scientists, and the public at large that collect GPS data can use CORS data to improve the precision of their positions. CORS enhanced post-processed coordinates approach a few centimeters relative to the National Spatial Reference System, both horizontally and vertically. The CORS network is a multi-purpose cooperative endeavor involving government, academic, and private organizations. The sites are independently owned and operated. Each agency shares their data with NGS, and NGS in turn analyzes and distributes the data free of charge. As of May 2010, the CORS network contains over 1,450 stations, contributed by over 200 different organizations, and the network continues to expand. Higgins provides a diagram that illustrates how CORS works (Figure 103).

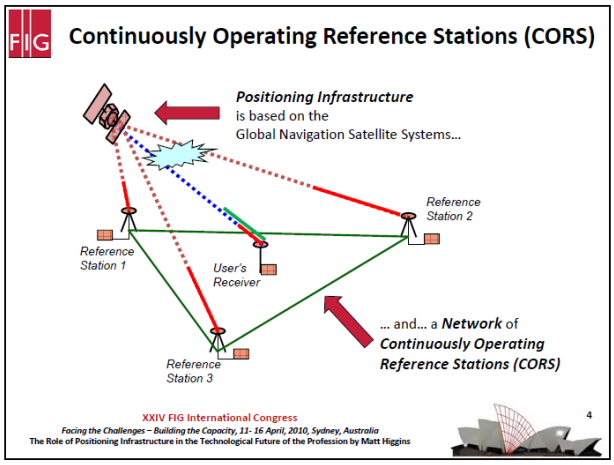


Figure 103 Schematic diagram of CORS (Higgins 2010)

3.3.4.3 The Virtual Reference Network (VRN).

In order to support the highest form of real-time positioning technology surveyor and construction operators use real time kinematic positioning by establishing a virtual reference network Kumar-Mills et al. (2005) provide the following description of the Virtual Reference Network:

> "The use of multiple reference station networks with real-time kinematic (RTK) positioning provides a high precision, centimeter level, satellite positioning service that is extremely reliable and accessible. This technology has gained wide acceptance in the geodetic, engineering, earth moving and public works communities. One proven implementation of the networked RTK is the Virtual Reference Station (VRS) network concept. The VRS concept is widely accepted as the most advanced approach for increased spatial separation of reference stations and error modeling. This system calculates network corrections for systematic errors based on real-time data from all reference stations and simulates a local reference station (or VRS) near a GPS rover station. Corrections for the VRS are transmitted through a communications link. This approach eliminates the need for actual reference stations on site as VRS data can be generated for any location within the network coverage area. The reduction in systematic errors allows increased spatial separation between the reference stations while increasing the reliability of the system and reducing the initialization time."

3.3.5 Global Navigation Satellite System (GNSS)

It is common to refer to the GPS system as part of a larger Global Navigation Satellite Systems (GNSS). The GNSS currently consists of the US NAVSTAR Global Positioning System (GPS) and the Russian GLONASS **GLO**bal **NA**vigation **S**atellite **S**ystem". GLONASS now has 24 of 24 satellites operational.

(http://www.navcen.uscg.gov/pdf/cgsicMeetings/47/%5B21%5D%20GLONASS%20CGSIC%20September%2024%20Fort%20Worth.pdf.)

It is also important to understand whether devices are able to receive and process the signals. According to Wikipedia http://en.wikipedia.org/wiki/GLONASS

> "Septentrio, Topcon, JAVAD, Magellan Navigation, Novatel, Leica Geosystems and Trimble Inc produce GNSS receivers making use of GLONASS. NPO Progress describes a receiver called "GALS-A1" which combines GPS and GLONASS reception. SkyWave Mobile Communications manufactures an Inmarsat-based satellite communications terminal that uses both GLONASS and GPS."

3.3.5.1 An Analysis Of The Prospects And Timeline For Additional Constellations.

During the next decade the GNSS system is scheduled to expand to four global systems. Higgins (2010) provides the following schedule:

- USA: Global Positioning System (GPS) - Now
- Russian Federation: GLONASS – during 2010
- European Union: Galileo – 5 to 10 years
- China: Compass – 5 to 10 years
- Plus at least 2 Regional Systems:
  - India: Regional Navigation Satellite System (IRNSS)
  - Japan: Quasi-Zenith Satellite System (QZSS)

3.3.6 Assisted (Versus Augmented) GNSS based on Communication

During his presentation Higgins (2010) also noted that advancements by companies such as Nokia in the mobile communication field are also going to have profound impacts on how positions are determined. He "declared that precise positioning is coming to your mobile phone" Applications that require knowing where users are in real time will drive the fastest technology changes.

In 2000 the Federal Communications Commission issued an order for an enhanced 911 system that would be able to provide emergency assistance to mobile telephone user. The requirement for Phase II of Enhanced 9-1-1 included this mandate:

> "The FCC requires wireless carriers, within six months of a valid request by a PSAP, to begin providing information that is more precise to PSAPs, specifically, the latitude and longitude of the caller. This information must meet FCC accuracy standards, generally to within 50 to 300 meters, depending on the type of technology used."

The FCC's directive influenced the consumer GPS market and may have created the LBS consumer industry. Feuerstein (2010) believes that the big E-911 buzz is about emerging hybrid systems that blend today's handset-based GPS with high accuracy network-based technologies to drive location performance indoors, where satellite systems don't work well, and where the majority of wireless calls are made.

At an extreme Skyhook determines a location simply based on the location of cellular towers and Wi-Fi hotspots – and does not use GPS. The following explanation is useful.

> "To pinpoint location, the Core Engine uses a massive reference network comprised of the known locations of over 100 million Wi-Fi access points and cellular towers. To develop this database, Skyhook has deployed drivers to survey every single street, highway, and alley in tens of thousands of cities and towns worldwide, scanning for Wi-Fi access points and cell towers plotting their precise geographic locations. Skyhook's extensive coverage area includes most major metro areas in North America, Europe, Asia, and Australia."

In order to create this network Skyhook employs its own version of mobile field data collection that is worth noting. (http://www.skyhookwireless.com)

This capability presents an interesting alternative for some parts of the nation. For example the map of Washington DC coverage is indicates a direct and logical association with population density (Figure 104).

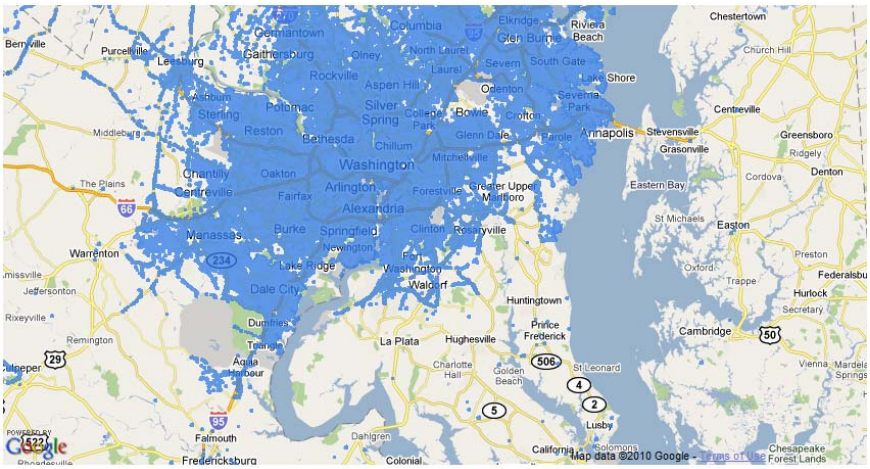


Figure 104 Map of Skyhook coverage for the Washington DC area (Skyhook.com)

3.3.7 Prospects for Future Developments from Nokia and other Cell Phone Manufacturers about Autonomous Cell Phone Positioning.

Many experts believe that Nokia will be the leader in the convergence of positioning technologies. According to Berg Insights (Malm, 2010):

> "Nokia was the first major vendor to declare a strategy for GPS-enabled handsets and location services. The company sees location and maps as a key enabler for numerous experiences and has decided to include free navigation services for all its GPS-enabled smart phones starting in March 2010. The other major handset vendors have shown less assertive strategies but are nevertheless expanding their portfolios of GPS-enabled handsets rapidly."

Also Nokia's business strategy has included a vertical integration approach that includes the acquisition of NAVTEQ in 2007 for $8.1 billion. Wirola (2008) in *High-accuracy positioning for the mass market* provided a glimpse of Nokia's plans that may soon result in centimeter level accuracies.

- The goal is a single, unique interface for providing handsets with assistance information over IP-networks.
- Up to 50% of Nokia's handsets could be equipped with AGPS in 2010 to 2012.
- AGNSS (Assisted GNSS) architecture will include a server either in a public internet or cellular network. The server will distribute data originating from GNSS receivers and external services (extended ephemerides etc.). This information will be distributed over a control channel of the cellular network or the User plane (IP-network)
- AGNSS standardized in 3GPP, 3GPP2 and OMA. The architecture opens the way for advanced assistance data and positioning methods. This will evolve with current standards include basic assistance data including broadcast ephemerides for GPS and Galileo.
- Compared to normal methods assisted GNSS techniques can improve user experience in terms of speed of position fix and accuracy. Assisted GNSS obtains position fix typically in 10-20 seconds.
- Real-Time Kinematics in mobile landscape will be supported by proposed standards enable high-accuracy relative positioning between two handsets OR between a handset and an assistance server.

Indications are that Nokia has patents for the type of Assisted GNSS described above. Wirola provides a schematic of how such a system would work (Figure 105)

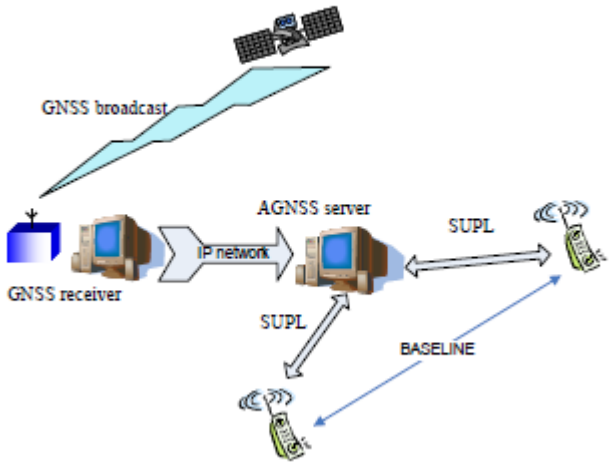


Figure 105 Nokia's schematic diagram for use of Supplemental User Plane Location (Wirola 2008)

3.3.8 Summary

The GSS Initiative will benefit from the incredible advancements in positioning technology that are rapidly emerging. With the convergence of an expanded GPS network and cellular technology an increasingly expanded variety of devices will be capable of recording the location of relevant features with accuracies that meet MTdb criteria. As a result of enlightened public policies, regulations, standards and consumer demand, all location oriented activities will be supported by a ubiquitous and basically free infrastructure. Evidence of the remarkable changes that are underway was included in our report *Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data* where we reported that scientists at the USGS have adopted the Apple iPad as their field data collection device and report a high level of satisfaction with the quality of the coordinates they are recording. The title of Dru Smith's (National Geodetic Survey) recent paper is a fitting conclusion *The Future of Global Navigation Satellite Systems: Anyone, Anywhere, Anytime, Any Accuracy*.

3.4 Identification of Approaches to Finding Candidate Areas for Targeted Address Canvassing

3.4.1 Proposed Model

The goal of this portion of the report is to provide the Geography Division with information regarding approaches to GIS based data mining and modeling that would help it select specific parts of the country for targeted address canvassing. The objective is to explore cost effective and systematic procedures that can aid the Geography Division in its efforts to evaluate the completeness and accuracy of the current MTdb as well as to discover additional features that need to be added or deleted. While the focus of this section is "targeted address canvassing" we believe that term does not reflect the true nature of the process. Any approaches to improve the MTdb are inherently nationwide. The goal is to quickly discover specific areas where there has been an "alert" that something has changed. The analogy to a battle field is appropriate where spatially explicit strategies are developed and deployed based on reconnaissance. This reconnaissance is obtained by monitoring activities. As noted elsewhere in this report, change detection is monitoring – whether it is conducted by an intelligence analyst, a radiologist or a Census Bureau employee. Therefore, our recommended approach to "targeted address canvass" is to establish a spatially explicit nationwide platform that can be constantly monitored. In this environment the entire "battlefield "is being observed but resources are directed where they are needed.

Data fusion methods coupled with data mining tools could constitute an executive dashboard that serves as a spatial decision support system. This "dashboard" could be a set of visualization tools that enable mangers to justify allocation and specification of resources. As noted in our deliverable *Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data*, there are now a host of assessable geospatial data resources. The challenge is to discover and utilize these resources effectively. Areas that need attention include new subdivisions, multiuse structures that may contain residential units, and areas where there may be significant over coverage of addresses. The overriding objective is to maintain an accurate MTdb while

minimizing the number of times that an employee is sent into the field. Just as in a battlefield sending personnel into the field is expensive and potentially dangerous.

In practice, the proposed system could start with a simple web based mapping system that provides visualization tools to “drilldown” to more detailed data. Analogies are the CDC West Nile Virus map (Figure 106).There are also several other census tract level systems that are prototypes for such a system. For example, Policy Map (http://www.policymap.com/) is used by HUD to map housing conditions (Figure 107). The New York Times interactive foreclosure mapping system is a particularly intriguing example (Figure 108). This Flex based API allows a user to start with the entire three state New York Metropolitan area and then drill down to the street level for any tract in the region. It even includes specific geocoded address points for foreclosed property.

Figure 106 Centers for Disease Control West Nile Virus Map (CDC)

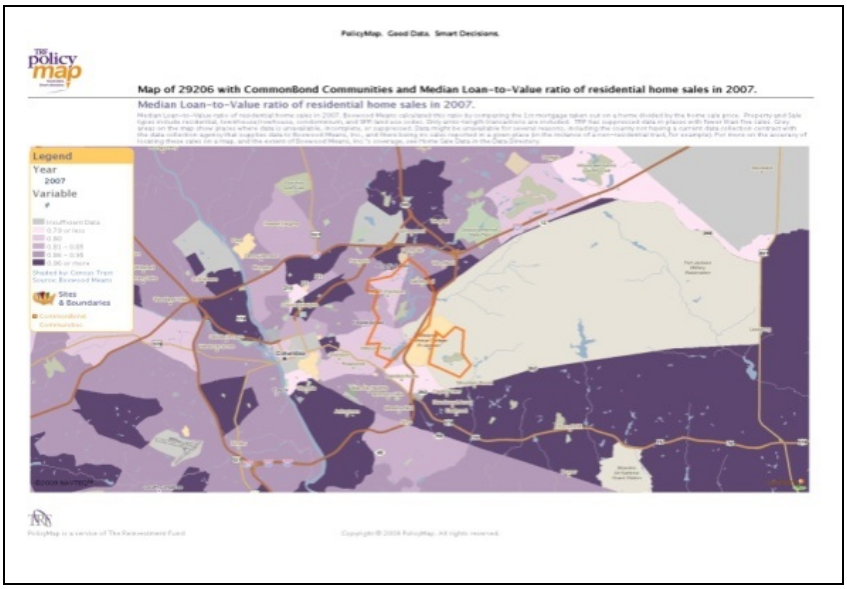


Figure 107 Example of Policy Map for Columbia, SC (Policy Map .com)

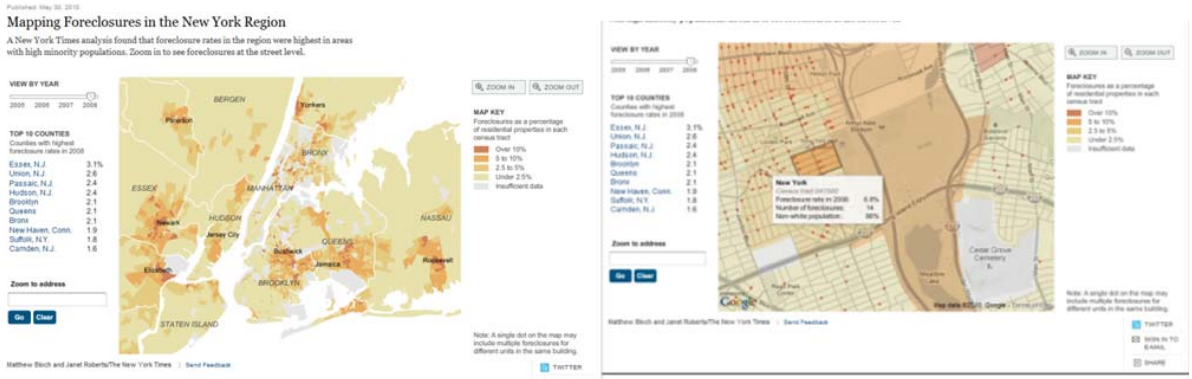


Figure 108 New York Times interactive foreclosure mapping system (//www.nytimes.com/interactive/2009/05/15/nyregion/0515-foreclose.html?ref=foreclosures)

In the current geospatially enabled environment almost every data source (new addresses, structures, roads, tax parcels, pixels address points, and other spatial data) is available in a digital GIS form defined by geographic coordinates or can be geocoded to obtain coordinates. This means that an infinite range of data can be harvested and aggregated at any level of census geography. Such a database of indicators of potential problem areas would form the basis for the decision support system. The full implementation would involve the following steps:

1. Selection of geographic monitoring units.

   The nested hierarchical system of Census geographic area is ideally suited for this. Tracts would be a logical starting point. However, data elements and appropriate metadata should be collected at the finest level of spatial resolution (point, pixel, parcel, road segment etc). This would support a full range of aggregation disaggregation (drill down) capability.

2. Selection of indicators. The following is a list of indicators that could be assembled:

- Internal
  - MTdb
    - Street Segments added between 2000 and 2010
    - Blocks added between 2000 and 2010
    - MSPs
  - 2000 / 2010 Census
  - Undeliverable as Addressed (UAA)
  - Change in number of housing units
  - Density of housing units
  - Etc.
- ACS
  - M2501 Percent of Housing units that are Mobile homes or trailers
  - M2502 Percent of Housing units that were built in 2005 or later
  - M2503 Percent of housing units that were built in 1939 or later
  - M2504 Percent of Occupied housing units that were moved into in 2005 or later
- USPS (Partnership)
  - DSF
  - Zip + 4 etc.
- Federal
  - Census/HUD new building starts
  - HUD/ USPS vacancy
  - HMDA Data
- External
  - Imagery
    - MDA Urban Change
  - Roads
    - Commercial
    - Public
    - Crowdsourced
- Dwellings
  - Address lists
    - Commercial
    - Public Partners
      - Building permits
  - Structures
    - Building footprints
    - Address points
      - Open Address.org
  - Parcels
    - Commercial
    - Public Partners (counties or states)
  - Demographic and housing data
    - Commercial
      - GeoLytics, CACI etc.

3. Build a Decision Support environment with analytical and visualization functions
    - Binary – Statistical determination of thresholds for alert
    - Data mining and visualization to discover associations and patterns
        - Predictors of dependent variables (e.g. M2502 Percent of Housing units that were built in 2005 or later)
4. Decision Making
    - Implementation of criteria
    - Examination of drill down resources
        - Free commercial Web Services
            - Google, Bing, Zillow etc
            - Esri Community Base maps
        - Public Web Services
            - Indiana Map
            - Virtual Charlotte
        - Licensed Commercial Web Services
            - Parcels
            - Image services (Pictometry etc.)
    - Allocation of resources
        - Non field
        - Field

3.4.1.1 Discussion of Components of the Decision Support System

3.4.1.1.1 Level of Geographical Analysis

The approximately 66,000 census tracts would appear to be the most logical unit at which to monitor change. In fact, Harris (2008) developed a prototype census tract level “prioritization of areas for change detection” based on tracts (Figure 109). Their prototype was based on trends in population growth, adjacency to areas of population growth and the robustness of the information available from the USPS DSF. These factors were conceived to omit Census Tracts (primarily in the urban/suburban areas) where existing methods are available to identify structure changes as well as those with the highest probability of rural/non-urban structure change.

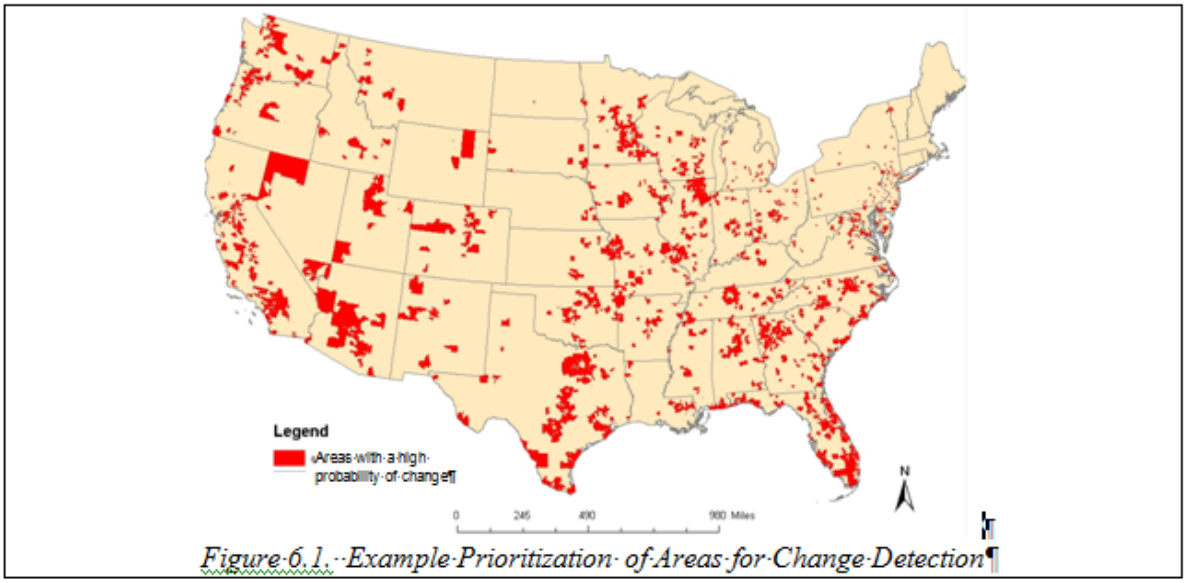


Figure 109 Census tract level map of prioritized areas for change detection (Harris 2008)

As noted in the Policy Map and New York times examples, it is possible to develop robust APIs at the tract level. Furthermore, the tract is the unit of analysis employed by the HUD / USPS vacancy data (Figure 110) that monitors the following data elements

Total Number of Addresses - This reflects all addresses (residential and commercial) that USPS has recorded in their database.

Total Vacant Addresses - These are addresses that delivery staff on urban routes have identified as being vacant (not collecting their mail) for 90 days or longer.

Total No-Stat Addresses - There are many reasons an address can be classified as No-Stat, including:

- Rural Route addresses vacant for 90 days or longer
- Addresses for businesses or homes under construction and not yet occupied
- Addresses in urban areas identified by a carrier as not likely to be active for some time (http://www.huduser.org/portal/datasets/usps.html)

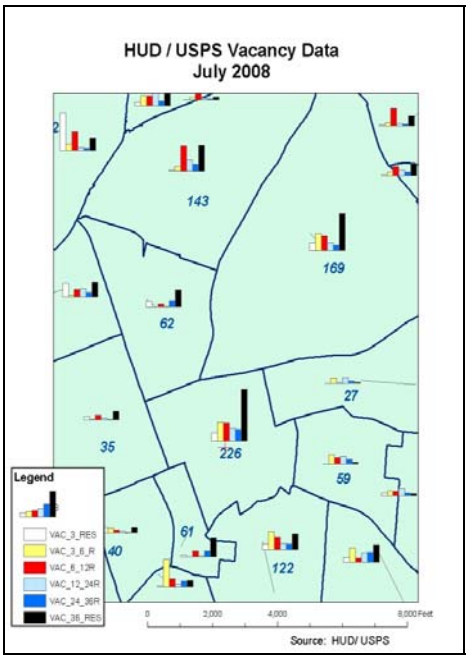


Figure 110 HUD / USPS tract level vacancy data (Cowen unpublished)

Tracts are also the reporting unit for the Home Mortgage Disclosure Act (HMDA) data that monitors mortgage activities. These data are compiled for urban areas and are available on line through the Federal Financial Institution Examination Council http://www.ffiec.gov/. These data provide an indication of the mortgage activity within a tract along with some census variables (Figure 111). The HMDA data provide an excellent source of annual property based activity with a tract. It may be assumed that the number or proportion of mortgage applications is a good indicator of stability or change within a tract.

TRACT	112.02
Count	254
Minimum Loan Amount	$5,000
Maximum Loan Amount	$362,000
Average Loan Amount	$136,772
Standard Deviation of Laon Amount	$75,490
Population	3544
% Minority	14.39
MSA Median Family Income	$58,200
Number of Owner Occupied Housing Units	1259
Number of 1 - 4 Family Units	1563
Minimum Applicant Income	$14,000
Maximum Applicant Income	$361,000
Average Applicant Income	$84,781
Tract to MSA Median Income Percentage	112.42

Figure 111 HMDA data elements for a census tract in Columbia, SC (Data from ffiec.gov)

3.4.1.2 Data Elements

The assemblage of indicators at the tract level should start by harvesting internal data from the MTdb. This could include measures of the number and quality of structure points and road segments. It could also include population and housing tabulations for 2010 and for measuring changes since 2000. Technically, these data could be harvested from Oracle tables, such as a feature accuracy table, a MTFCC or road classification accuracy table, a MAF table or housing indicator table. Tracts could be flagged if they contain unique land uses such as a federal prison, a university, parks, national forest, etc. An interesting exercise would be a forensic examination of census questionnaires that were Undeliverable as Addressed (UAA). At a minimum, they would have a five digit zip code for geocoding. This UAA data would provide a potential indicator of areas where the address information in the MAF is erroneous and highlight the most problematic tract. Tracking these addresses and their associated history file could reveal sources of address information that are less reliable in specific areas than other sources. Further, longitudinal analysis of the address histories may reveal sources that are no longer as relevant as in prior years.

Density measures would also provide a simple starting point for estimating the probability of change. Simple tract level density measures could be generated for 2000, 2010 and the changes. This would reveal areas of relative activity. As the ACS develops it should be possible to tabulate ongoing housing changes. Other potential indicators include commercial data from vendors such as CACI or GeoLytics that include current estimates and five year projections. In a similar fashion, the Department of Homeland Security's Landsat based National Urban Change Indicator (NUCI) would provide a measure of land cover change (Figure 112) that could be summarized at the tract level (e.g. proportion of land area converted).

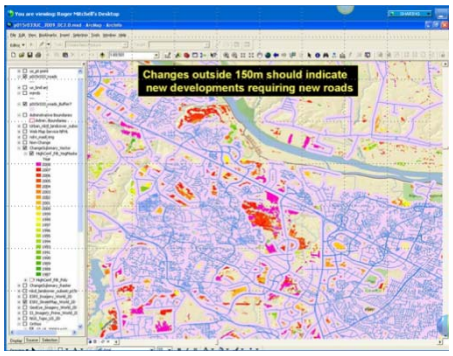


Figure 112 MDA correlated change data (Mitchell 2010)

3.4 2 Public Sources of Roads

As part of E911 programs most local governments maintain a current set of street centerlines. Many of these were utilized by Harris during the MAF/ TIGER Accuracy Improvement Program Many states such as New Mexico are coordinating this information into a central repository to support emergency response. The South Carolina GIS Council's status map (Figure 113) identifies the current update cycle for the authoritative source for roads and also identifies those counties where little change has occurred. As noted elsewhere, data from these files can be used

to automatically discover and separate new roads that can be summarized at the tract level (Figure 114).

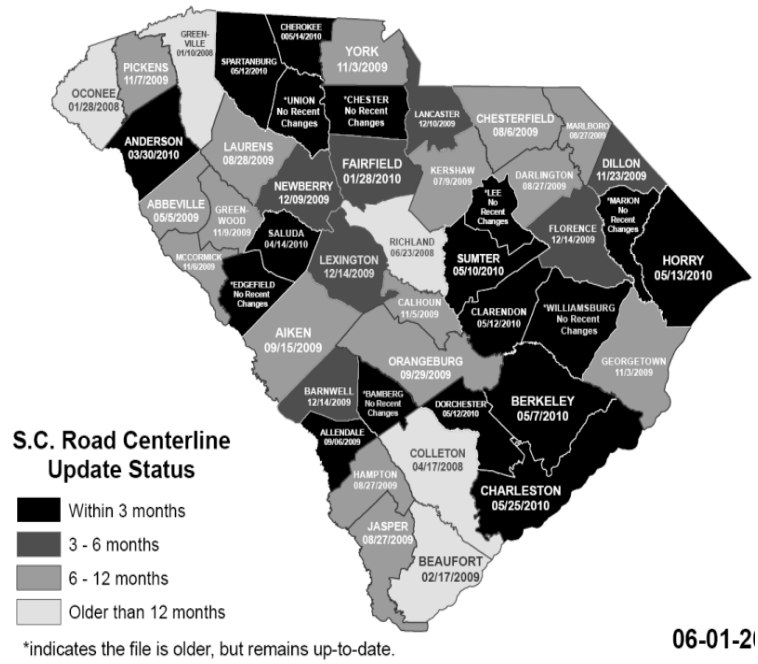


Figure 113 South Carolina road centerline status map (Courtesy of DeTroye)

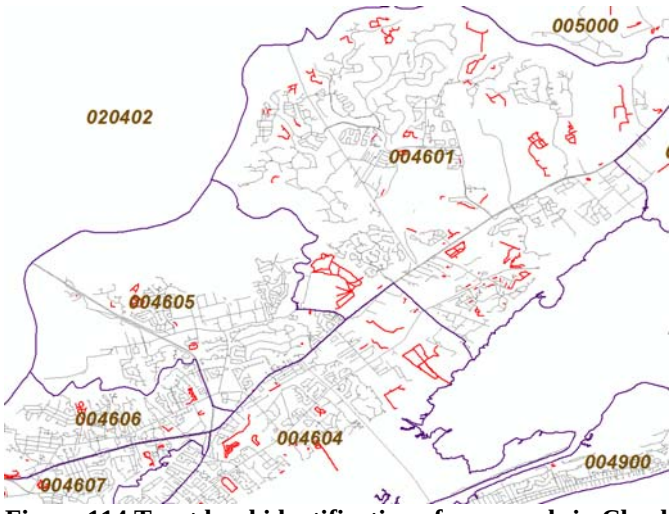


Figure 114 Tract level identification of new roads in Charleston County, SC (Cowen unpublished)

3.4.2.1 Crowd Sourced Streets

As discussed in section 3.5 of this report (as well as in *Reporting the State and Anticipated Future Directions of Addresses and Addressing*), crowdsourcing is being effectively utilized by Open Street Map as well as commercial firms to update road databases. In a manner similar to the previous example, roads provided by crowdsourcing could be tabulated. For example, passive crowdsourcing or probe-based procedures could reveal where vehicles are travelling streets not shown in the current version of TIGER. While not being an authoritative source,

crowdsourcing is an effective, though not comprehensive, indicator of spatial change (Figures 115 and 116). In Figure 115 we show an example of passive community input showing the delineation of a new street and active community input indicating the correct street name and the address of the house on this "new" street. Areas in which GPS traces were recorded for streets not found in TIGER would presumably be candidates for further field examination. In addition, large clusters of streets that were missing from TIGER could be used to highlight areas where compilation sources, including imagery, were either incomplete or out-of-date.

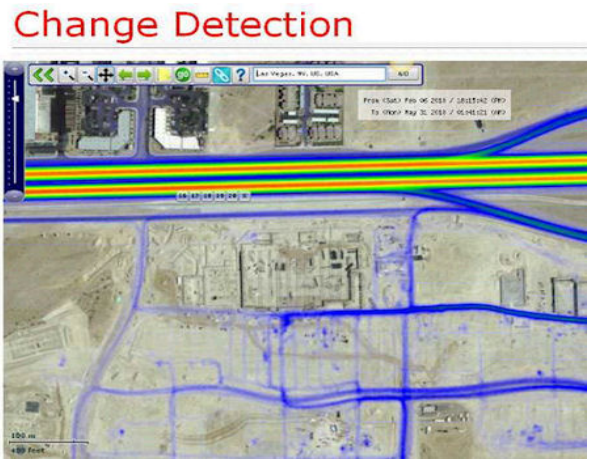


Figure 115 The image is overlaid with GPS traces from TomTom Personal Navigation Devices. Of note is the area to the south of the major highway that was not yet in the Tele Atlas database. (Image courtesy of Tele Atlas.)

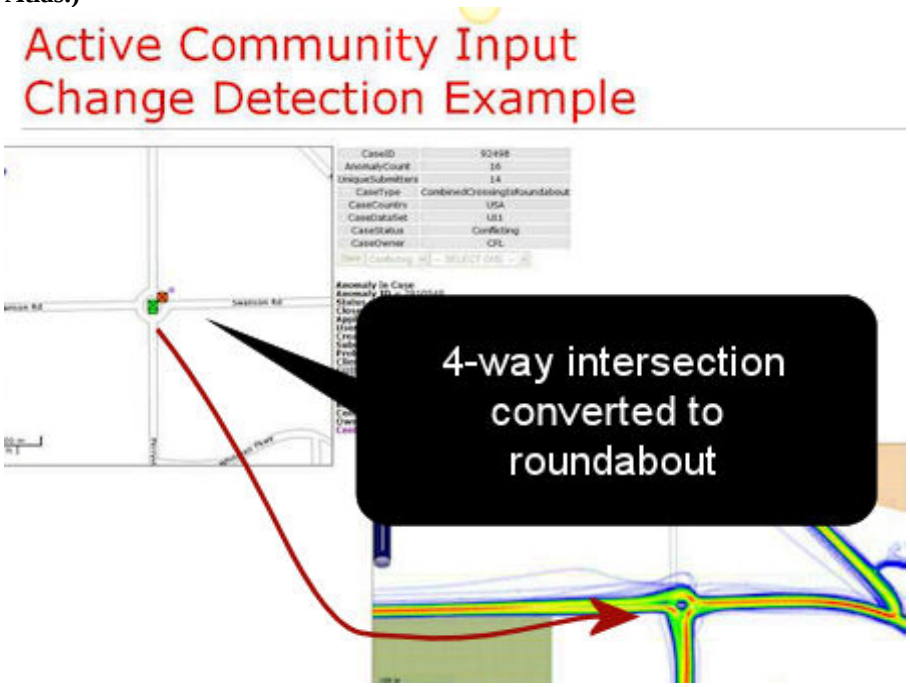


Figure 116 Shown is an example of both active and passive crowdsourcing. Through MapShare, TomTom received a number of map corrections reporting that an intersection shown as a four-way in the Tele Atlas database was now a roundabout. At the bottom of the image are the GPS traces of vehicles equipped with TomTom PNDs, navigating a roundabout, as shown by the flow and the circular median in the center to the roundabout. (Image courtesy of Tele Atlas.)

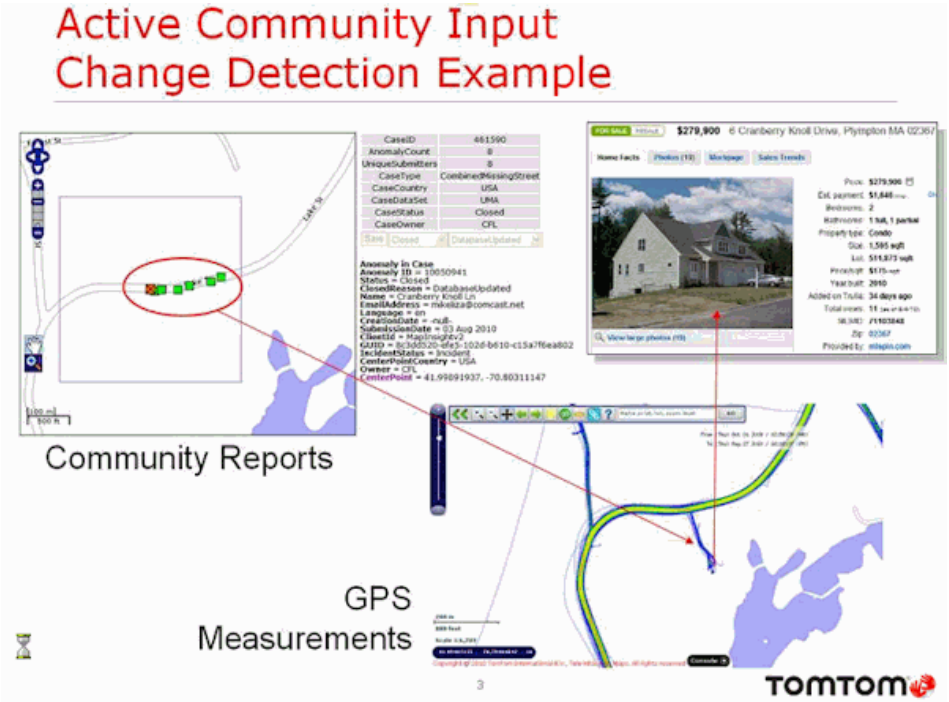


Figure 117 In this example TomTom users actively reported a missing street, as well as the street name and the correct address of the house on the street. In addition, GPS tracks showed that a street existed, but little other detail. (Image courtesy of Tele Atlas.)

3.4.3 Address Lists

As noted in our report *Reporting the State and Anticipated Future Directions of Addresses and Addressing* there are many sources of address data. In addition to the USPS Delivery Sequence File, address lists are maintained by municipalities, counties, regional governments or at central state offices. The process of address creation varies dramatically among local governments. For example, the Loudon County, TN GIS department provides a good example of the submission and approval process (http://www.loudoncounty911.org/mapping.php).

In Deliverable 2 on Addresses and Addressing, we noted that local government agencies should be the most authoritative source of addressing information and the LUCA program is a reflection of this authority. We understand that the LUCA program is difficult to administer, but creating a program that had to the goal of polling local sources of address information more than once a decade could be a powerful tool for improving the accuracy and currentness of addresses in the MAF. While the current LUCA procedures involve maps, mapping software and a number of complex interactions involving the security of address data, it is possible that less aggressive systems could be developed to provide useful information with less effort.

For example, the Geography Division could set up a secure website to allow local governments to view the address information that is currently in the MTdb. The local governments, through a secure, controlled access Wiki, would be allowed to enter, for areas under their jurisdiction, comments about addresses that appeared to be missing or circle areas on the display where

addresses are believed to exist but are not represented in the MTdb.[3] While such a system is clearly not as powerful as the LUCA software, it might not have the usability problems associated with LUCA. In essence local source could be used to point to areas where the MTdb might need updating, or areas that might be candidates for address canvassing.

The coordination of these address lists at state repositories would greatly improve the current practices. For example, the state of South Carolina GIS Coordinator reports that by using the FME Safe Software interoperability tools he was able to easily standardize the addresses for forty one counties in just a few days. As noted in our *report Reporting the State and Anticipated Future Directions of Addresses and Addressing* at least twenty three states are currently maintaining state wide address lists. County comparisons of these tables with the current MAF should be a simple batch process. Other interesting address sources are building permits (Figure 118) that would provide an indicator of future change at the tract level. These addresses could be flagged for future updates.

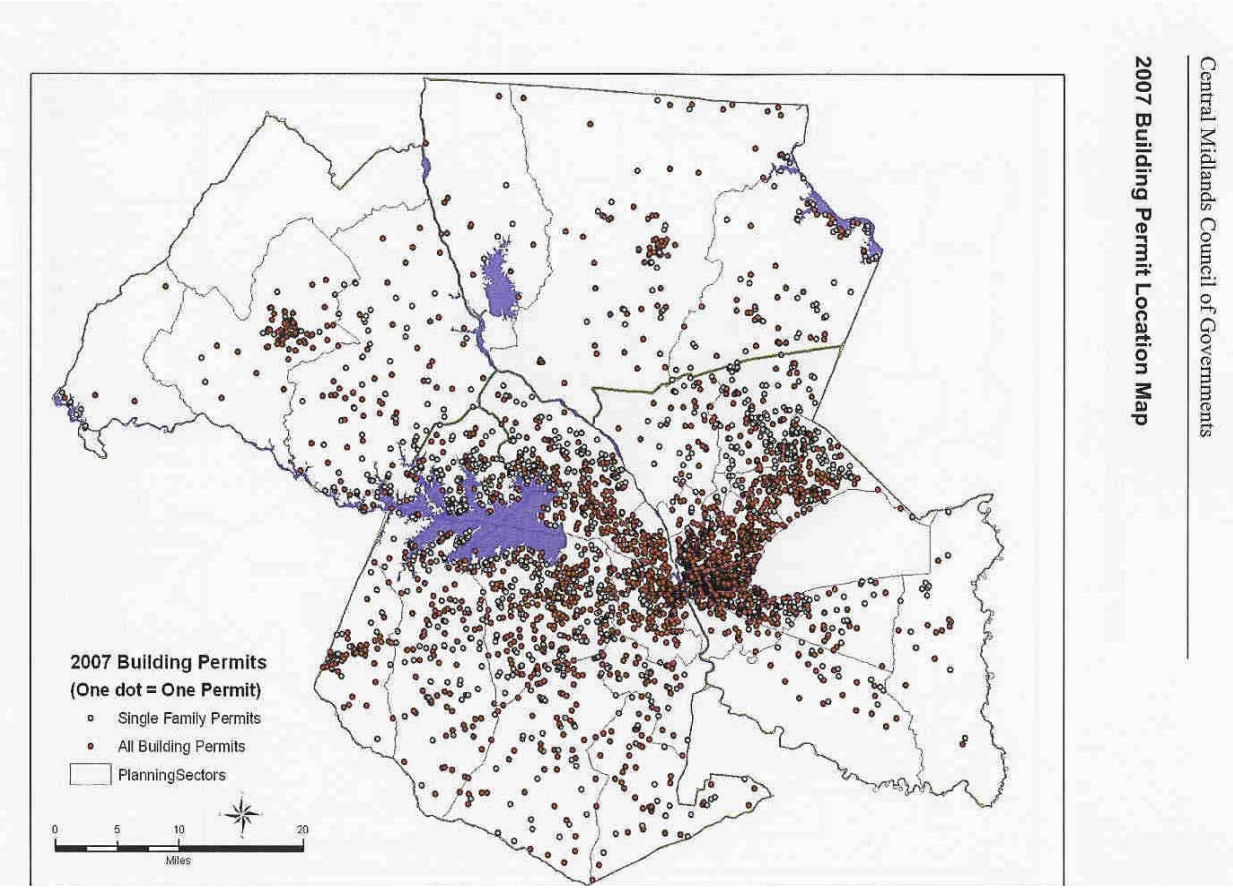


Figure 118 Map of building permits in four county area of South Carolina (Central Midlands Council of Governments)

[3] The Wiki approach would need to support the confidentiality required by Title 13.

3.4.3.1 Use of Mass Mailers

Most mass mailers use CASS[4], the DSF2[5] and a host of programs available from the USPS on a monthly basis to ensure that their address lists are as current as possible because qualifying their addresses software with the USPS allows them to take advantage of the lowest postal rates available. It is our suspicion that the quality of the mailing lists of mass mailers benefits from this constant application of addressing "hygiene" tools. We recommend that the Geography Division adopt USPS tools that might help improve the address hygiene of the MAF. An alternative strategy, if allowed under Title 13, would be to periodically submit the mailable addresses in the MAF (or a suitable filtered extract) to a mass mailing house to test the addresses for "deliverability". Mapping the results of these tests should reveal areas in which the addresses in the MAF appear less reliable than other areas in respect to deliverability. In addition to helping pinpoint the reason that the addresses are undeliverable, the analysis could serve to identify areas where address accuracy is less than desired and these areas could be viable candidates for address canvassing. Finally, it might be possible to conduct block level count reviews between the MAF and the commercial lists and use the results as a measure for change detection.

3.4.4 Temporal Analysis of DSF Data

In a similar fashion, to that mentioned above the DSF could be monitored to see how the "geography of address accuracy" evolves. Each iteration would provide a measure of change. The rate/number of changes would change the status in the Decision Support System. It important to note that this process would constitute a complete (rather than targeted) address canvass. In the DSF provided to the Census Bureau there are fields that identify:

- Delivery ZIP Code
- Delivery Carrier Route ID
- Delivery Point Type Code (distinguishing residential and business addresses)
- ZIP Add On Records (ZIP+4)

To initiate a temporal analysis of the DSFs, an initial baseline dataset would be created by selecting the residential records, then sorting and tabulating the data by ZIP, Carrier Route, and ZIP+4 values. This tabulation process could be repeated for each subsequent DSF provided to the Census Bureau. We suggest that the currentness and robustness of this analysis would be

[4] CASS is available to all mailers, service bureaus, and software vendors who want to evaluate their address-matching software and improve the quality of their ZIP + 4, carrier route, and 5-digit coding accuracy. The USPS® National Customer Support Center administers and grades the mailers' tests and then returns the results so mailers can diagnose and correct deficiencies. However, CASS does not measure the accuracy of ZIP + 4 delivery point, 5-digit ZIP, or carrier route codes in a mailer's existing files. CASS enables mailers to measure and diagnose internally written, commercially-available address-matching software packages. The effectiveness of service bureaus' matching software can also be measured.

[5] Featuring over 165 million addresses, the DSF2 claims to contain every deliverable address in the United States and is the most accurate, -up-to-date address data available to commercial firms from the USPS. The DSF2 is used to identify deliverable and undeliverable postal addresses, business versus residential addresses, commercial mail receiving agencies, seasonal address, addresses vacant for over 90 days and address delivery types such as curb, door slot, box, etc, The DSF2 includes delivery point validation, postal coding and delivery/walk sequence numbers.

enhanced by getting monthly DSF updates. Once a number of datasets have been acquired, more sophisticated temporal analysis procedures can be used on the data to detect trends. Each of these tabulation units corresponds to a finer level of geography.

The USPS does not delineate boundaries (i.e. polygons) for these units, but rather denotes segments of the road network that make up the ZIP code and other postal service areas. Commercial firms (such as Maponics and Datasheer) build ZIP code polygons from the individual addresses and align the service area boundaries with streets. There are about 40,000 ZIP code areas within the United States, covering tens to hundreds of square miles. Figure 119 shows a ZIP code area in Syracuse that covers all or parts of seven Census tracts. Approximately 300,000 Carrier Routes nest within a ZIP code and traverse the road network in a way that can be circumscribed with polygonal areas. An example is shown in Figure 120. The approximately 40 million ZIP+4 service areas are difficult to symbolize using a polygon map (since each side of a street might have a different ZIP+4 code), so these service areas are sometimes located using a geographic average of the locations covered by the ZIP+4 identifier (see Figure 121 and 122).

The geographic specificity provided by the smaller postal service areas needs to be balanced against the increased amount of “noise” that will be introduced into the data caused by the frequent changes made to the system by USPS. Changes in a tally based on Carrier Route or ZIP+4 could be the result of reallocation of mail delivery (route expansion or consolidation) rather than additions caused by new houses or deletions by demolitions.

It is recommended that this system be tested on DSF data for a rapidly growing state with a mature geospatial infrastructure, thus allowing for validation of the test results (e.g. North or South Carolina). The test would help to determine the effectiveness of this method for detecting areas of change and improve our understanding of the geographic sensitivity of the system to “false” changes introduced by USPS operations.

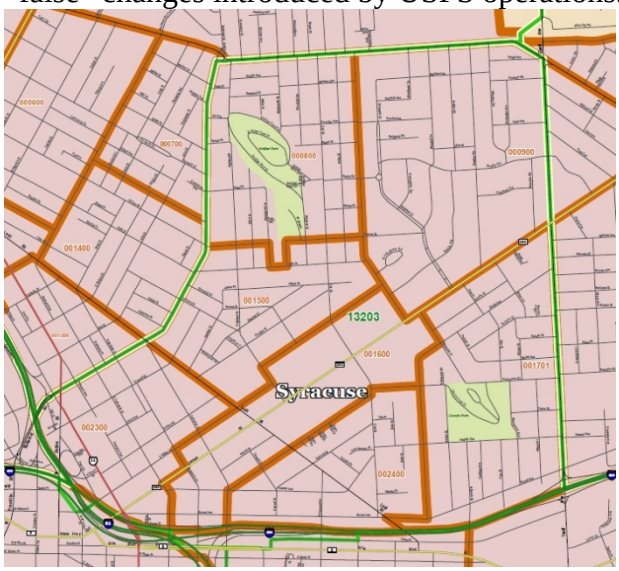


Figure 119 ZIP Code area (delineated in green) shown with Census Tracts (orange) and roads (Maponics.com)

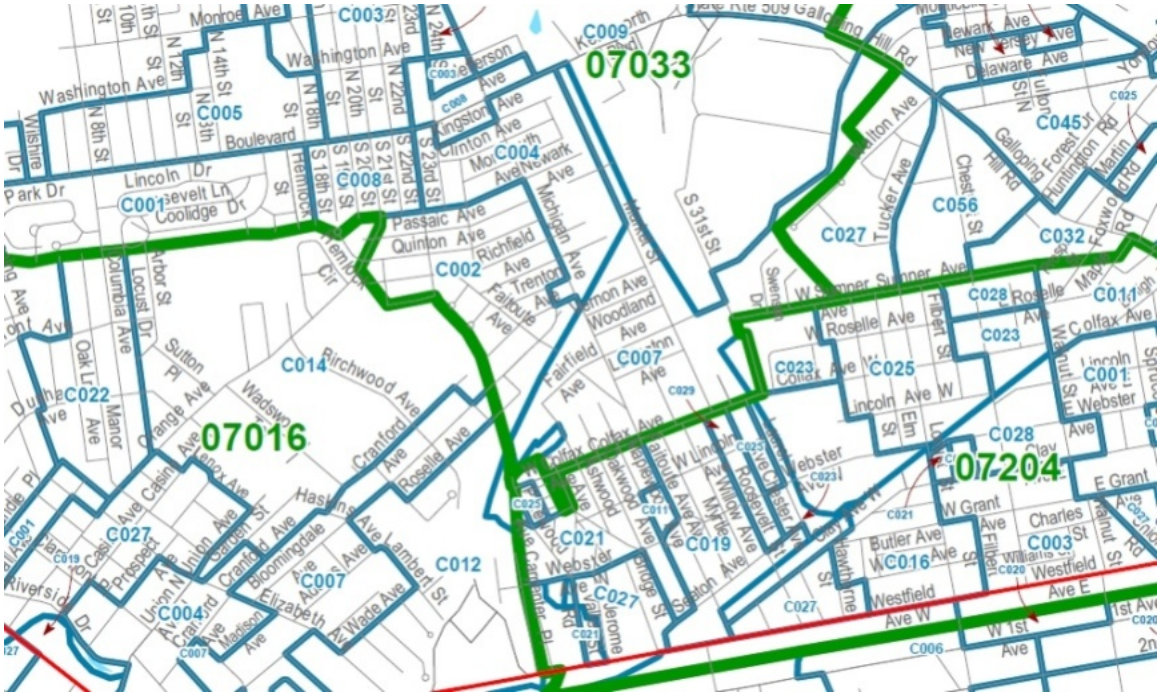


Figure 120 Carrier Routes (outlined in blue) shown with ZIP Code Areas (green) and roads (Maponics.com)

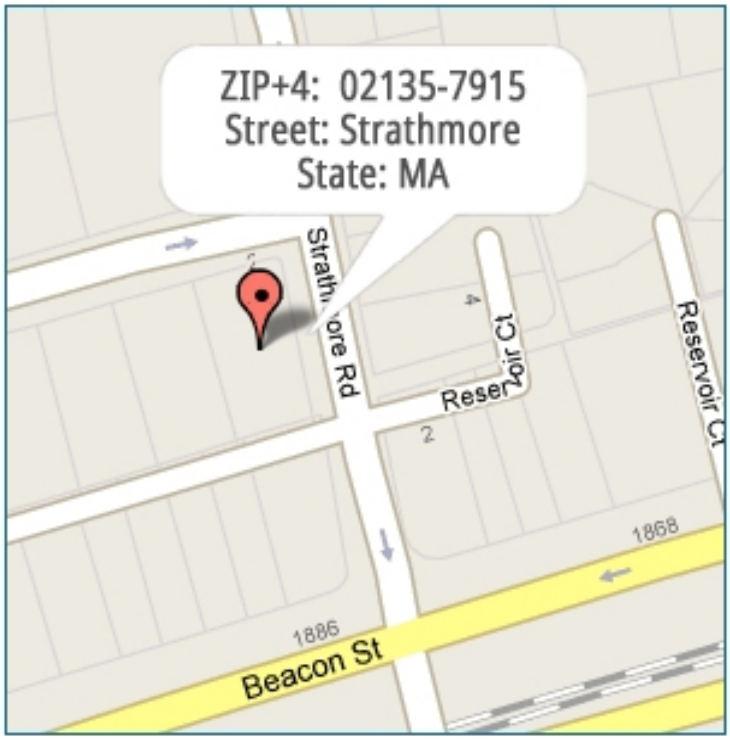


Figure 121 Location of a ZIP+4 centroid along a street segment (Maponics.com)

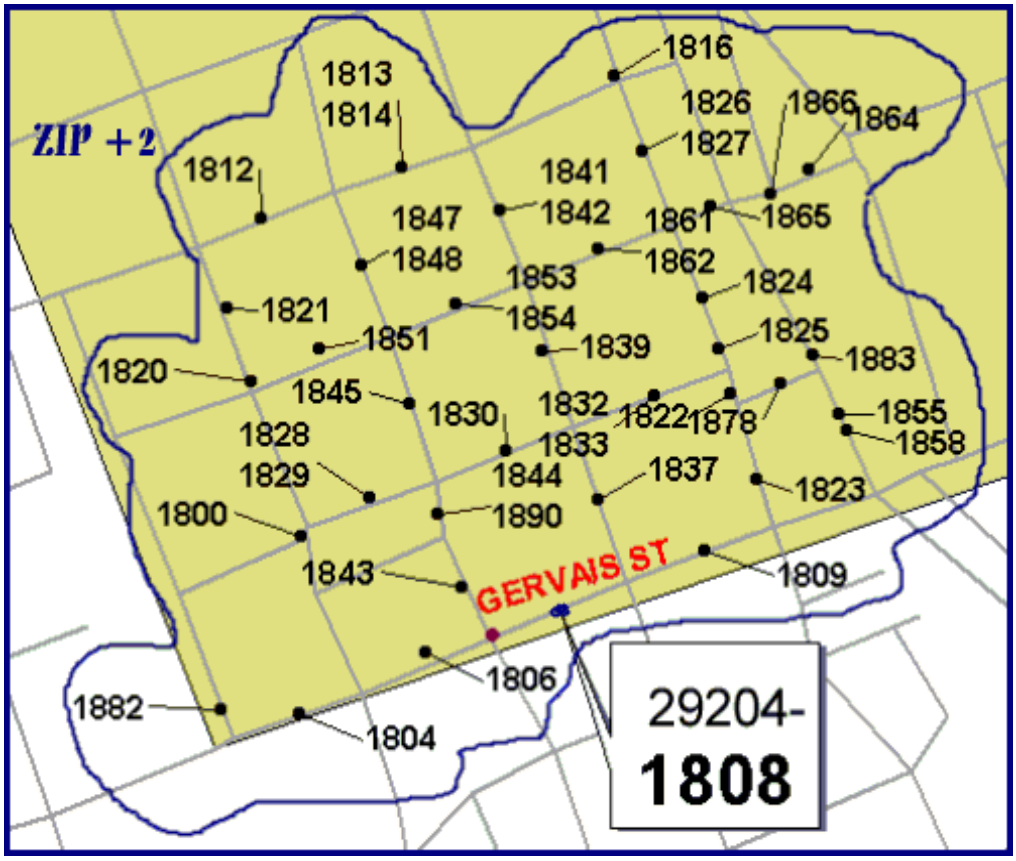


Figure 122 Example of zip +4 points for zip code 29204-1808 (Cowen unpublished)

3.4.5 Parcel Data

Tax parcels represent the most appropriate geographical unit for studying the use, value and ownership of land. These parcels can be assembled within a tract and data from the Nome Mortgage Disclosure Act (HMDA) can be assembled to monitor mortgage activity. Furthermore, any local government program that monitors the addition of new parcels and the split of existing parcels would be an excellent source of information for targeted address canvassing. However this would be very difficult to measure on a national scale. Nevertheless, it is important to note that following the 2007 The National Research Council (NRC) report *National Land Parcel Data: A Vision For The Future*, there has been a great deal of interest in coordinating parcel data at the federal level. The recommendations of that report have been endorsed by several groups including the National Geospatial Advisory Committee (NGAC) and the National States Geographic Information Council (NSGIC). Furthermore, the 2009 The Federal Geographic Data Committee (FGDC) Annual Report focused on parcel data and Karen Siderelis, the interim chair of the FGDC stated:

> "Land parcel data combined with other geographic information are essential to such functions as the management of emergency situations, development of domestic energy resources, management of private and public lands, support of business activities, and monitoring of regulatory compliance. The feature story of this year's report underscores

> the need for a coordinated system of land parcel information across the country." (FGDC, 2009)

While there is no federal coordination of parcel data, several federal agencies including the FBI are licensing parcel data from Core Logic, and HUD is conducting a year-long study to evaluate its potential role as the coordinator of parcel data. Furthermore, the recent financial regulatory act may require lending institutions to report parcel identification numbers as part of HMDA. Such a parcel level database of mortgage activity would provide an excellent indicator of change for the GSS Initiative.

3.4.6 Decision Support System - Analytical functions

A tract level assemblage of various indicators would provide a robust starting point for implementing the GSS Initiative. Armed with the correct tools these data could serve as the basis for decisions about allocation of resources to validate or update the MTDB. The primary goal would be to use a series of metrics to classify tracts into different "bins". These bins would establish the priority for review. Working with Bureau statisticians the Geography Division could establish criteria and thresholds on different measures (accuracy of addresses, completeness of road segments, likelihood of new address etc.). In all cases there would be documented procedures and criteria to justify the decisions. Again, it should be emphasized that these data and procedures would be implemented across the nation. It is the outcomes that are targeted.

This tract level data base would also be ideally suited for spatial data mining and visualization. The GeoVISTA Center**,** at Penn State University has developed an excellent set of open source tools (GeoVIZ) (Figure 123) for this type of task. Within this type of environment various exploratory data analysis tools could be used to test relationships and identify patterns. For example, regression procedures could be deployed to identify independent variables that explain variations in mortgage activities or change in number of housing units. Visualization tools would then highlight tracts with large residuals.

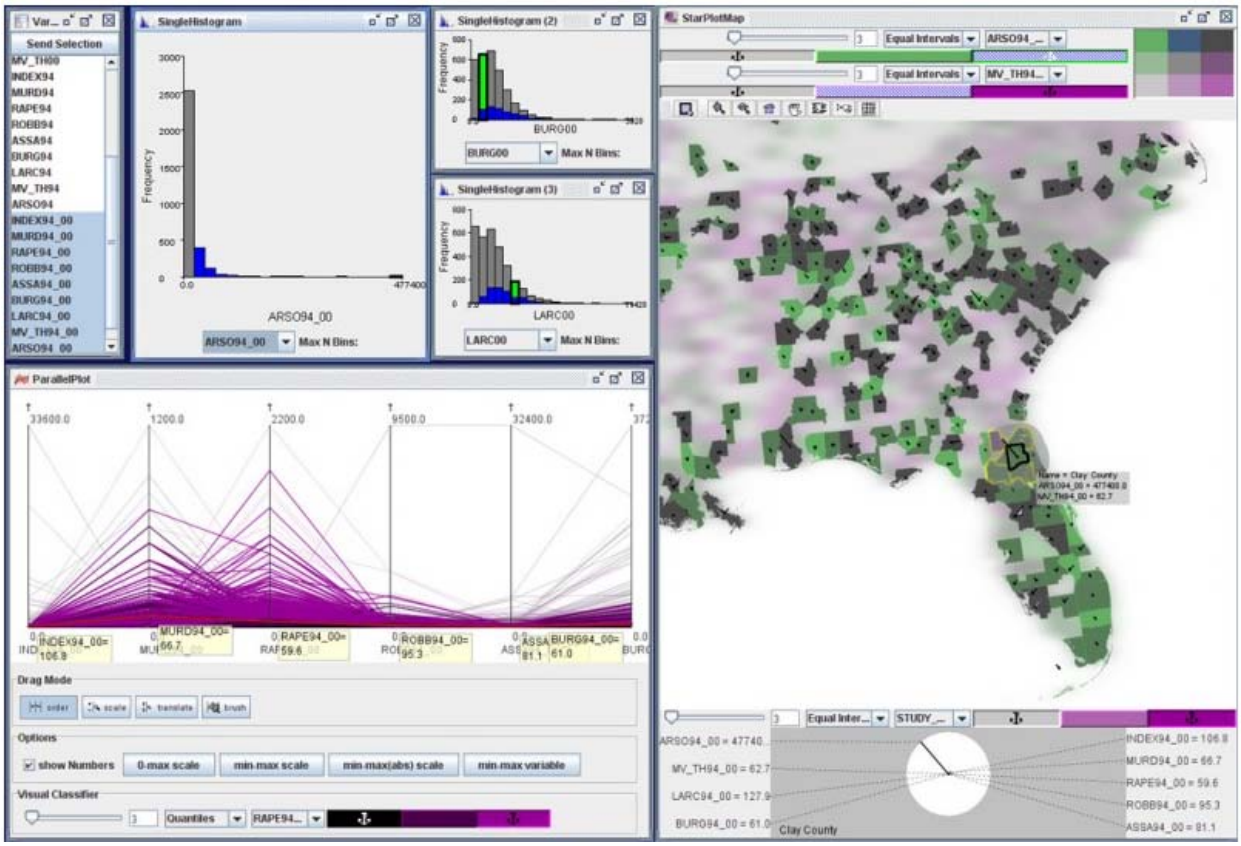


Figure 123 Example of GeoVIZ tools (www.geovista.psu.edu)

3.4.7 Decision Making

The final part of the executive dashboard monitoring system is the actual deployment of resources. The output of the decision support stage would be a list (map) of census tracts that fall into different bins based on the models calibrated by the statisticians. The deployment stage of the GSS Initiative requires a detailed plan to diagnose and remedy the elements of MTdb. Regardless of how the decisions are reached the deployment would involve a comparison of the current MTdb to current high resolution data sources.

As noted throughout all of our reports, there are vast arrays of geospatial data to support this comparison. While some of these resources are provided across the nation, many key resources are only available in specific areas. Consequently, there is no single approach to discover the best sources of data to support a “drill down” exercise. However, the GSS Initiative should explore the most cost effective and convenient way to access these services. As noted in our deliverable, *Reporting on the Use of Handheld Computers*, the South Carolina Gas and Electric Company (SCE&G) has a similar set of data maintenance needs to those of the Census. In fact, it maintains a decision support system that functions like the one we have proposed. In order to expedite the assemblage of data it contracts for imagery, parcel data and street centerlines. It closely integrates field based data collection via wireless communication and updates web services in a near-real -time manner. The final step in our model would deploy employees to the field to conduct inspections. Following the example of SCE&G the Decision Support System is

used to identify the best available resources to be transferred to a mobile device before an employee is sent into the field.

3.4.8 Other Related Topics

The following section includes a brief review of topics relating to targeted address canvassing.

3.4.8.1 National Urban Change Indicator (NUCI): MDA Federal Cross-Correlation Analysis (CCA).

The Department of Homeland Security is implementing a National Urban Change Indicator (NUCI) program that has direct bearing on the GSS Initiative. As noted above, these data could be used as a valuable indicator of where land cover changes are occurring. The data for this program is generated by MacDonald Dettwiler and Associates (MDA) from 30 meter Landsat data. Using a Correlated Land Change™ (CLC) procedure MDA claims to be able to "detect cultural features such as new roads, subdivisions and larger buildings". The CLC process looks for significant multispectral changes that persist. In a presentation to the National Geospatial Advisory Committee Mitchell made specific reference to the value of the CCA product for targeting change for updates to TIGER (Figure 124). It is important to note that within the Modernization Roadmap for the Geospatial Platform is a action item to "Determine the feasibility of acquiring and publically sharing a National Urban Change Indicator (NUCI) dataset using existing Federal data center investments"(Modernization Roadmap for the Geospatial Platform Version 3.0 August 2010).

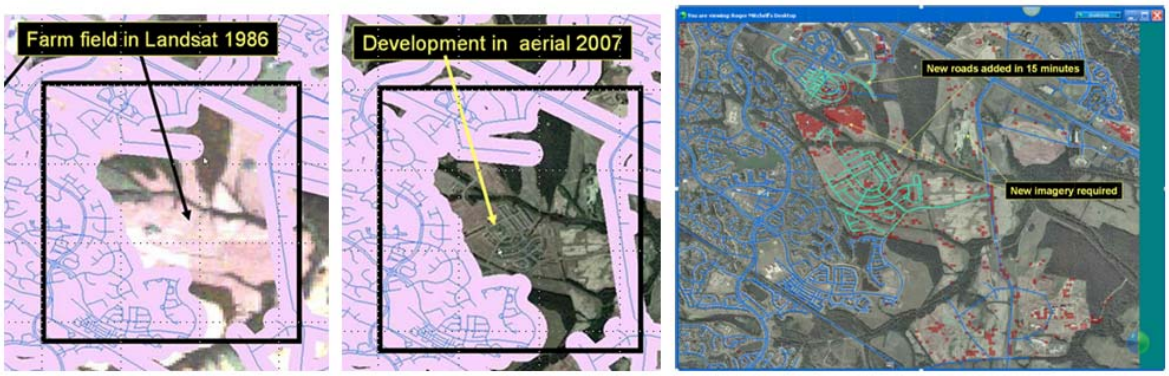


Figure 124 Example of MDA Correlated Land Change to identify change (Mitchell 2010)

3.4.8.2 NOAA Digital Coast Coastal Change Analysis Program C-CAP

http://www.csc.noaa.gov/digitalcoast/data/ccapregional/

The Coastal Change Analysis Program (C-CAP) produces a nationally standardized database of land cover and land change information for the coastal regions of the U.S. C-CAP products provide inventories of coastal intertidal areas, wetlands, and adjacent uplands with the goal of

monitoring these habitats by updating the land cover maps every five years. C-CAP products are developed using multiple dates of remotely sensed imagery and consist of raster-based land cover maps for each date of analysis, as well as a file that highlights what changes have occurred between these dates and where the changes were located.

3.4.8.3 Population Growth Projection Research

Targeted address canvass involves the identification of areas where the Geography Division believes change in housing has occurred. Traditionally, the Decennial Census of Population and Housing has served as the benchmark for the distribution of population and housing units. An active area of research involves estimating the changes in population and housing between the census periods. Commercial data providers such as CACI and Geolytics make a business of estimating the current demographic data and predicting the future. They utilize a number of public record sources to develop these estimates and predictions. As noted above these data could be used as one of many inputs to the decision support system.

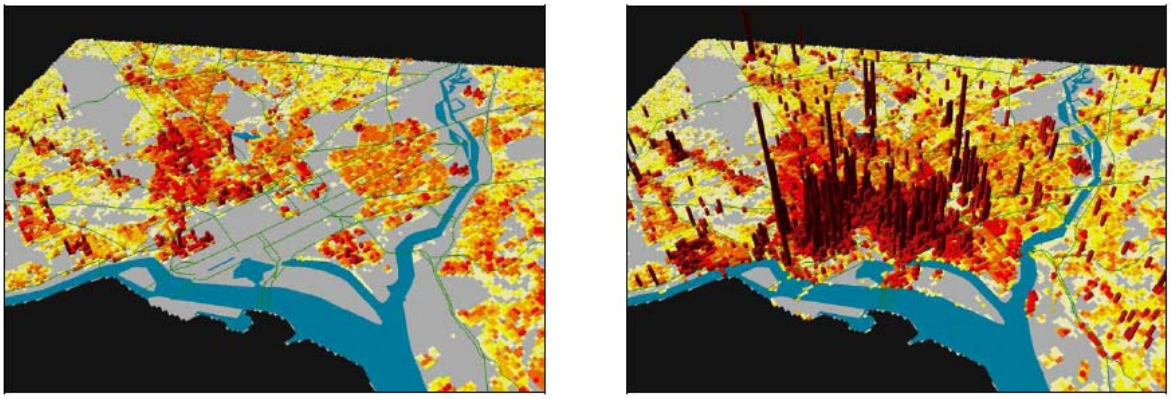
Developing models for small area population projection is an active academic research field. A theme common to several recent research projects has incorporated geographically weighted regression. For example, Lo (2008) developed a geographically weighted regression procedure to generate tract level population estimates for Atlanta. Yu (2007) also used geographically weighted regression to estimate housing values from remote sensing information in Milwaukee. Luo et al (2008) used a logistic geographically weighted regression to model spatial variations in urban growth. Huang et al (2009) incorporated a logistic regression model and exponential smoothing techniques to explore the effects of various factors (population density, slope, proximity to roads, and surrounding land use) on land use change. Qui et al (2003) utilized TIGER and remotely sensed data to estimate census tract level population in Dallas.

3.4.8.4 Day and Night Time Population Estimates

A final area of research that could yield some beneficial outcomes for the GSS Initiative involves the estimation of changes in population within a city between the daytime and nighttime (Figure 125). Researchers working on the Oak Ridge National Lab LandScan Global Population Database are leading proponents of this line of research. The motivation for this work is the need to estimate the population along transportation routes that may be used to move hazardous materials. The general description of this work is:

> "ORNL is currently developing a very high-resolution (90m cell) population distribution data (LandScan USA) for the US. LandScan USA includes nighttime (residential) as well as daytime population distributions. LandScan USA is more spatially refined than the resolution of block-level census data and includes demographic attributes (age, sex, and race). The potential benefit of LandScan USA has been demonstrated for 29 counties covering coastal Texas and Louisiana including the Houston metropolitan area. Locating daytime populations requires not only census data, but also other socio-economic data including places of work, journey to work, and other mobility factors. The combination of both residential and daytime populations will provide significant enhancements to

geospatial applications ranging from homeland security to socio-environmental studies." http://computing.ornl.gov/cse_home/about/LandScan%20long.pdf



Washington DC Nighttime Washington DC Daytime

Figure 125 Maps of daytime and nighttime population in Washington DC (Oak Ridge National Lab)

3.4.9 Summary

We believe that the GSS Initiative should evolve into a robust decision support system that would treat targeted address canvassing as part of an ongoing surveillance system. In the section we have outlined a series of steps that we believe can be implemented using many of the Geography Division's existing resources and partnerships. We recommend that a diverse set of internally and externally provided indicators be assembled and fused at the tract level. This rich data system could provide the information required to conduct an efficient assessment of the quality of the MTdb across the nation and highlight those areas in need of attention. We believe that a prototype of this system should be part of the pilot test in preparation for the next generation of address canvass.

3.5. Investigate the Current and Potential Use of Crowdsourcing/User Generated Data for Collecting Spatial Data

We reported on the use of crowdsourcing (also known as Volunteered Geographic Information (VGI) and Citizen Science) as a method of collecting address information in Deliverable 2 "*Reporting the State and Anticipated Future Directions of Addresses and Addressing.*" We recommend that parties interested in crowdsourcing and its potential application by the Geography Division may want to read Sub-Task 3 of that document for a more thorough introduction to crowdsourcing than is presented here. In the present report, we will focus our analysis on the use of crowdsourcing for collecting data that could be used in augmenting or extending the types of spatial data found in TIGER rather than the specific address data found in MAF.

Crowdsourcing is often thought of as an architecture of participation in which the concepts of sharing, cooperation, and collective action allow crowds of people of ordinary skills to build and publish databases without the need for active coordination, management, operational budgets, or an overarching business structure. Wikipedia, the free, World Wide Web-based, collaborative, multilingual encyclopedia is, perhaps, the most successful of the crowdsourced database efforts. The map database known as OpenStreetMap is the prime example of the crowdsourced compilation of spatial data. The Tele Atlas Multinet Database and the Google map and navigation database are the foremost examples of commercial entities using crowdsourcing to supplement more traditional forms of compilation used to build spatial databases.

In a very real sense the Web's role as a laboratory for social networks reflects the conclusion of James Surowiecki on "The Wisdom of Crowds" that "…under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them.[6] He continues providing example after example that "The simplest way to get reliably good answers is just to ask the group each time."[7] If Surowiecki is right, then, online social network websites such as OpenStreetMap, whose members (the group) are tied together by their interests in spatial data might be expected to produce a map database of a reasonable quality level based on their field observations and knowledge of local geographies.

Crowdsourcing has been enabled by the mass collaboration made possible by social networking and by the asynchronous connectivity that is the hallmark of the Web. Online collaboration benefits from collective intelligence and knowledge aggregated from decentralized sources and low cost. More specifically, the crowdsourcing of spatial data has benefitted from the above tools, but also from: the availability of handheld devices capable of recording GPS traces while a person or a vehicle moves between locations on a transportation network, the public availability of high resolution satellite and aerial imagery used for reference, and the development of open sourced Geographic Information System software.

The application of crowdsourcing to spatial data was initiated in 2004 by Steve Coast in the United Kingdom, in part as a response to the high cost of and restrictive licenses for the use of Ordnance Survey data in that country. The goal behind OpenStreetMap was to provide free

[6] James Surowiecki, 1967 "The Wisdom of Crowds", New York: Random House, p XIII
[7] Ibid.

geospatial data that could be edited, updated, and augmented by anyone, as long as the changes that they made to the data were available to everyone under a Creative Commons Share Alike license.[8]

In addition, OpenStreetMap was a response to the notion that local maps were often out-of-date, incomplete, and non-comprehensive in respect to the types of data they included. The Web, though social networking, has made it possible for people to share their insights and information about their local community and doing so often requires a basemap to show the location and attributes of the information to be shared. For example, AOL's Patch uses OpenStreetMap as a basemap on which they display community events and news. Through its MapQuest subsidiary, AOL is able to update the maps shown on Patch every 10 minutes, and even allows local updates by Patch users.[9] As a consequence, the concept of "neighborhood" sharing enabled by social networking, OSM, and similar efforts could be useful to the Geography Division as sources for data discovery and change detection in local geographic areas. (See Figure 126.)

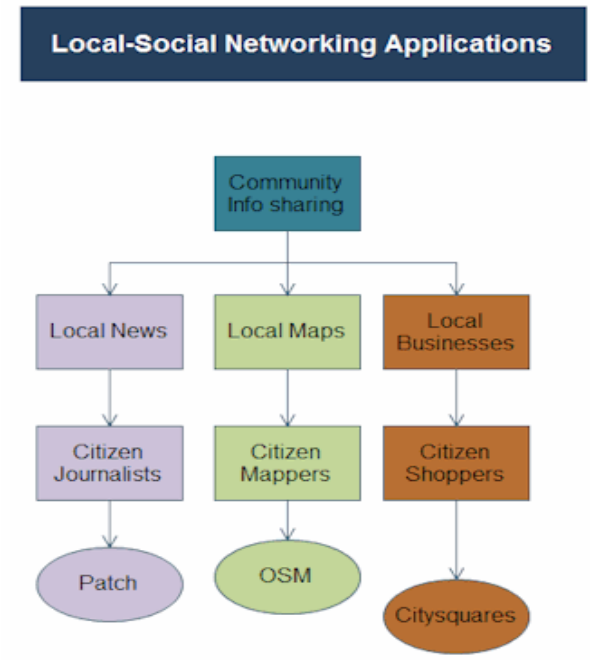


Figure 126 The power of social networking and the concept of neighborhood sharing are powering the creation of local spatial data on the Web.

The issue of the data quality provided by crowdsourced spatial databases is a concern that we will weave through our discussion of the topic. The nature of crowdsourcing suggests that the integrity of the data depends on the law of large numbers. At the heart of crowdsourcing is the thought by Linus Torvald, initiator of the Linux movement, who expressed the concept, that in the development of crowdsourced software "... with more eyes all bugs become shallow."[10] As

[8] See http://creativecommons.org/licenses/by-nc-sa/2.0/ . The OSM database is now licensed under the Open Database License (ODbL). For more information see http://www.opendatacommons.org/licenses/odbl/
[9] See http://hermosabeach.patch.com/search/listings?cat=1998198544
[10] See Wikipedia http://en.wikipedia.org/wiki/Linus%27_Law

applied to crowdsourced map databases, such as OSM, the idea is that with more eyes all map errors become visible and candidates for correction.

Due to their "always editable" status, crowdsourced spatial databases are constantly changing and, as a consequence, their error signatures are considered (at least theoretically) to be self-healing over time. Whether the healing actually takes place in a uniform and helpful manner is a complex issue that involves the number of participants contributing spatial data, their intentions and motivations, their interest in contributing data over long periods of time, and the spatial distribution of these contributors that is required for comprehensive map coverage. In essence, it is an open question whether or not the spatial data quality of crowdsourced mapping efforts can be managed to meet specific requirements on a reliable basis. For example there is a website devoted to pointing out errors in OSM called OpenStreetBugs which is designed to allow interested parties to find and fix bugs. Unfortunately, the site does not appear have gained much traction with uses, perhaps reflecting the requirement of local knowledge to correct errors in road geometry or attributes.[11] It is important to note that there is not a focused driving force urging OSM contributors to complete x coverage by y date. Instead, the schedule of coverage is rather like topsy, it continues to grow at its own pace, where and when contributors feel interested in adding, correcting or augmenting data. Time frames and formal correction cycles are not a part of the crowdsourcing world.

Recent significant research by Girres and Touya on the quality of the French OpenStreetMap Dataset raised questions on the heterogeneity of the crowdsourcing process, the scale of production and the compliance of contributors to standardized and accepted specifications.[12] The authors concluded that OSM has great promise, but that its data was of variable quality and would remain so until the tension between "openness" and standardization of requirements was restructured with specific requirements for data entry and attribution.

Regardless of the potential limitations of crowdsourcing in terms of data quality, commercial businesses and government agencies have adapted crowdsourcing for use as tool to enhance their efforts to build spatial databases. In most of these cases, however, crowdsourcing is used in conjunction with other compilation techniques in a hybridization process which mixes both controlled and uncontrolled data in a process using methods that attempt to quality control all data.

For instance, the Federal Office of Topography (Swisstopo) has integrated user generated content into their national databases - Topographic Landscape Model (TLM and Digital Cartographic Model (DCM).[13] Swisstopo apparently has taken the view that crowdsourcing could provide a new source of content for upgrading their national databases, as well as the possibility of opening new relationships with their customers. The system allows the uploading of GPS-tracks, attachments, and notifications of errors, as well as surfaces and lines. All input is funneled through an online-interface connected with a central revision layer whose dispatcher distributes the comments to the appropriate database. The crowdsourced data are considered to

[11] See http://wiki.openstreetmap.org/wiki/OpenStreetBugs
[12] See Girres, Jean Francois and Guillaume Touya, 2010, Quality Assessment of the French OpenStreetMap Dataset, Transactions in GIS 12(4), pp 435-459
[13] See Guélat http://www.eurosdr.net/workshops/crowdsourcing_2009/presentations/c-4.pdf

be sources of change detection that may or may not require field investigation before their acceptance into the national databases.

In the commercial arena, Google was an early adopter of crowdsourcing to build map data, principally through its application Map Maker[14] that allowed its users to create and update maps of areas that showed locations of interest to them, as well as allowing them to update maps and add features that were missing from the Google maps (at the time licensed from both NAVTEQ and Tele Atlas). Managers at Google depended on their commercial map suppliers to provide them accurate, up-to-date, and comprehensive spatial data that could help support their local advertising business, as well as to serve as the foundation for their aspirations in the world of mobile commerce, which they believed, would be map-centric and focused on the local areas surrounding mobile users.

The Google Map Maker application originally was developed to help provide map data in countries where neither NAVTEQ nor Tele Atlas had map coverage. Google, whose license with its users gave the company a license in and to the data that its users created, often used the maps created in Map Maker to form the basis for their own map base. In essence, the data contributions of their users were: quality engineered, compared to high resolution imagery, conflated with other geographic sources, and then included in Google Maps, its corporate map franchise for consumer use.

Google Maps has now replaced the map data for the United States and Canada from NAVTEQ and Tele Atlas with spatial data that they have compiled through a variety of methods. Although the majority of these data have not been created through crowdsourcing, Google has added a "report a problem" button on the bottom of Google Maps that allows users to suggest changes for Google Maps.

Tele Atlas is the commercial map database provider most tied to crowdsourcing, although the impetus for this conversion actually started with TomTom, a company that provides Personal Navigation Devices (PNDs), a lower priced, highly efficient alternative to in-car navigation systems. TomTom had always used Tele Atlas as the supplier of its navigation data and was dissatisfied with length of the process required for Tele Atlas to respond to customer complaints about map inaccuracies. Even worse, although the reporting-feedback cycle on errors was approximately two-months, it often took two to three times as long to include these changes in the distributable version of the TA databases (Multinet). In response, in 2007 TomTom developed a product known as Map Share[15] that encompassed crowdsourcing using two very different, but complimentary methods that we classify as Active Community Input (ACI) and Passive Community Input (PCI).

The crowdsourcing of spatial data is likely to become a key component of all major spatial database compilation efforts. We examine the promise of crowdsourcing and more detailed examples of how it is being used in mapping and database building in the following section of this report.

---

[14] See http://www.google.com/mapmaker/mapfiles/s/support.html

[15] TomTom announced its Map Share technology in 2007 as part of its announcement of the TomTom Go model of PND.

3.5.1 An Examination of Crowdsourcing in the Form of Active and Passive Community Input (User Generated Content) for Tracking Map Changes Including Roads and Addresses.[16]

Our interest in canvassing the major methods of community input are to provide the Geography Division with examples of how crowdsourcing has been implemented in the compilation of spatial data. In this section we describe the two major methods of collecting spatial data, active and passive community input.

3.5.1.1 Active Community Input

Active Community Input is a change to a map database made by a contributor as the result of a process in which they must take the initiative to identify the desired change and take purposeful action to convey the change to the provider of the spatial database. Examples of this method would be OpenStreetMap and TomTom's Map Share software.

Active contributors serve both as compilers and editors of the work they contribute, as well as editing the work of other contributors. In general, the types of operations available to active contributors take the following forms.

Move features (relocate position of point, line, area)
Delete features (remove feature or segment)
Insert Features (add new point, line, area or associated attributes)
Modify Features (change non-positional attributes attached to points, lines, areas.)
Connect Features (connect disconnected paths (navigation))
Block Features (disconnect connected paths (navigation)
Personalize Features (not available in all systems - make changes or additions across categories that reflect individual contributor's tastes and preferences. This action may be focused on POIs, but it is sometimes used to rename roads with common, local identifiers.)

Table 2 Selected database functions available in active crowdsourcing[17]

	Shared Information								Private
User Actions	Move (new position)	Delete	Insert	Modify	Connect	Block	Unblock	Personalize	Personalize
Features and attributes									
Roads and Ferries	x	x	x	x	x	x	x	x	x
Named Areas	x	x	x	x				x	x
General Features	x	x	x	x				x	x
Land Cover/Land Use	x	x	x	x				x	x
Services/POIs	x	x	x	x				x	x
Structures	x	x	x	x	x			x	x
Waterways	x	x	x	x	x	x	x	x	x
Railways	x	x	x	x	x			x	x
Road Furniture	x	x	x	x				x	x
Administrative Areas	x	x	x	x	x			x	x
Public Transport	x	x	x	x	x			x	x

[16] As noted previously, the use of crowdsourced data for addressing was described in Deliverable 2 on Addresses and Addressing. The focus of this discussion will be on spatial data other than addresses, although addresses are referenced occasionally.

[17] Road Furniture is a term used in the GDF (Geographic Data File Format, now ISO14825:2004.) used to describe road signs, traffic lights and other official road management features found along road edges and often related to traffic management. See this page for a non-technical overview of GDF - http://www.ertico.com/gdf-geographic-data-files/

Although the commercial providers of databases who use active crowdsourcing do not reveal specific details on their use of crowdsourced data, our conversations with representatives of these companies leads us to postulate that the active crowdsourced information they receive is mainly used for purposes of change detection, that is, to indicate where their database is either incomplete or incorrect. Further conversations have revealed that geocoding and address errors are the most commonly reported errors, followed by a focus on points of interest (POIs), missing streets, and/or incorrect street names. See Figure 127 for more details on this breakdown.

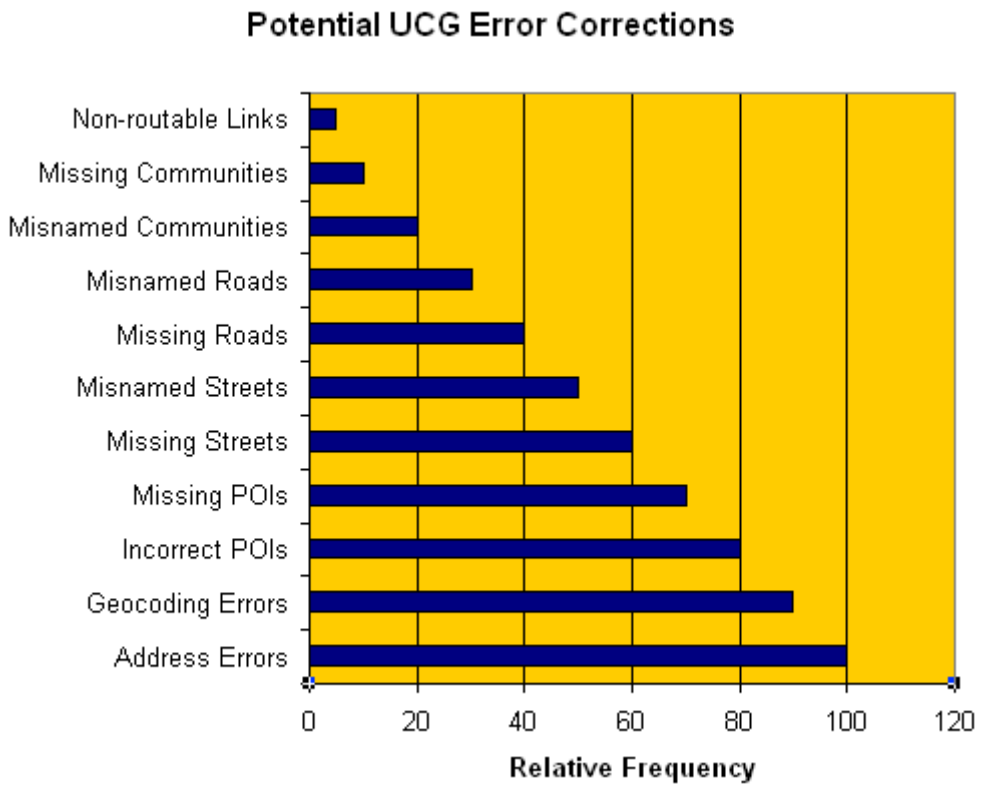


Figure 127 Error components resolved by "active" crowdsourcing.

.

It is important to note that most commercial map database providers consider crowdsourced data to be a validating agent for checking the integrity of their existing database, rather than as a source of readily useable content. The reasons for this orientation often relate the vagueness of the data found in active crowdsourced input and an ongoing problem with negative consumer intent in the form of “malicious” (i.e. purposefully erroneous) suggested changes. In addition, some active changes that are reported are not associated with GPS readings and often are only loosely associated with the actual location of information to be corrected. It is fair to say, that most, but not all, organizations that work with crowdsourced data prefer to work with passive community input since it is difficult, although not impossible, to contaminate passive community input.

Tele Atlas, for instance, shared with us that people often use active input in a manner designed to achieve a goal of interest to them. Apparently, Tele Atlas receives a modest number of requests to change speed limits from individuals who feel that the local speed limits around their homes

are too high and indicate that the limit posted by Tele Atlas is incorrect in a misguided attempt to promote neighborhood safety.

OpenStreetMap, however, regards active community input as a core asset, although even here while the input is actively provided, the common form of providing data to OSM is GPS traces. The traces are recorded by the OSM contributors using a GPS receiver, such as those provided by Garmin or Trimble. OpenStreetMap however, must rely on its contributors to attribute the GPS traces they provide and this is where the OSM database is extremely weak. Haklay and Ellul analyzed the OSM database of the United Kingdom and found that compared to their reference (The Meridian 2 database from the Ordnance Survey) the OSM database had complete "cells" (meaning roads and attributes) only in 21% of the country (covering 45% of the population), although most of the country was covered with roads that were usually unattributed and often unnamed.[18] (See, also, Figure 145).

3.5.1.2 Passive Community Input

Passive Community Input is a term generally denoting the use of GPS devices to collect what is known as floating path or probe data. In essence, the PND or GPS receiver used by the contributor tracks their paths of movement (usually vehicular) during their daily journeys.[19] The points comprising the GPS traces are logged based on a predetermined distance or recorded and logged based on time. (We discuss of the data exchange mechanisms used to receive authenticate and ingest data from the contributors to various crowdsourcing systems in the next section of the report.)

The most well-known and successful example of passive community input is TomTom's Map Share application, which automatically tracks TomTom users who have elected to allow the GPS traces of their activities to be anonymously contributed to TomTom.[20] In turn, TomTom uses these data to update the Tele Atlas navigation database, as well as to derive traffic and travel time statistics for streets and roads to improve the routing capabilities of their products. While individual tracks may not be extremely accurate, TomTom believes that aggregating them can provide data accuracies rivaling the most precise field collection units.[21] TomTom has now collected approximately three trillion GPS points using Passive Community Input. See Figure 128 for an example of mapped GPS traces.

[18] Haklay and Ellul, Completeness is volunteered geographical information, Journal of Spatial Information Science, 2010. p8.

[19] Although OSM contributors use GPS receivers to collect street data, the use is part of an active decision to capture data along a specific road or section of a road network. In passive collection, the user has no pre-planned courses of action related to the collection of road data.

[20] The actual tab on the TomTom Home site, which implements this action, reads as follows, "TomTom would like to gather anonymous statistical data to improve your map quality and navigation experience. We would retrieve this data whenever you connect your TomTom device to TomTom. It will not be noticeable, will not cost anything and will be completely anonymous. Setting this preference will allow us to retrieve this data."

[21] Personal conversation with Christopher Wilson, Director of ADAS, TomTom/Tele Atlas.

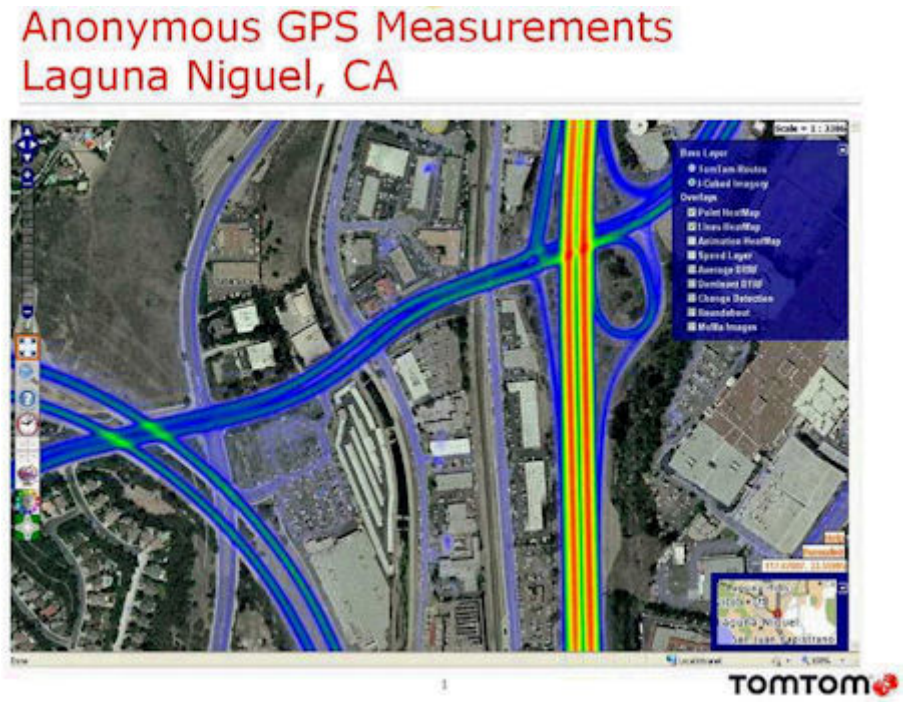


Figure 128 Heat map of GPS traces in Laguna Niguel, California. The red lines on the right define the I-5, while the crossing route from northeast to southwest is a local arterial named Crown Valley Parkway. The roads shown in light blue indicate a lower number of probe traces in areas of industrial parks, while a local shopping center is shown in the lower right. Image courtesy of Tele Atlas.[22]

It is important to note that collecting probe data does not include reports on data attributes, as it is limited to describing the geometric traces of streets and roads along which the GPS units travel. While some attributes can be inferred from the traces (number of lanes, elevation, travel direction, distance, slope, etc.), they are associated with the physical characteristics of the road centerline that is built up by the aggregation of traces along roadways. Street names, house addresses, municipality, or other characteristics of streets and roads that might be of interest to the Census Bureau are not part of the information available from passive data. However, the passive data does demarcate where people drive and a comparison of probe data across a time series can be of significant benefit for determining change detection.

Tele Atlas was skeptical about the use of these data when they were provided shortly after the company was acquired by TomTom in 2008. The Tele Atlas researchers spent a year finding out how to best use the data and how to integrate these data into their compilation workflow. The company now believes that crowdsourced data has turned out to be of incredible value and it has become their preferred indicator of change detection. It is the belief of Tele Atlas management that community input helps the company detect and target crucial road information that is hard to discover otherwise, but easy to exploit using Passive Community Input.[23]

[22] See http://blog.telemapics.com/?p=323 for more information on TomTom's Map Share.
[23] Personal conversation with Patrick McDevitt, Vice President of Crowdsourced Data, Tele Atlas.

Figure 129 Tele Atlas claims that the miles of GPS traces it receives in the United States in one day are equivalent to the total miles of roads in the United States and the data collected in Europe represent four times the total kilometers of road in Europe.

Originally, Tele Atlas thought that the probe data might allow them to create better street centerlines than they currently had in their database. While acknowledging that there are problems with mobile GPS data collection quality (it can be of variable quality, being impacted by the influence of urban canyons, scattering, foliage, etc.), the company soon found out that aggregating the large numbers of signals that are reported to them averages out these types of errors. They commented that the coordinates reported from the PNDs are less accurate than those captured by their mobile mapping vans, but that the sheer numbers of points they collect make data about most locations more and more accurate over time. In mid-2010 a Tele Atlas representative told one of the authors of this report that they had collected approximately three trillion GPS points through Passive Community Input and every day they collected enough traces to cover the North American Road Network and that in Europe, the collected road data represented four times the amount of roads in the European road network.. Although the analogy is slightly misleading since the coverage of the traces is not comprehensive across the reference areas, the numbers can be taken as an indication of the potential value of probe data to create spatial databases.

The use of passive crowdsourcing to compile map data is constrained by the number and distribution of the users willing to voluntarily contribute anonymous GPS traces. To date, the efforts appear spatially autocorrelated with population density. GPS traces are more common in cities and larger cities are covered by more traces than smaller cities. Rural areas suffer from low density of traces, due to the diminished size of the audience available to volunteer their information.

While NAVTEQ and Google uses probe data to advantage their database building efforts, neither approach is substantially different from the probe data approach taken by Tele Atlas. NAVTEQ has partnered with a number of companies that use its data for logistics and provide the GPS traces of their units back to NAVTEQ. Although probe data is now being gathered by Google through mapping applications on its Android cell phone platform (and eventually by NAVTEQ through its association with Nokia), there is some question as to the relative accuracy of probe data generated by the less sophisticated GPS receiver and inferior antennas provisioned in cell phones compared to PNDs.

3.5.1.3 Summary

It is our opinion that the combination of conventional compilation techniques, driving streets with instrumented vans, active crowdsourcing and passive crowdsourcing will soon become the

established norm for compilation of the spatial data used by Google, Tele Atlas and NAVTEQ (see Figures 130 and 131). OSM will continue to rely on Active Community Input, although its dependence on GPS traces provides some of the advantages of true probe-based systems.

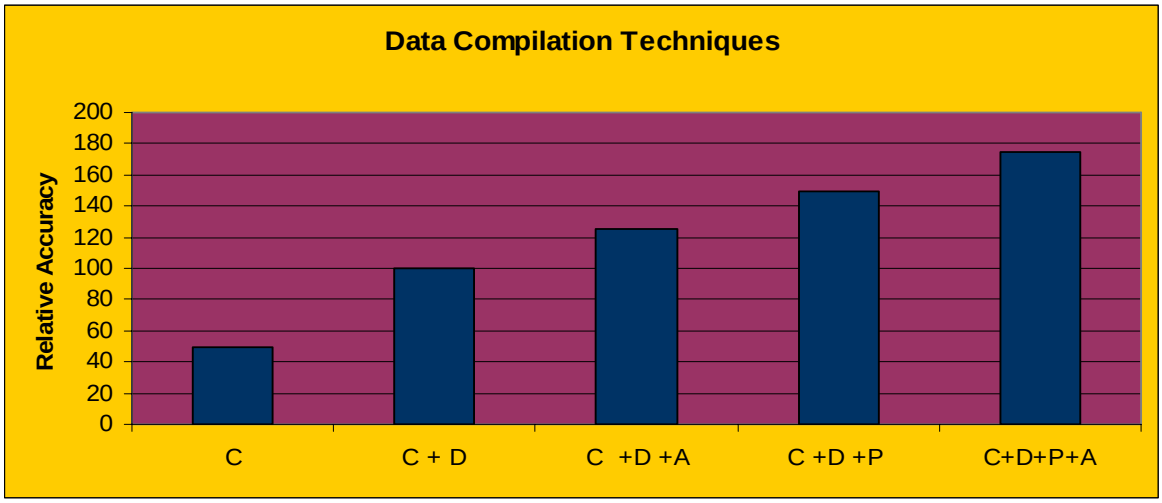


Figure 130 Adding crowdsourcing capabilities to the compilation model results in increased accuracy. Using the Hybrid model (CDPA) will generate more robust and relevant databases for navigation and mapping than other alternatives. C = traditional compilation methods, D = Driving roads/streets with instrumented vans, A = Active UGC, and P= Passive UGC

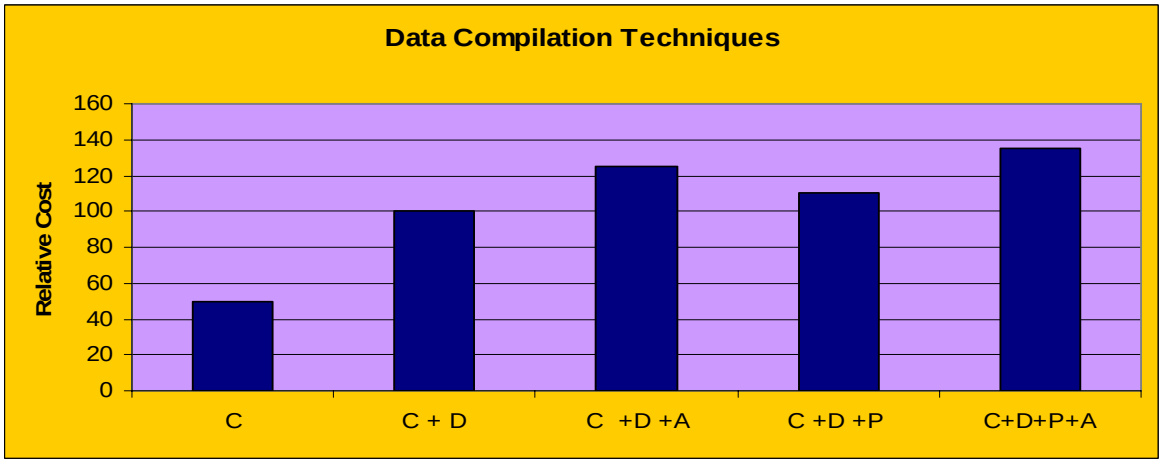


Figure 131 It is likely that using all data compilation techniques will add to the expense of data collection, but will also add to the accuracy and comprehensiveness of the data that result. It is likely that UGC will play a significant role in spatial database compilation efforts as a cost conservation measure, if for no other reason. C = traditional compilation methods, D = Driving roads/streets with instrumented vans, A = Active UGC, and P = Passive UGC.

The use of active or passive community input is impacted by the ease of use of the method. It is our belief that the ease of use directly correlates with the frequency of contributions. Continuing active contributions, like those required for OpenStreetMap, are likely to attract map “geeks” who love maps and have the time to collect data using a GPS, upload them, attribute them, and then edit the results. Passive applications require little interest in maps and only require docking

your PND with your computer. While some might question why anyone would bother docking their PND, the answer is quite simple. You need to dock your PND to get the community updates collected by others and vetted by Tele Atlas. In addition, when you dock your TomTom PND, you also are able to download updated ephemeris data that helps your PND to fix satellites much quicker than a similar device that has not been updated. However, the main observation here is that over time, passive community input should provide significant benefits to those capable of managing such a system (see Figure 132).

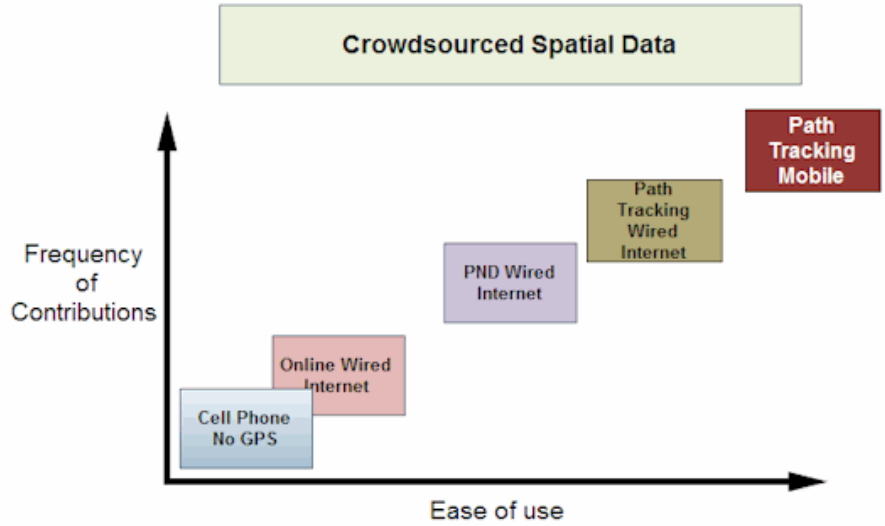


Figure 132 The ease of use of passive contributions increases the frequency of contribution by its users.

3.5.2 A Description and Evaluation of the Data Exchange Mechanisms Used to Receive, Ingest, and Authenticate Data from Volunteers.

The data provided to various systems from the user community results from the application of a wide range of techniques. We describe a variety of approaches for collecting and ingesting these data in this section of our report. Regarding issues of data quality, it is the philosophy of some organizations that solicit crowdsourced data that the data ingested should be accepted as reported and changed by other users, over time, if it is not accurate. OSM, for example, adheres to this philosophy, although the commercial users of crowdsourced data apply it in a hybrid solution where it is compared to an existing database and then researched based on an evaluation of its promise. In some sense, the commercial approach is to use the crowdsourced data as a measure of change detection that are subsequently examined using the preferred techniques of the database providers. In essence, there are a variety of approaches to collecting, transmitting, and evaluating crowdsourced data.

A useful general model of the data exchange mechanism used in crowdsourced systems is provided by Hall et. al.[24] in their overview of creating the crowdsourced-based MapChat system designed to facilitate citizen input of volunteered geographic information.

[24] See Hall, G.B., Raymond Chipeniuk, Robert D. Feick, Michael G. Leahy and Vivien Deparday, 2010, Community-based production of geographic information using open source software and Web 2.0. International Journal of Geographical Information Science Vol. 24, No. 5.

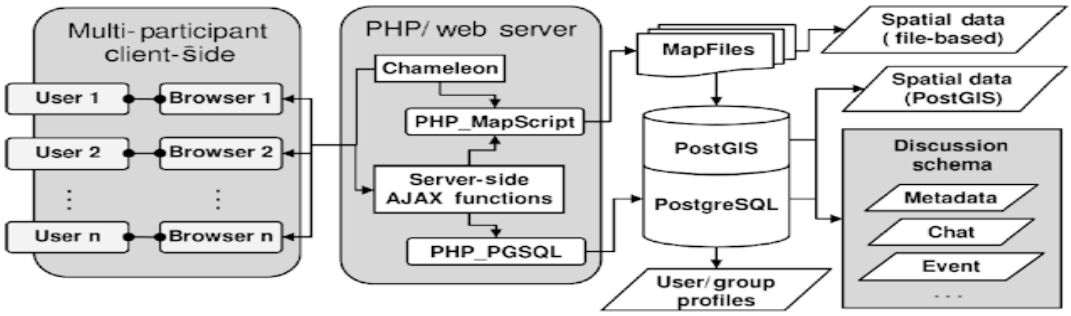


Figure 133 High level architecture of Map Chat (source Hall and Leahy 2008). For each user's contribution, a discussion schema is generated to record and manage user data generated while interacting with the software.

All active crowdsourcing platforms use some variant of the diagram above to capture the data provided by their contributors. Data are entered directly or indirectly through a map-based, Web browser application that interacts with a spatial database through a variety of tools. While metadata are collected and history files are kept by all systems, it is in the processing of the crowdsourced data that methodologies diverge.

3.5.2.1 TomTom/Tele Atlas Data Ingest and Evaluation

TomTom has taken a unique approach to data collection in that the spatial data provided by the users may be entered using a Personal Navigation Device or a browser, although all change data must be communicated through a browser interface.

TomTom's Map Share product is a useful example of active community input. The user of TomTom PND can enter data while interacting with the screen of their PND, while at the location where they believe data is erroneous or missing, or they can simply mark the location on the screen by touching it and make the correction later using their desktop computer and the TomTom Home extension of Map Share.

TomTom's MapShare lets the user make six types of changes to the map that result in an immediate change to the local database resident on the PND (See Figure 134). The changes that can be made to the operation of the database on a user's device are:

- Block or unblock a street in one or both directions
- Reverse direction of a street
- Name or rename a street
- Add a POI (Point of Interest)
- Edit a POI (including removal
- Report other errors

In addition to these changes that are immediately operational on the individual's device, the user can report 27 other types of map errors that go to Tele Atlas for vetting.

As a TomTom user, you can determine your level of trust in the corrections made to the database by allowing your device to be updated only by specific sources, such as:

Changes made by the user
Changes verified by TomTom (researched and approved by Tele Atlas)
Changes to POI lists to which the user subscribes
Changes from trusted sources (e.g. other trusted users)
Changes reported by many (use all of the changes reported to TomTom) Changes reported by some (One the sources actively selected are downloaded)

Clearly, TomTom has mined its users' preferences and created MapShare as a method of addressing the most common "inconveniencing" problems that erroneous map data create. Most of the "direct" corrections allowed by MapShare are aimed at solving the "I can't get there from here" problem, which is usually caused by incorrect addresses location, impeded access (road not a through road), incorrect directionality, or the inability use the database listings information to locate and travel to a Point Of Interest (POI). Allowing the user to immediately benefit from the changes they suggest is accomplished through the creation of a change layer that interacts with the local navigation database on the user's PND.

Figure 134 TomTom PND screen for beginning the map correction process

Once the screen has been marked for correction (see Figure 135), the user can make the change immediately, or postpone the action until they return home and connect the PND to their computer to initialize the Map Share program that is used to process and upload edits, as well as to download updates to improve the efficiency of the PND. The computer-screen view of MapShare (shown below) mimics using the PND and brings up a map of each of the areas marked previously for correction. Working on the computer screen version of the product is much easier and more satisfying than attempting to make changes on the PND itself, though TomTom indicates that many users prefer to update on the device.

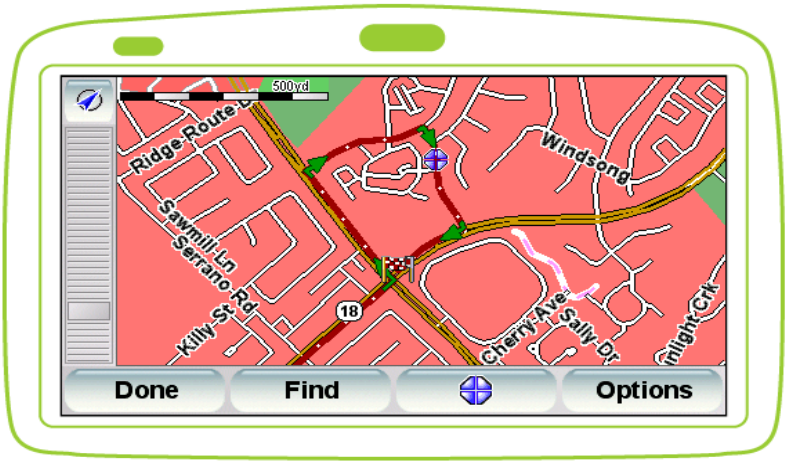


Figure 135 Adding a road to the TomTom database resident on the PND through use of the Map Share interface (which mimics the PND screen).

Passive Community Input from TomTom's PNDs is collected as an automatic upload only when the PND is docked with a computer running Map Share. Map Share automatically recognizes the PND and starts the data upload which is transferred to TomTom. TomTom preprocesses the data and strips the first two minutes of traces, as well as the last two minutes of traces from the GPS string in order to ensure the anonymity of the user. After preliminary processing, the data is transferred to Tele Atlas.

Both the active and passive data are provided to Tele Atlas by TomTom. Note that there is no formal error checking or correction when the data are transferred from the user to TomTom. The crowd sourced view of the world is that all input should be considered accurate until corrected by newer input. TomTom and Tele Atlas, however, use a hybrid approach by setting the current version of the TA Multinet database as the standard of accuracy. All changes are compared to the current content of the Multinet database and any differences are considered indicators of change to be verified by targeted research.

New roads or other information suggested by active input are compared to the most up-to-date imagery. Streets and roads discovered by examining probe data initiate data mining to find sources containing attribute detail, followed by field examination if data mining does not produce an unequivocal identification.

Tele Atlas has a number of partnerships with companies that provide it with probe data recorded as these companies' drivers make their daily rounds. The data are usually uploaded to TomTom as a purposed data feed and are unrelated to Map Share, although these data are examined and verified using the procedures described above.

Finally, Tele Atlas has an online application called Map Insight which it provides as a way for non-TomTom users of Tele Atlas data to make online corrections to errors found in the database.[25] Map Insight contributes a small portion of the updates processed by Tele Atlas and

[25] See the Tele Atlas website for more details on Map Insight http://mapinsight.teleatlas.com/mapfeedback/index.php

we will not describe it here, as it is similar to the NAVTEQ online error correction service we describe below.

3.5.2.2 NAVTEQ Data Ingest and Evaluation

NAVTEQ is a data supplier that does not have the strong customer facing presence of Tele Atlas through its association with TomTom. As a consequence, NAVTEQ currently does not have a strong source of either active or passive community information, at least on the part of consumers using its data. It is expected that Nokia and NAVTEQ will remedy that issue in the future by NOKIA sharing GPS traces from users of its OVI Maps application that runs on the Nokia smart phone platform. While NAVTEQ does have a number of commercial partners in the logistics field who do provide it with valuable probe data, its major source of crowdsourced data is its online NAVTEQ Map Reporter.[26]

Figure 136 NAVTEQ's Map Reporter input page

The Map Reporter is an example of active community input since the interested party must go the site, activate an account and enter the type of correction of interest. The site is geared to allowing: the entry of Points of Interest; changes to the location of a house or a building; or adding, editing or removing streets, roads, and road features such as signs, one-ways, and restrictions (See Figure 136 above and Figure 137 below).

[26] See the NAVTEQ Map Reporter at http://mapreporter.navteq.com/. In order to look under the hood, you must register, which requires identifying the device whose data is incorrect.

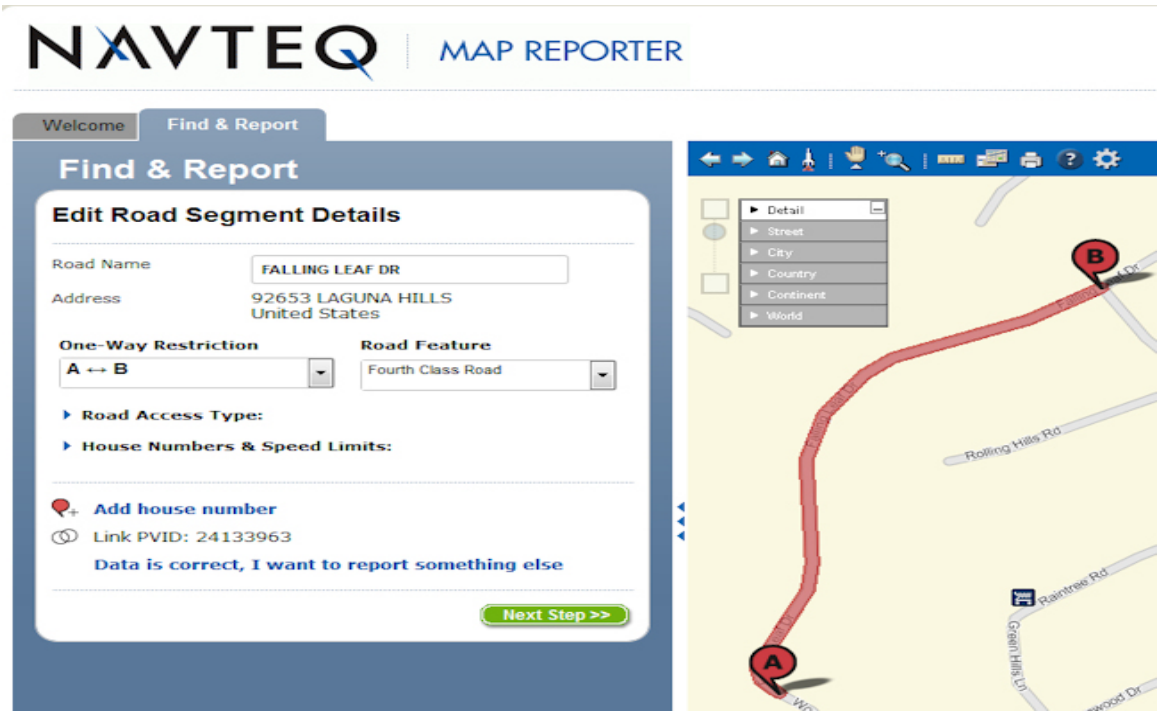


Figure 137 Detailed view of the changes that you can make to streets in the NAVTEQ Map Reporter

The method of screen interaction with the Map Reporter is poorly executed. However, the user can simply fill out a quick report in an on-screen form that only requires a brief description of the problem of interest.

NAVTEQ accepts all data without verification at the data input level, but then assigns the evaluation of the corrections to an internal group that attempts to verify the input against the current state of the NAVTEQ database. Data mining is the initial verification step and data that pass this hurdle are either accepted based on internal hurdles or examined in the field through use of the company's extensive field research operations.

3.5.2.3 OpenStreetMap Data Ingest and Evaluation

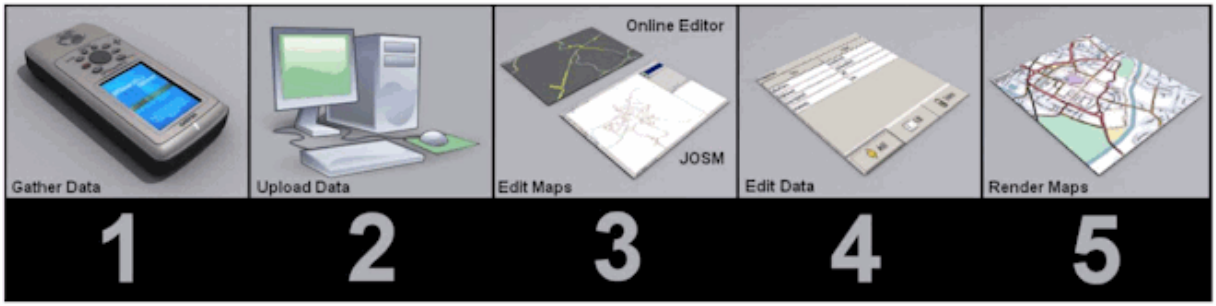


Figure 138 OSM's data collection and edit process (source: OSM)

OpenStreetMap's data flow is almost as simple as it appears above (see Figure 138). Users are free to collect data using GPS, personal knowledge of an area, public domain imagery (or

imagery donated to OSM for the purposes of mapping), or personal maps not based on copyright protected sources.

Data from GPS receivers are uploaded to OSM in GPX (a lightweight XML-based GPS Exchange Format). There are a number of open source editors, such as GPSBabel, that can be used to convert GPS tracks to GPX from the internal file format of most GPS receivers.[27] The GPS traces are then uploaded online to OSM from the main OSM interface page (http://www.openstreetmap.org/) and can be marked for public sharing or kept private. Once uploaded, the data (in either raw or mapped format) can be downloaded and edited.

3.5.2.3.1 Edit Maps And Data

JOSM (Java OpenStreetMap Editor), Potlach (Flash-based editor) and Merkaartor (OSM editor for UNIX, Windows and Mac OS X) are three of the over 50 free software packages that exist to capture and edit data for OSM.[28] More recently even more user-friendly editors have been developed, including Mapzen, an editor provided by CloudMade (formed by Steve Coast, the founder of OSM) that utilizes the Adobe Flash Player 10. It includes Yahoo imagery and establishes a highly dynamic editing environment that displays all the nodes that comprise lines and polygons in the selected area (Figure 139). Topological edits are made by selecting existing nodes and links and splitting or snapping.

For example, the Mapzen documentation demonstrates how to split or break a road, as follows:

> "To split or break a road or other way (streets and roads are called ways in OSM jargon), first select the road, then select a node that is part of the road. A hint box will appear when you have selected a node that is part of a way. Choose "split" to split a one road into two joined roads. Choose "break" to break a road into two different, unjoined roads."

Figure 139 Mapzen from CloudMade is an easy to use, Flash based OSM editor

OSM has made a significant impact on the mapping world, so much so that Esri provides an ArcGIS extension for downloading, editing and uploading OSM data (Figure 140).

[27] See http://sourceforge.net/projects/gpsbabel/files/
[28] Comparisons of the capabilities of the major editors can be found at http://wiki.openstreetmap.org/wiki/Comparison_of_editors

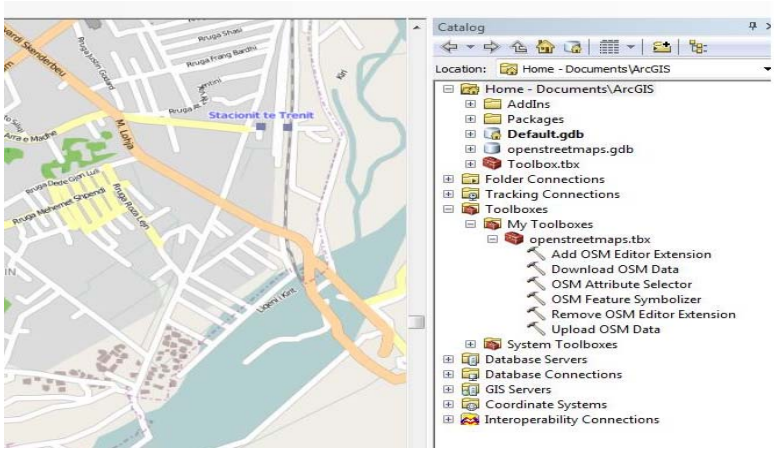


Figure 140 Esri's OSM editor for ArcView

OSM data is uploaded and tagged as geographic primitives, called GeoObjects. There are three data types that may be tagged: nodes, polylines (paths), and areas (closed paths). Roads are polylines with tags such as “highway = primary, oneway = no, and name = M40.” There are 4 different types of tags that frequently appear:

- Element Tags: These tags state one bit of information each (e.g. oneway=yes, lanes=2).
- Pre-defined Tags: These tags are shortcuts which have the value of many element tags. (e.g. “highway=motorway”, which would be the same as “lanes=3 oneway=yes maxspeed=65 etc.”).
- Tweaking Tags: These tags can be used to make a pre-defined tag correction for specific situations (e.g. “motorcycle=no” may be used along with “highway=motorway” where motorbikes can’t go. “lanes=2” may be used for smaller motorways, such as ones commonly found in Scotland).
- Rendering Tags: These are tags that are unrelated to what is being mapped, but rather are there to tell the renderer how to render something (e.g. layer=1).

Unfortunately, contributors to OSM are free not only to use the recommended conventions for tags that have been established by OSM, but also may use tags that are non-standard, but perhaps meaningful to the individual contributor. In fact, contributors are requested to map the topics of interest to them and tag them as they feel is appropriate. OSM’s recommended feature tagging catalog can be found at http://wiki.openstreetmap.org/wiki/Map_Features. The variability of tagging and the UK orientation of the OSM suggested tags are two of the complications in OSM that hinder its use internationally.

3.5.2.3.2 Error Correction

OSM is a pure crowdsourced system and regarded as self-healing over time. In general, the data viewed are the latest data contributed, even though the view may contain erroneous data that replaced accurate database information. In the case of conflict, those entering data are asked to make a choice between the data they entered and the current data being used by OSM. While data entry can be a collaborative effort in which people discuss why a specific change is the best,

it is not necessary to be collaborative, although someone may revert OSM back to an earlier state to correct particularly egregious errors.

Automated edits (e.g. Xybot[29], etc.) are part of the OSM process for cleaning data, but are governed by a voluntary "Code of Conduct" whose main rule is to "respect the work of others".[30] Not all of the OSM community is in favor of automated edits and they are used sparingly because the basic notion of crowdsourced data is that someone will find and fix errors in the data. In addition, the life blood of the system is the work of contributors who might become irked if hours of work were removed by a "bot" that was not carefully written.

A number of external parties track the bugs in OSM and keep track of them for those wanting to find and correct problems in the map database. Skobbler[31] is one example; although there are others (see Figure 141).

Figure 141 Skobbler is an example of an application that tracks bugs (errors) in the OpenStreetMap database

3.5.2.3.3 More on OpenStreetMap

A final note about OpenStreetMap relates to the fact that OSM is a form of active community input. It takes a considerable degree of effort and interest for a contributor to become capable with the system. One measure of an active crowdsourced system is the number of contributors it attracts and in this respect OSM has been very successful.[32] (See Figure 142) Another measure is how quickly the map coverage increases and by this measure OSM is also successful.[33] (See Figure 143) However, an increase in coverage and more registered users should mean more contributions and adherence to the thought that more eyes make map errors shallow.

[29] See http://wiki.openstreetmap.org/wiki/User:Xybot
[30] See http://wiki.openstreetmap.org/wiki/Automated_Edits/Code_of_Conduct
[31] See http://www.skobbler.co.uk/
[32] See Steve Coast's presentation at the USGS VGI Meeting, 2010 - http://cegis.usgs.gov/vgi/results.html
[33] See Haklay and Ellul, 2009

Unfortunately, research by Haklay has shown that the majority of data contributed to the OSM database of the United Kingdom comes from limited number of contributors.[34] (See Figure 144) In fact, as of 2009, fifty percent of the nodes in the OSM UK database were created by less than 30 contributors. While it is possible that these super-users have considerable knowledge of the geography of the UK, it is more likely that they were simply digitizing the imagery donated by Yahoo for use in the project. One of the benefits touted for crowdsourced data used for the creation of spatial databases is that it benefits from the contributions of local people with knowledge of local places. The statistics for OSM contributions, argue that large parts of the OSM UK database have not been created by those with local knowledge and, thus, may not reflect the best that crowdsourcing has to offer.

Perhaps more discouraging is Haklay and Ellul's finding that spatial attributes are sorely lacking in the OSM's UK database.[35] (See Figure 145) It is likely that the relative absence of attributes is related to GPS traces being the preferred method of capturing and providing data to OpenStreetMap.

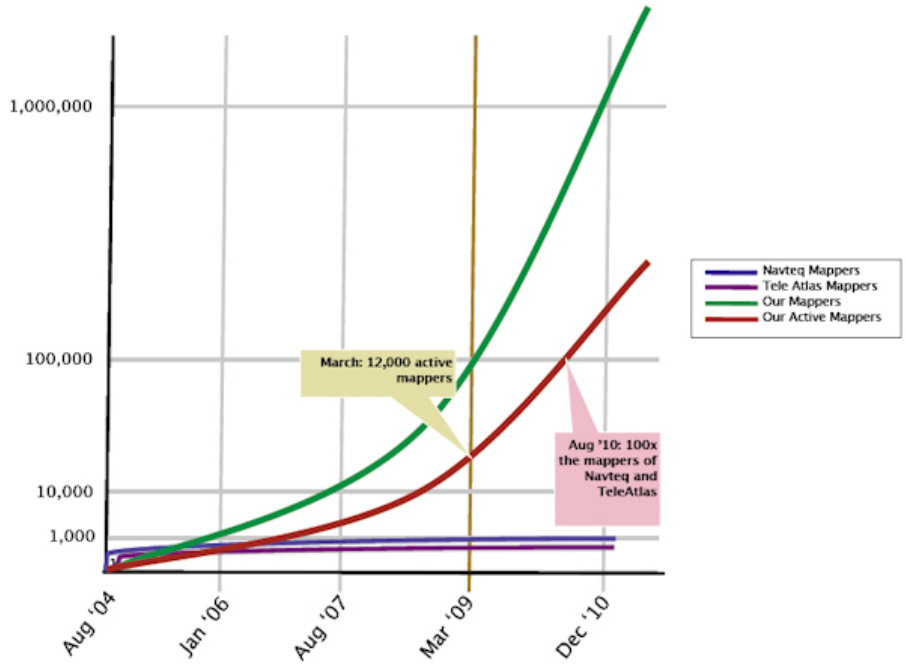


Figure 142 Number of contributors to OSM and other mapping services (Steve Coast, 2010). Coast did not point out that the vast majority of contributors to OSM contribute once and do not contribute ever again. We were unable to verify his claims for the number of users of Tele Atlas and NAVTEQ, but believe them to be erroneous by several orders of magnitude.

[34] Haklay, M, 2009, http://povesham.wordpress.com/2009/01/12/osm-quality-assessment-s4-presentation/
[35] Haklay and Ellul, 2009.

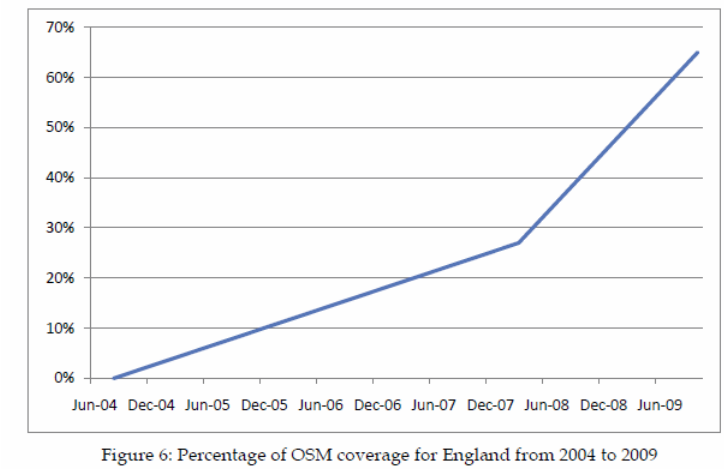


Figure 143 The amazing growth in coverage exhibited in the OSM United Kingdom database (source Haklay and Ellul, 2009l)

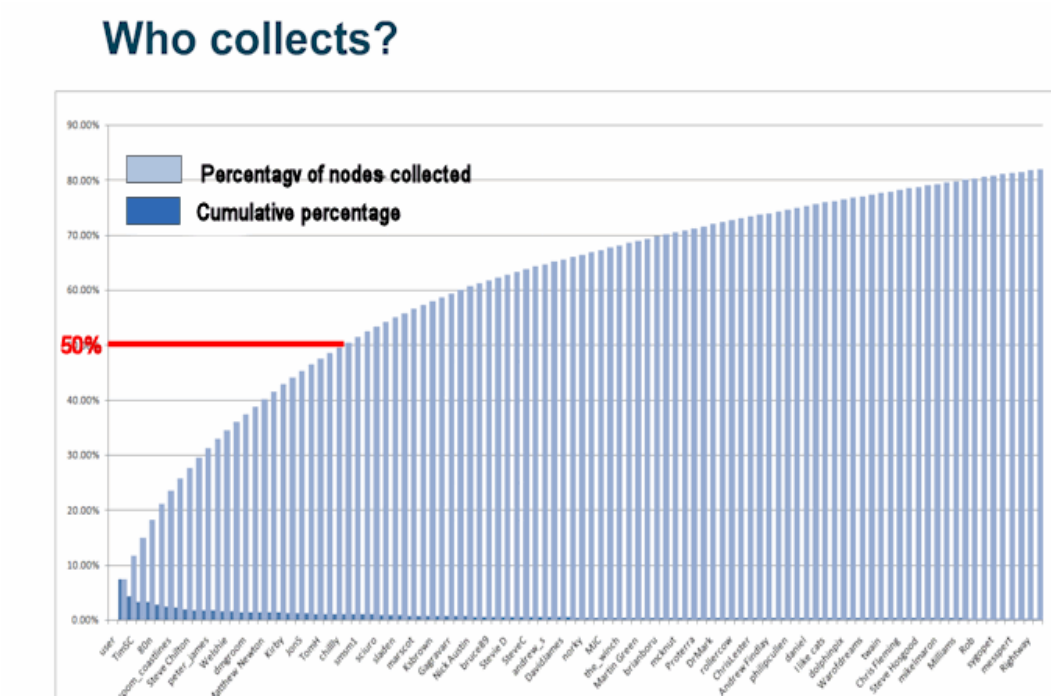


Figure 144 It appears that approximately 50% of the data in the OSM United Kingdom database were contributed by less than 30 of the contributors, raising questions on whether OSM is a good example of the notions behind crowdsourcing. (Graph from Haklay 2009)

	Meridian 2	OSM 03/08	OSM 03/09	OSM 10/09
Total length (m)	261,678,924	169,608,142	267,950,190	324,127,812
Total length – with attributes	205,621,010	54,581,704	95,332,646	123,632,213
Empty grid squares	20,935	56,481	29,444	20,700
Number of 'complete' cells	112,385	30,702	57,622	73,092
Number of 'complete' cells – with attributes	92,467	6,670	13,595	19,719

Figure 145 Haklay and Ellul compared the OSM UK database with the Meridian 2 database provided by the Ordnance Survey and found that the number of "complete" cells that had been attributed by OSM contributors was relatively minor compared to the number complete cells that were attributed in the Meridian 3 database. (Chart from Haklay and Ellul, 2010, emphasis added by authors of this report.)

3.5.2.3.4 Summary

OSM has a simple data model and is relatively easy to use, although more complicated than entering data into the TomTom or NAVTEQ solutions. Although OSM is focused on community maps and community solutions, using the data for map display is fairly complicated. Most often, those interested in displaying OSM data, including commercial parties, often turn to companies like CloudMade[36] and Geofabrik[37] that provide easy-to-use mapping solutions for those interested in leveraging OSM data.

3.5.2.4 Google

Google uses crowdsourcing to support two distinct, but interrelated, product lines: Google Maps and Google Map Maker (by far the more interesting of the two processes). Google Maps features a "Report a Problem" button that lets the user of the map input a correction on-screen while viewing the map. Clicking the Report a Problem button opens a window that asks you to drag a marker to the location of the error and then to select a map element from a limited list (street, address, satellite imagery, one of the search results in the search panel, another place or landmark, or other). Corrections are limited to providing a short description of the problem. Google attempts to resolve these issues within 30 days and is fairly good about responding with a resolution to the issue, if you enter your email address and agree to be contacted by them (see Figure 146 below).

[36] See http://www.cloudmade.com
[37] See http://www.geofabrik.de/

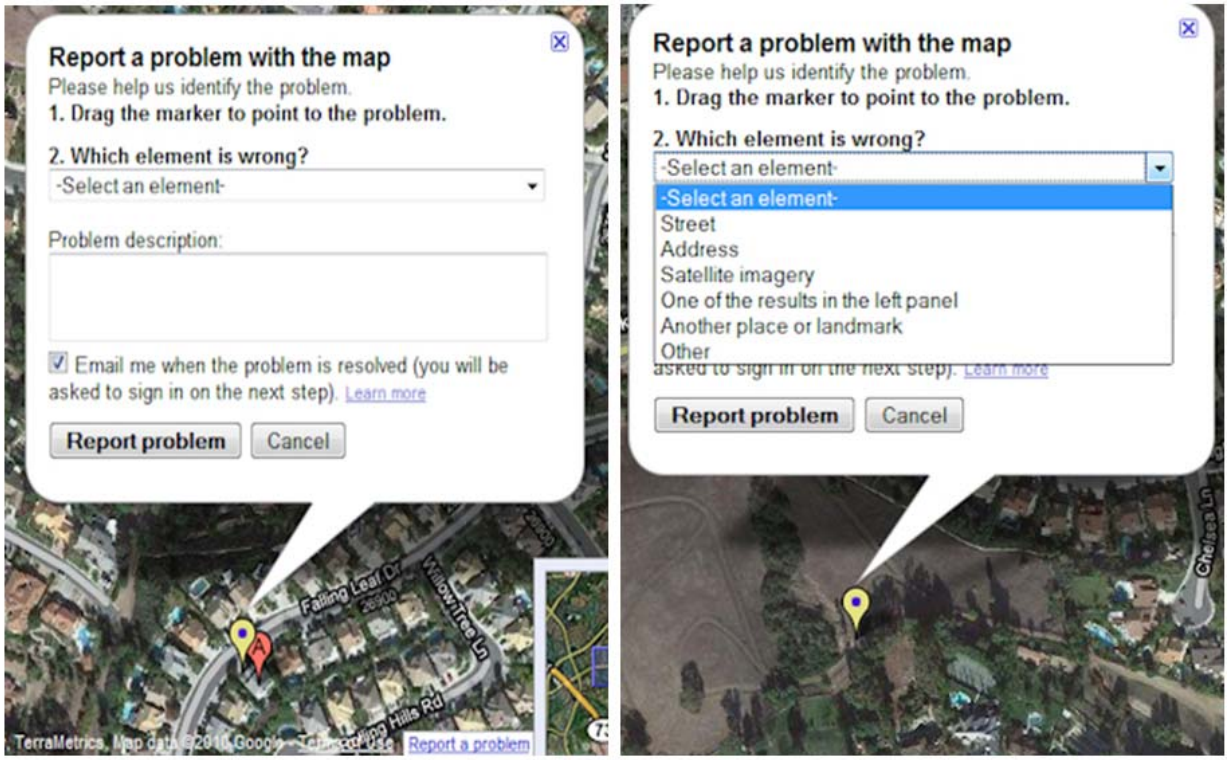


Figure 146 Example of the Report a Problem function on Google Maps

Google Map Maker is Google's attempt to use crowdsourcing to help compile maps of countries for which their existing providers lack coverage.[38] Currently, Google is using crowdsourcing to provide map coverage in 181 countries around the word (Figure 147).

Figure 147 Google Map Maker is available for 181 countries around the world.

Google Map Maker is designed to allow local experts to create new map features by adding roads, buildings, businesses, parks, lakes and other features that would make a richer map (see Figure 148). Map Maker is designed to be a collaborative process much like OpenStreetMap, but

[38] See the Google Map Maker at http://www.google.com/mapmaker . If your map location is zoomed in on the United States you may not see an image, since the U.S. is not covered by Map Maker.

the difference is that Google holds a copyright in and to each of the works that make up My Maps, so the individual's contributions benefit Google, perhaps, more than they benefit the users. However, users we have queried want their country to be shown in detail in Map Maker (and possibly in Google Maps) and are willing to contribute their time and effort to achieve that goal.

Figure 148 The 2009 version of Google Maps (left) was less detailed than the 2009 Map Maker version of the city. In 2010 the coverage and detail appeared identical on both platforms.

Google Map Maker provides its users with selection tools that allow the contributor to add features to blank pallets or to edit map information that needs correction, augmentation, or deletion. See Figures 149 – 152 for more details.

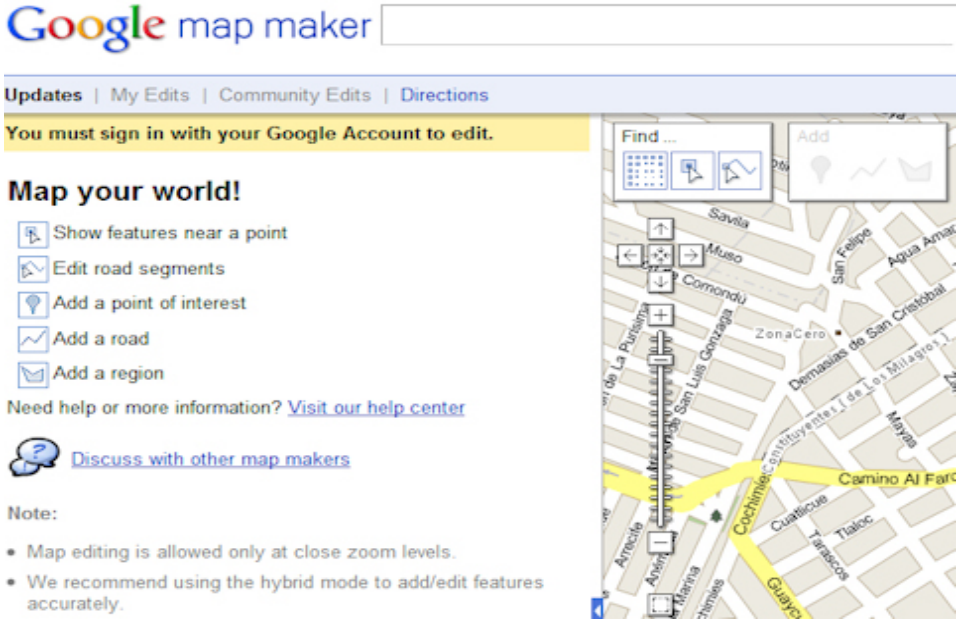


Figure 149 The Google Map Maker tools give the user significant abilities to create a full set of map features

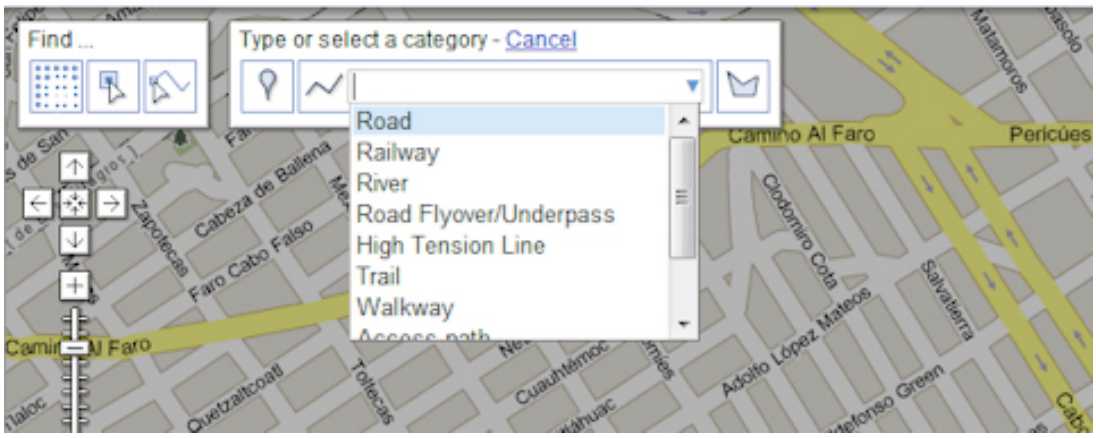


Figure 150 More details on the Map Maker Tools showing the types of line work that can be entered

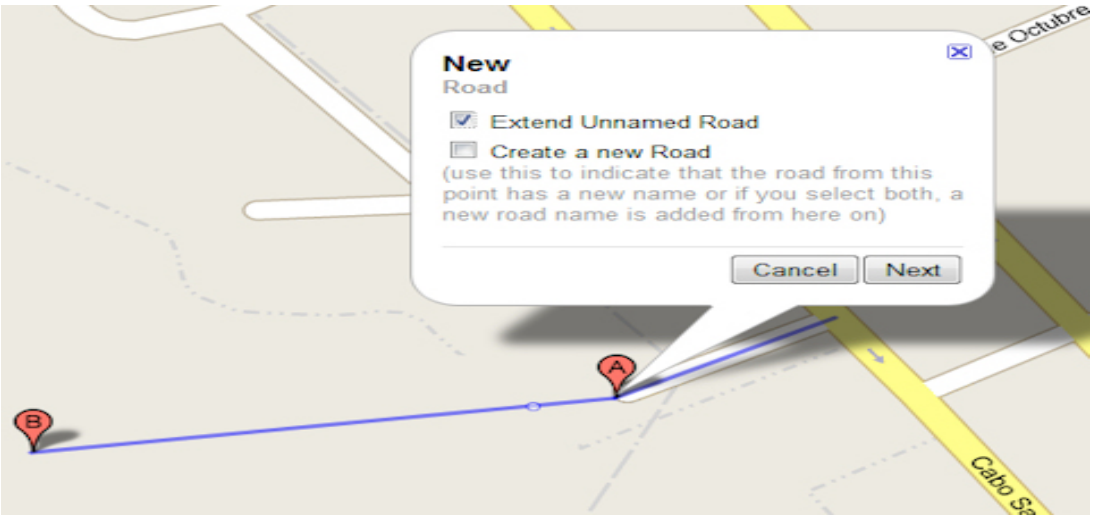


Figure 151 Extending a road using Map Maker

Google Map Maker maps are designed to be edited by the community and include overview information about the features, attributes, description, events (held at a location) and the edit history of features. Once you have edited and saved your work it may appear instantly, or if there is a moderator for the area, your changes must be approved by the moderator (you cannot moderate you own work). Moderators are members of the community who claim to have significant knowledge about specific areas or neighborhoods. Edits that occur in the "neighborhood" of a moderator must be approved by them before it goes live and the contribution will have a pending status until that time.

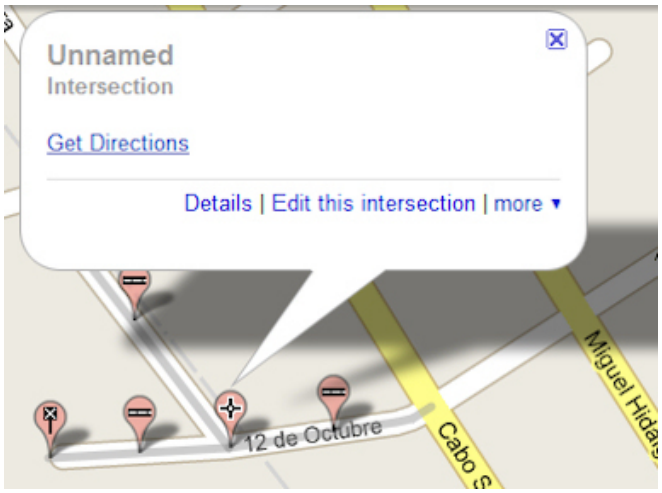


Figure 152 Editing a road intersection in Map Maker

The Map Maker application, also, has a navigation functionality that was put in place mainly to assist in editing. The requirements for routing can be used to find broken roads, roads with incorrect attributes, disconnected places and other map errors sensitive to the practicalities of computing connected routes. Google has provided its Map Maker users with a variety of robust compilation and editing tools. In addition, Google provides the best imagery that they can source for the areas covered by Map Maker in order to assist the crowdsourcers' desires to create detailed, up-to-date maps of their local areas.

As noted previously, Google has reserved the right to publish any content from Google Map Maker in any venue, although Google Maps and Google Earth are the only products where it is actively used.

Before Map Maker maps cross the fence to Google Maps, Google compares the data provided by Map Maker with its resources for an area and may add or edit the data further before it decides to include the data in Google Maps. Google Map Maker has proven an exceptionally valuable tool for Google in that it helps Google incent user generated content in a manner that is pleasing to those who contribute data, as well as by creating reasonable quality maps that can be used by Google in its Google Maps application.

3.5.2.5 Summary

The data flows of each of the systems described above differ in subtle manners and comparing and contrasting these flows helps understand the advantages and weaknesses of various methods used to implement crowdsourcing.

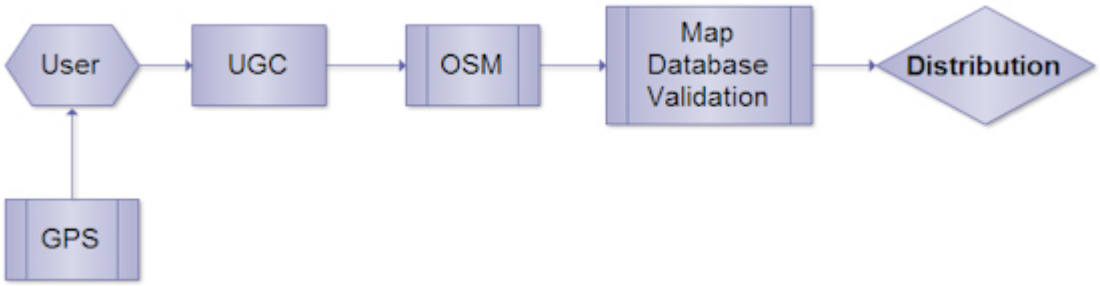


Figure 153 OpenStreetMap data flow

The OpenStreetMap dataflow (Figure 153) shows that the system benefits from its simplicity. However, it suffers from lack of local database for error corrections in the field, as can be accomplished with a TomTom PND. A competitive issue for OSM is that it does not have an option that can be used to collect data by cell phone, as is done by Google, TomTom, and soon by Nokia/NAVTEQ. Cellular GPS phone-based tracks may become an extremely useful addition to the map compilation tools of major map database developers.

A critical issue for OSM is that is appears separated from those who use its data. The OSM data model allows external companies (e.g. CloudMade and GeoFabrik) to manage its relationship with users of OSM data and this lack of a direct, wide-scale distribution of OSM data by OpenStreetMap may become a significant problem in the future. While it is in close contact with its database builders, OSM is not a use-driven database and does not necessarily receive the benefits accrued from correcting the types of map errors often discovered and communicated by map users. While distribution companies like CloudMade and GeoFabrik may elect to edit data errors they (or their users) discover and report these changes to OSM, it is not a requirement of the OSM use license to do so. While companies electing to use OSM data have a duty to provide the data they add to OSM under the same license guiding their user of the data, there is some question whether these “versions” of the OSM database will be discovered by OSM or its users, so that the new data can be incorporated into the OSM database.

At this early stage, the distribution mechanism for OpenStreetMap data is underdeveloped, which limits the use of the data for applications and limits the use of crowdsourcing (from a broad audience of users) from correcting the data. The recent association of OSM and AOL through Patch and the ability of individuals to edit OSM through Patch may become one of the standard contributor models for OpenStreetMap.

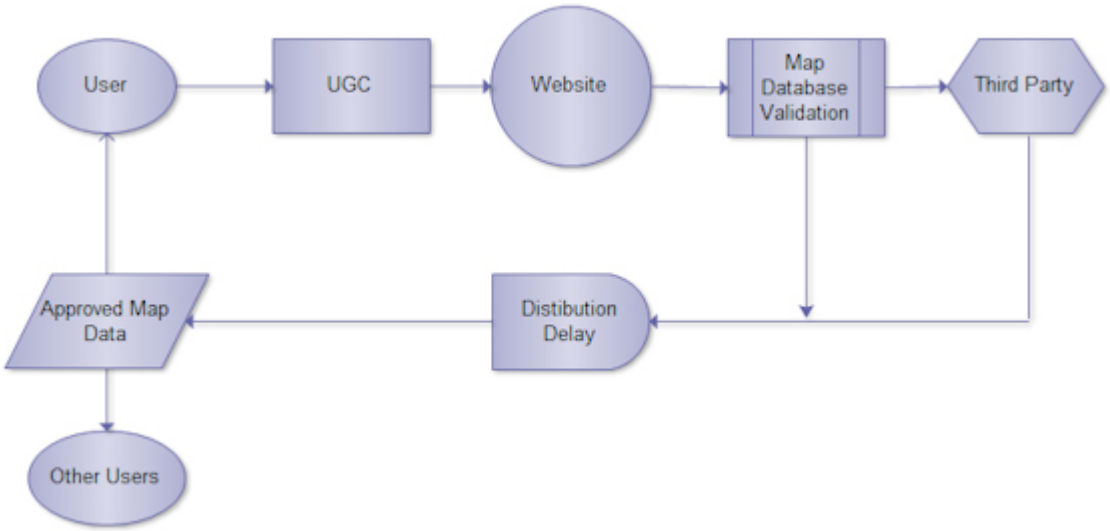


Figure 154 The "online" dataflow (used by NAVTEQ and Google)

The online dataflow (see Figure 154) used by NAVTEQ and Google is much less flexible than the PND solution used by TomTom (Figure 30). Contributors to online systems have to remember where the error they want to report occurred in the real world and, then, find it on a map and retrieve the spatial details from their memory in order to make a correction. In addition, the feedback loop is delayed due to the use of moderators. Further, users of some commercial services that use NAVTEQ data (e.g. Bing Maps) may not see their suggested changes, for considerable periods of time. Other NAVTEQ users (e.g. owners of a PND powered by NAVTEQ data, or an embedded in-car navigation system) may not see the changes unless they buy a license to a new version of the database. Finally, the online form of a crowdsourced system relies on the altruism and good will of the users. Conversely, the potential delays in “seeing” the results of their contributions may not be a good, long-term incentive. Google avoids this issue, in part, by allowing the contributed map data to be displayed in My Maps almost immediately.

We think that TomTom’s Map Share is an elegant method for processing crowdsourced data. Using this software, the user can see any map changes they make as they make them and the changed data is immediately available for navigation or map viewing. Alternatively, contributors can choose to wait for the change to be confirmed by Tele Atlas, or wait to view changes supplied by other TomTom users. In all cases, the feedback loop is more immediate and pleasing than an approach that is limited only to online correction of map errors. Perhaps the greatest benefit of the TomTom solution is that it involves a portable version of the map database, allowing the user to note errors or omissions while they are navigating in a specific geography and to mark the location of the error while it is being observed or experienced.

Unfortunately, it is unlikely that TomTom and PNDs will continue to be an important part of the map database market due to pressure from GPS-equipped mobile devices. However, Map Share

may survive in some form after being ported to mobile devices.[39] Whatever its future, its elegant method of collecting crowdsourced data is noteworthy. (See Figure 155)

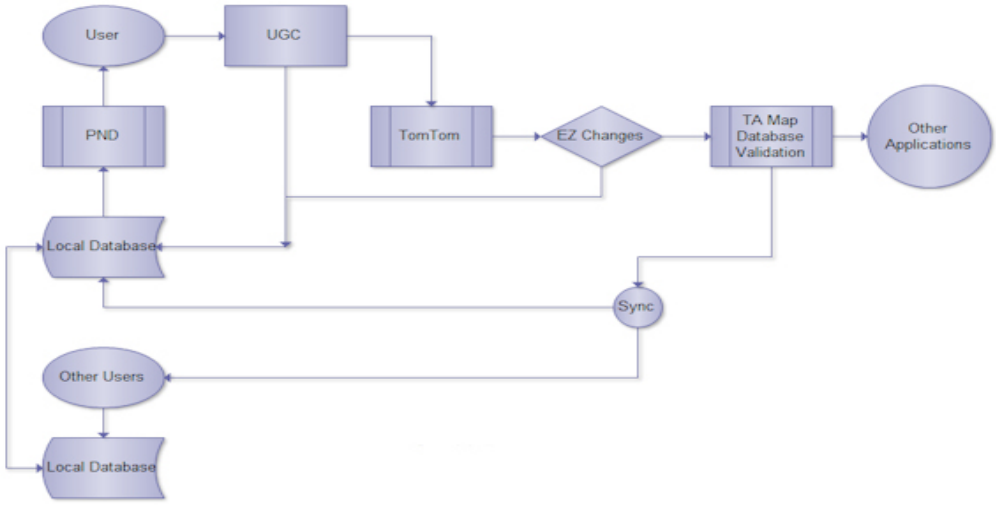


Figure 155 The data flow used by TomTom in its Map Share application for PNDs

Based on these three dataflow models, there are at least three different approaches to map compilation using the crowdsourced approach:

1. Create a unique map database from scratch based on User Generated Content (see the OpenStreetMap website for the UK).

2. Create a "new" map database by augmenting an existing public domain database with crowdsourced data (see the OpenStreetMap database for the U.S., which is based on TIGER).

3. Augment an existing commercial/proprietary database with User Generated Content, for the sole purpose of providing a more up-to-date version of the database for users of these data (See Tele Atlas, Google Maps or NAVTEQ).

The Geography Division would likely benefit from attempting to use crowdsourced data to augment TIGER. The choices it faces, however, are complex. It could decide that augmentation should come from importing or (in some form) sharing the OSM database of the United States and using it as a compilation tool for improving the MTdb. Conversely, the Geography Division could develop its own online TIGER-based crowdsourcing platform to attract crowdsourced corrections to its maps and addresses. It would seem to us unlikely that the Geography Division would consider licensing data from a commercial supplier whose data had been augmented by crowdsourcing due to potential problems in distributing these data, although redistribution of OSM data might be difficult under the ODbL, not to mention the restrictions on address data

[39] Map Share currently runs on the iPhone, but works optimally only when you buy the TomTom in-car mount that actually contains a higher-grade GPS receiver than those included in most cell phones.

imposed by Title 13. Another approach to the use of crowdsourced data might be found in pursuing a collaborative Federal spatial database that allowed for crowdsourcing, but from which, the Geography Division could reasonably withhold its address data, but benefit from the data gathering capabilities of other Federal agencies.

In any case, crowdsourced data is becoming a fundamental part of map compilation and its values and limitations should be considered by the Geography Division as it moves forward with its development plans for the 2020 Census, as well as its plans for continuous update of the MTdb.

3.5.3 A Report on the Status of Crowdsourcing Efforts by Federal Agencies

As noted previously, crowdsourcing is most successful when there is a natural relationship between a potential contributor and the website desirous of collecting and managing the contributed spatial data. For example, OSM interfaces directly with those who want to create a free, worldwide map database, while TomTom's Map Share is an interface with their customers who want to contribute corrections to the maps they use for navigation. In each example the contributor is taking an action that produces benefits for them in the form an improved free map (OSM) or improved routing and address finding (TomTom). The question for agencies of the Federal Government is "how will their use of crowdsourcing directly benefit the contributors, as well as the agency?"

Budhathoki et. al (2010) in a study of the motivation behind contributors to OpenStreetMap indicated that concerns for the usability and sharing of local knowledge with social networks, self-learning, and the desire for free and open map databases, were key motivators.[40] Whether these types of motivation could or would be applicable to helping meet the needs of Federal Agencies to improve map databases remains unclear and a potentially fertile area for further research.

Alternatively, Federal Agencies might consider using the data collected by organizations such as OSM as part of their updating strategy, rather than mount their own effort at using crowdsourcing to update their spatial databases. Licensing issues for using OSM data are discussed in more detail in the section on crowdsourcing in Deliverable 2.[41]
Our report on the use of crowdsourcing below is focused on the U.S. Geological Survey, an organization that appears to have taken the lead position in examining the potential applications of user generated content in the maintenance and updating of its spatial databases.

3.5.3.1 United States Geological Survey

The USGS held a meeting on Volunteered Geographic Information (VGI) in January 2010 to gather information on the potential use of crowdsourcing to support *The National Map*, the

[40] Budhathoki, et. al. 2010. "Who Map in OpenStreetMap and Why?"
[41] Syneren Technology, 2010, Reporting the State and Anticipated Future Directions of Addresses and Addressing, A Report to the Geography Division, U.S. Bureau of the Census Deliverable #2 – Syneren Technologies Contract – Task T005.

USGS source of topographic information for the nation (Figure 156).[42] The USGS National Map Corps, a volunteer group who used an online system to contribute location and attribute information about point features had been put in hiatus in mid-2008 due to difficulties in managing this effort. In part, the intent of the organizers of the January meeting was to solicit input from others to see if the knowledge would be of any relevance for resurrecting the National Maps Corps in a manner that would benefit the USGS and the contributors. In essence, the meeting was part of a USGS process aimed at developing a new business plan for effective and efficient incorporation of volunteered geographic information into *The National Map,* given the rapidly changing technical landscape and the mandates for more transparency in government.

The VGI Meeting was sponsored by the Center of Excellence for Geospatial Information Science (CEGIS) of the USGS. Established in 2006, the mandate of CEGIS is to conduct, lead, and influence the research and innovative solutions required by the National Spatial Data Infrastructure (NSDI), as well as to monitor the continuing emergence of the Geospatial Web.

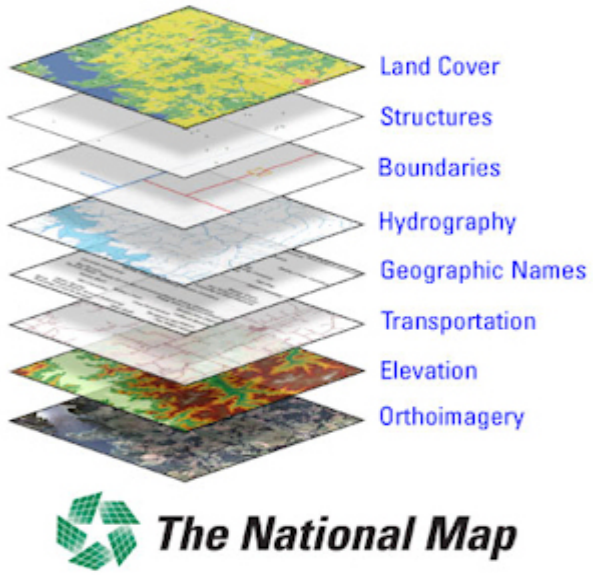


Figure 156 Contents of The National Map (TNM)

One of the interesting issues discussed at the USGS meeting on VGI was the potential problems of using OSM due to licensing issues. The possibility was noted in a presentation by Larry Moore[43] that OSM data cannot be used directly because the OSM license (Open Database License (ODbL))[44] is more restrictive than TNM. (See Figure 157)

[42] For details on the meeting and some of the presentations at the meeting see this USGS website http://cegis.usgs.gov/vgi/
[43] Moore. L, 2010, OSM Implementation Diagram, USGS VGI meeting. See http://cegis.usgs.gov/vgi/results.html
[44] See http://www.opendatacommons.org/licenses/odbl/

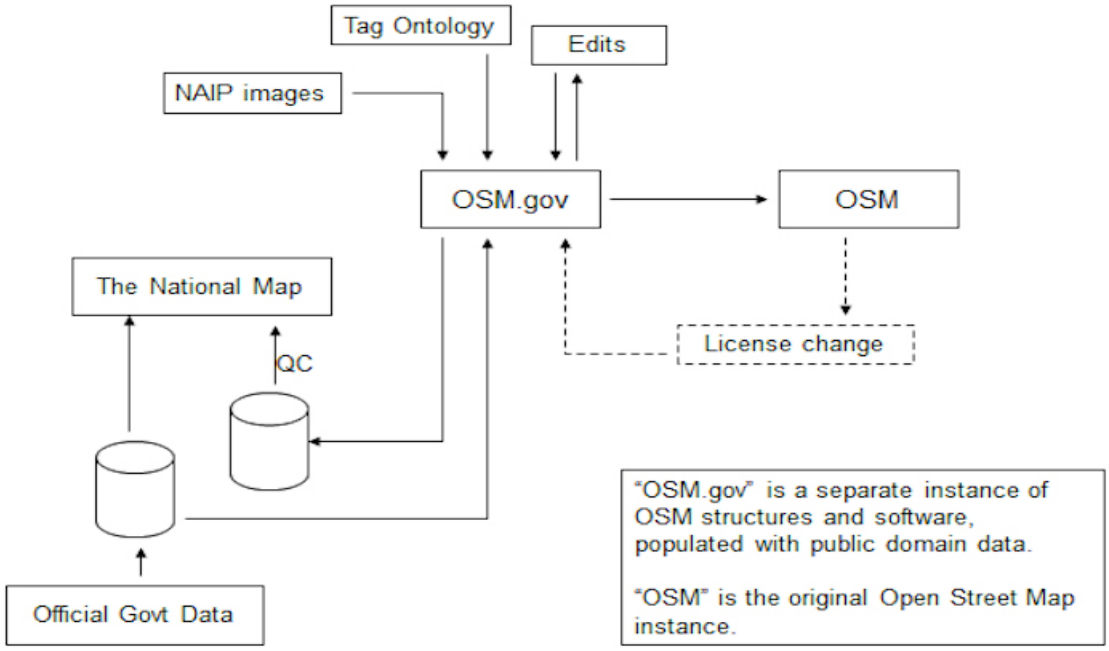


Figure 157 Proposed OSM implementation Diagram (Moore, 2010)

In addition, there was discussion of the issue of "Federated" identity concerning how to manage users and identities across the OSM technology stacks.

3.5.3.1.1 USGS VGI

Based on the meeting, the USGS proceeded with what was called an *OpenStreetMap Collaboration Project* (OSMCP) to determine whether two or more government agencies could collaborate on editing roads using OpenStreetMap (OSM) technology while applying common standards. The USGS investigated this issue in collaboration with the Kansas Data Access and Support Center (DASC) to learn what potential value crowdsourcing might have to both groups. The USGS tested the capability prototype in Johnson and Douglas counties Kansas. The result of the effort was that the participants were able to jointly edit the state roads in Douglas and Jefferson counties to the agreed specifications using a version of the OSM Technology Stack.

We were briefed by the USGS[45] on the project in June 2010. While the USGS refers to project as VGI and more specifically as the "OSM Capability Prototype", the project as described to us was focused on collaborative editing using OSM technology, but did not involve, at that time, crowdsourcing (see Figures 158 through161).

[45] Kari Kraun, Director of the USGS National Geospatial Technical Operations Center graciously arranged the presentation.

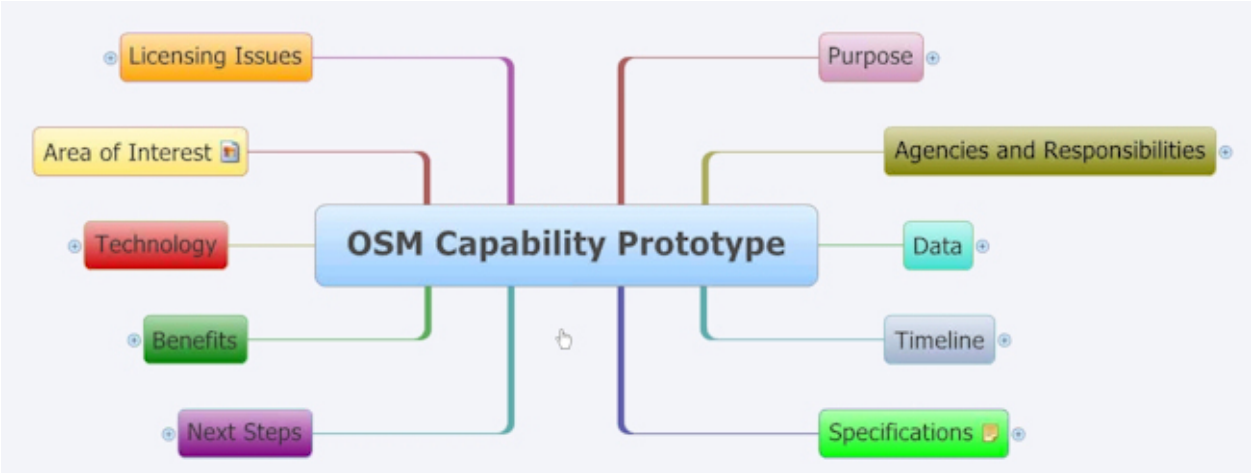


Figure 158 MindMap overview of the OSM Capability Prototype

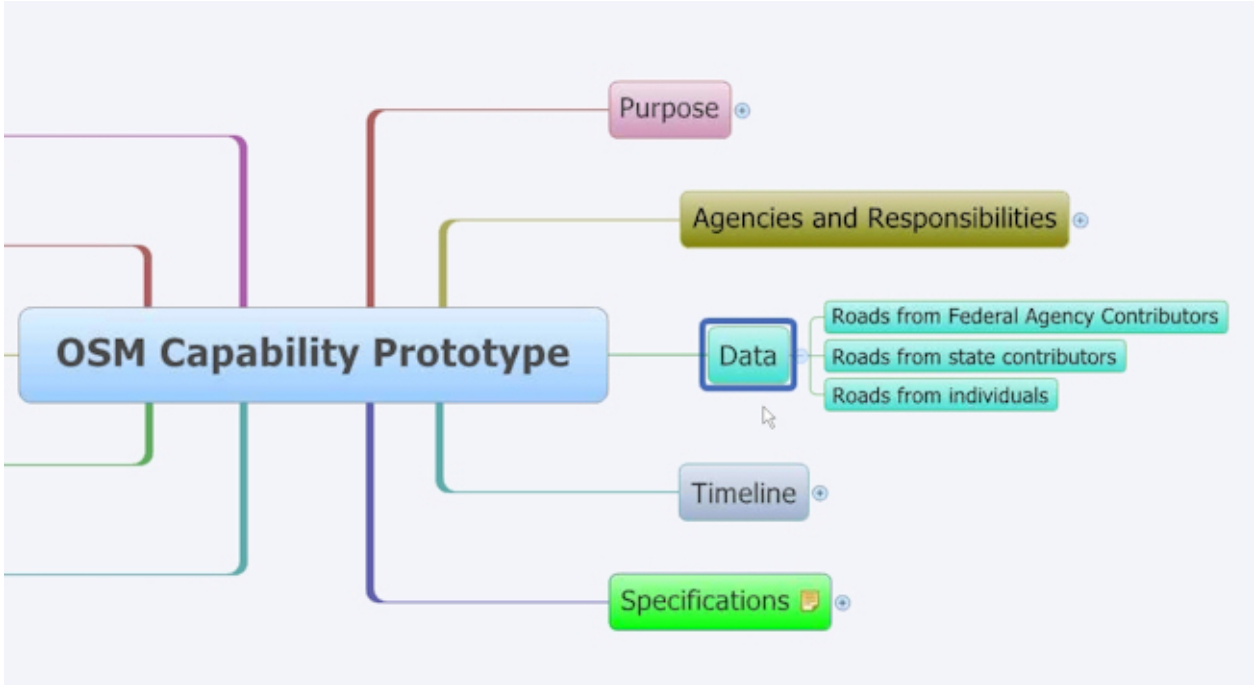


Figure 159 Although the project might include data from individuals in the future, its focus was on agency to agency collaboration

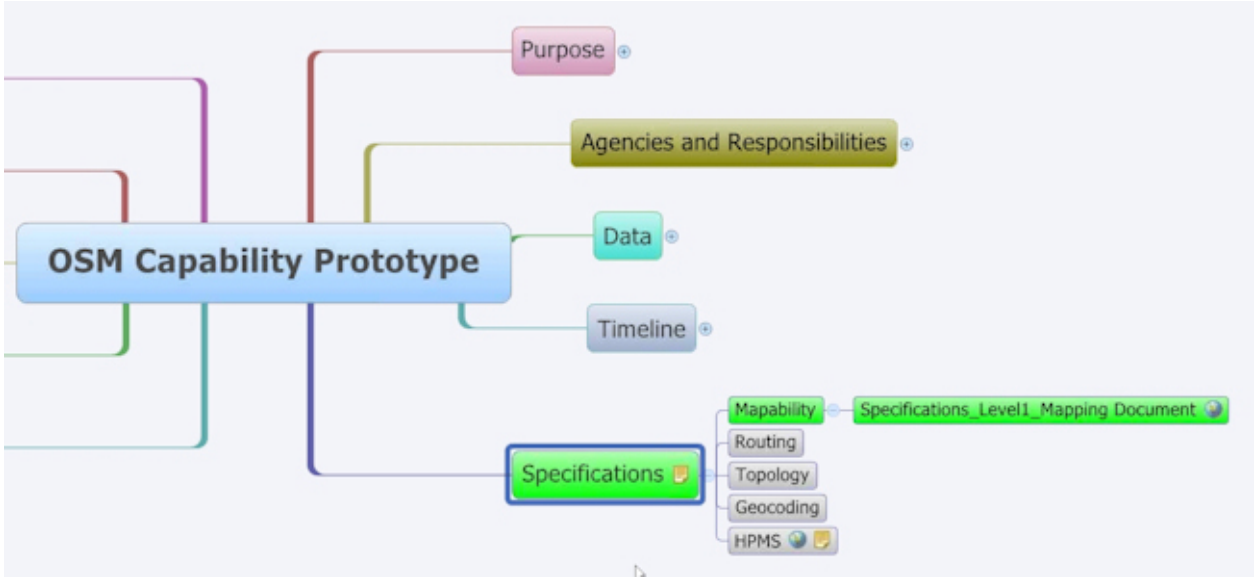


Figure 160 Details of the specifications for data interchange. The Level 1 Mapping Document was the basic road specifications for The National Atlas.

.

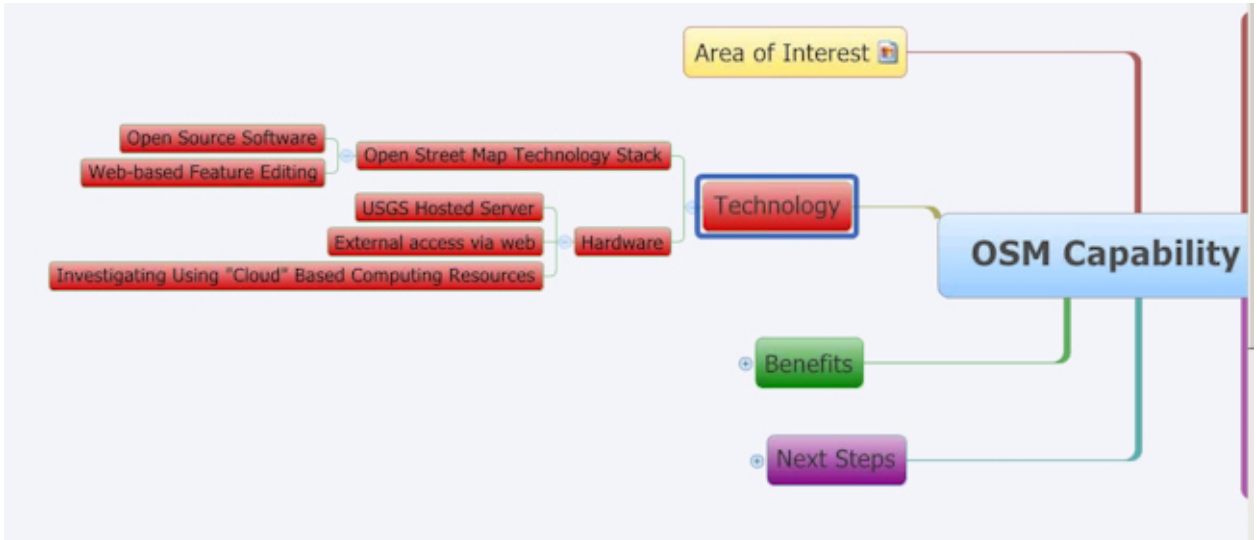


Figure 161 Details of the OSM Stack and the technology used in the project.

Based on the successful results of the OSM Capability Project it is our understanding that the USGS VGI group is planning a new project to collect structure data in the City and County of Denver. Apparently, the USGS plans to incorporate crowdsourcing provided by GIS clubs in Denver, Colorado in a project that is designed to collect what is described as 28 different Community Anchor Points (e.g. names of schools, fire stations, etc.). The goal is that the data from the project will end up in *The National Map*. USGS is said to have some interest in creating a collaborative version of *The National Map*, but our understanding is that its activities in this area will be hindered by budget limitations in 2011.

3.5.3.2 The FCC, NTIA and Crowdsourcing

In addition to the work by the USGS, we note that numerous states are mapping addresses associated with the FCC's Broadband Initiative based on grants provided by the National Telecommunications and Information Administration (NTIA).[46] In general, the FCC's goal for the address mapping is not uniform national coverage, but sampling to determine where broadband coverage does or does not exist, as well as its characteristics (download speed, upload speed, jitter, etc.). Interestingly, some portion of the address data collection effort appears to be based on crowdsourcing, where individuals can provide their address and have their broadband speed tested.[47] In response to the NTIA's grant program, three states indicated that they would use some form of crowdsourcing to gather address data. Utah proposed using a crowdsourcing application for address development through a public-private relationship with Blue Stakes of Utah. The State of Colorado indicated that it planned to collect and add to current county data and deploy a crowdsourcing tool to improve address file quality. Rhode Island indicated that its Office of Information Technology would collect and improve current county data and deploy a crowdsourcing tool to improve address file quality. Unfortunately, the methods that these states would be using were not documented further and substantial progress had not been reported by the time this document was finalized.

As noted in Deliverable 2, as part of its Broadband Initiative, the FCC offers the consumer a test of their broadband connection speed at http://www.broadband.gov/. The website visitor is required to provide their current address of the location where they are accessing the Internet and, in turn, receive an evaluation of the speed of their broadband connection. This win-win situation provides the user information about the speed of resource and provides the FCC with an address and an Internet speed record that will help them map the state of broadband across the United States.

3.5.3.3 FEMA

Through the GSA, the Government Challenge platform[48] is an attempt to use a crowdsourcing partnership between the government and citizens to bring the best ideas and top talent to bear on national problems. While this attempt at crowdsourcing does not appear successful (the levels of response were extremely low), we note that FEMA pages include a contributed discussion in the section on *Preparing our Communities Before a Disaster Strikes* titled "*Encourage Community Mapping*." [49]

[46] See the FCC Broadband Initiative page at http://www.broadband.gov. See the NTIA website for more information on the NTIA grants at http://www2.ntia.doc.gov/ . See also Deliverable 2 for more information on this project.
[47] See this page at Broadband.gov for more information http://www.broadband.gov/qualitytest/about/ .
[48] See http://challenge.gov/about
[49] See http://challenge.gov/challenges/87/forum_topics/28

3.5.3.4 Summary

Crowdsourcing appears to be of significant interest to various agencies of the Federal Government, but it is not yet a functionality that is commonly practiced by the agencies. It is likely too early in the game for the wholesale adoption of crowdsourced methods to assist government agencies with meeting their mission goals.

We recommend that the Geography Division further consider the uses of crowdsourcing as a component in its spatial database building efforts. (See Figure 162) Successful compilation models for spatial data are based on harnessing the advantages of a variety of data gathering and assessment techniques. Data partnerships are an extremely important critical success factor and crowdsourcing should be considered a viable component of this mix. While crowdsourcing is not a perfect method and perhaps should not even be considered anything other than a secondary method of data collection, it does provide useful data that can inform change detection programs. In closing, we note that one major concern with the use of crowdsourcing for the compilation of spatial data is that the efforts of crowdsourcers seem to be focused on urban and not rural areas. In part this is a reflection of the fact that those who are most interested in crowdsourcing spatial data have an interest in maps and some facility with spatial data. Citizens with these talents are a small part of the population in general and more likely to be found in urban than in rural areas, simply as a function of population distributions patterns.

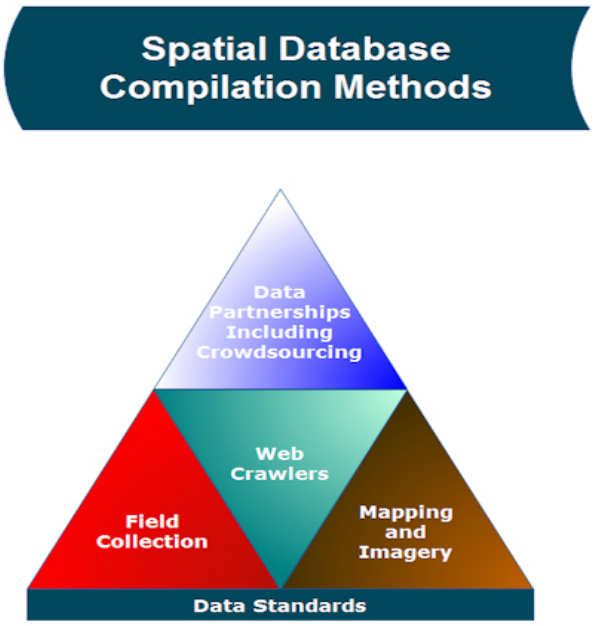


Figure 162 The compilation of Spatial Data can benefit from crowdsourcing

References

Anderson, J. et al. 1976. *A Land Use And Land Cover Classification System For Use With Remote Sensor Data.* Geological Survey Professional Paper 964 Washington DC: United States Government Printing Office.

Awrangjeb, M. et al. 2010. *Automatic Detection Of Residential Buildings Using LIDAR Data And Multispectral Imagery,* ISPRS Journal of Photogrammetry and Remote Sensing 65 2010. 457-467.

Blaschke, T. 2010. *Object Based Image Analysis For Remote Sensing* ISPRS Journal of Photogrammetry and Remote Sensing 65 2010. 2-16.

Budhathoki, Nama R., Muki Haklay and Zorica Nedovic-Budic. 2010. *Who Map in OpenStreetMap and Why?* State of the Map 2010. Atlanta, GA, 2010 (PowerPoint presentation).

Butkiewicz, T , Chang R, Wartell Z, and Ribarsky W. 2008.*Visual Analysis and Semantic Exploration of Urban LIDAR Change Detection.* Eurographics/ IEEE-VGTC Symposium on Visualization Volume 27, Number 3.

Budhathoki, Nama R. and Zorica Nedovic-Budic. 2010. *How to motivate different players in VGI?* Zurich, SZ: GIScience.

Butcher R., 2010. *Assisted Change Detection with MATISSE.* EuroSDR Workshop - Automated Change Detection For Updating National Databases Southampton, UK 4th to 5th March. (http://www.eurosdr.net/workshops/cd_2010/)

Champion, N. 2007. *2D change detection from high resolution aerial images and correlation digital surface models*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI, Part 3/W49A: 197–202.

Champion, N., Rottensteiner F., Matikainenc L., Liangc X, Hyyppäc J., and B. Olsend 2009. A Test Of Automatic Building Change Detection Approaches Commission III, WG III/4

Cheng, W., Zhang, X., Wang, K., and Dai, X. 2009. *Integrating classification and regression tree (CART) with GIS for assessment of heavy metals pollution*. Environmental Monitoring and Assessment, 2009. 158:419–431.

Chiang, Y. 2010. *Harvesting Geographic Features From Heterogeneous Raster Maps*. Dissertation Computer Science, University Of Southern California.

Congalton, R. 2010. *Remote Sensing: An Overview*. National Research Council Workshop on New Research Directions for the National Geospatial-Intelligence Agency White Papers May 17-19, 2010, Workshop National Academies of Sciences.

Delafontaine, M., Nolf, G., van de Weghe N., Antrop, M. and de Maeyer 2009. *Assessment Of Sliver Polygons In Geographical Vector Data.* International Journal of Geographical Information Science, 23: 6, 719 - 735.

Devillers, R. et al. 2007. *Towards Spatial Data Quality Information Analysis Tools For Experts Assessing The Fitness For Use Of Spatial Data*. International Journal of Geographical Information Science Vol. 21, No. 3, March 2007, 261–282.

District of Columbia Geographic Information System Steering Committee, 2010 Geographic PowerPoint Presentation to Information System Steering Committee May 14, 2010

Dobson, Michael W. 2007. *Local Search Meets Social Search*. MindCommerce Publications.

Domenech, E. et al. 2010. *Automatic Image Classification Techniques Applied To Land Use/Land Cover Cartography*. Workshop - Automated Change Detection For Updating National Databases Southampton, UK 4th to 5th March. (http://www.eurosdr.net/workshops/cd_2010/)

Doucette, P., Grodecki J. Clelland R. Hsu.A. Nolting J. Malitz S., Kavanagh C. Barton S., and Tang, M., 2009. *Evaluating Automated Road Extraction in Different Operational Modes, in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery* XV. Edited by Shen SS. Lewis, Paul E. Proceedings of the SPIE, Volume 7334 2009. pp. 73341A-7334A-12.

Elhami, S. 2010. Email correspondence. April 4, 2010.

Federal Geographic Data Committee. 2009. Annual Report. Available at http://www.fgdc.gov/fgdc-news/2009-fgdc-annual-report.

Fitzgerald, B. 2003. *Hybrid approach to infrastructure change detection using high resolution satellite imagery, texture measures and GIS Ancillary Information*, MS Thesis Department of Geography, U of South Carolina

Feuerstein, M. 2010. cited in *What Facebook Places Really Means to LBS*. Kevin Dennehy http://www.gpsworld.com/lbs/what-facebook-places-really-means-lbs-10438

Frontiera, P., Larson R. and Radke J. 2008. *A comparison of geometric approaches to assessing spatial similarity for GIR'*, International Journal of Geographical Information Science, 22: 3, 337 -360.

Girres, Jean Francois and Guillaume Touya, 2010. *Quality Assessment of the French OpenStreetMap Dataset*, Transactions in GIS 12(4), pp 435-459

Grodecki, J. 2010. Email correspondence. November 26, 2010.

Guelat, Jean-Christophe, 2008, *Integration of user generated content into national databases – revision workflow at swisstopo*. (PowerPoint presentation). See http://www.eurosdr.net/workshops/crowdsourcing_2009/presentations/c-4.pdf

Hall, G.B. and Leahy, M.G., 2008. *Design and implementation of a map-centered synchronous collaboration tool using open source components: the MapChat project*. In: G.B. Hall and M.G. Leahy, eds. Open source approaches in spatial data handling. Berlin: Springer-Verlag, 221–246.

Hall, G.B., Raymond Chipeniuk, Robert D. Feick, Michael G. Leahy and Vivien Deparday, 2009, *Community-based production of geographic information using open source software and Web 2.0*. International Journal of Geographical Information Science, Vol. 24, No. 5, May 2010, 761–781.

Haklay, Muki, 2009. *Understanding the quality of user generated mapping – comparing OpenStreetMap to Ordnance Survey* Geodata, (PowerPoint Presentation) http://povesham.wordpress.com/2009/01/12/osm-quality-assessment-s4-presentation/

Haklay, Mordechai (Muki) and Clarie Ellul. 2010. *Completeness in volunteered geographical information – the evolution of OpenStreetMap coverage in England* (2008-2009). Journal of Spatial Information Science

Harris Corporation, 2008. *Rural Structure Change Detection*. U.S. Census Bureau Contract No: 50-Yabc-2-66005 Document Control Number: 7018305.

Heipke, P., Woodsford and M. Gerke 2008. *Updating Geospatial Databases From Images* Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book – Li, Chen & Baltsavias (eds)

Higgins, M., 2010. *The Role of Positioning Infrastructure in the Technological Future of our Profession* XXIV FIG International Congress. (Power Point Presentation)

Hodgson, M.E., X. Li, and Y. Cheng. 2004. *A Parameterization Model for Transportation Feature Extraction*," Photogrammetric Engineering and Remote Sensing, 70(12): 1399-1404.

Hodgson, M.E., Jensen J.R., Tullis, J.A, Riordan, K, and Archer, C. 2003. *Synergistic Use of Lidar and Color Aerial Photography for Mapping Urban Parcel Imperviousness*, Photogrammetric Engineering & Remote Sensing Vol. 69, No. 9, September 2003, pp. 973–980.

Hodgson, M. and Jensen J. 2010. *Airborne and Terrestrial LiDAR Technology for Assessing Topography* 41st International Binghamton Geomorphology Symposium Columbia, SC (PowerPoint presentation).

Holland, D., Sanchez-Hernandez, C. and C. Gladstone 2008. *Detecting Changes To Topographic Features Using High Resolution Imagery*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Commission VI, WG VI/4Vol. XXXVII. Part B4. Beijing 2008.

Holland D, 2010., *Change Detection Overview*, EuroSDR Workshop - Automated change detection for updating national databases Southampton, UK 4th to 5th March. (http://www.eurosdr.net/workshops/cd_2010/)

Huang, B., Zhang, L. and Wu, B. 2009. *Spatiotemporal Analysis Of Rural-Urban Land Conversion*, International Journal of Geographical Information Science, 23: 3, 379 - 398

Im, J., Rhee J., Jensen J. R., and M.E. Hodgson, 2007. *An Automated Binary Change Detection Model Using Calibration Approach*, Remote Sensing of Environment, 106(1):89-105.

Im, J., Rhee, Jensen J. R., and M.E. Hodgson, 2008, Optimizing the Binary Discriminant Function in Change Detection Applications, Remote Sensing of Environment, 1122008.:2761-2776.

Im, J., J.R. Jensen, and J.A. Tullis, 2008. *Object-based Change Detection Using Correlation Image Analysis and Image Segmentation Techniques*, International Journal of Remote Sensing, 29(2): 399-423.

Im, J., Jensen J., Hodgson M., 2008. *Object-based Land Cover Classification using High Posting Density Lidar Data*, GIScience and Remote Sensing, 45(2):209-228.

Jensen, J. R., 2007. *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd Edition, Upper Saddle River: Prentice-Hall, 592 p.

Kavzoglu, T., and S. Reis 2008. *Performance Analysis of Maximum Likelihood and Artificial Neural Network Classifiers for Training Sets with Mixed Pixels* GIScience & Remote Sensing, 2008, 45, No. 3, p. 330–342.

Kühnen A. and Sohlbach M. 2010. *Object Oriented Change Detection and Support for Map Update with Definiens eCognition*, EuroSDR Workshop - Automated change detection for updating national databases Southampton, UK 4th to 5th March. (http://www.eurosdr.net/workshops/cd_2010/)

Kumar-Mills et al. 2005. *Efficient RTK Positioning by Integrating Virtual Reference Stations with WCDMA Network* Journal of Global Positioning Systems 2005. Vol. 4, No. 1-2: 48-55.

LaMarca A. and de Lara E., 2008. Location Systems: An Introduction to the Technology Behind Location Awareness Synthesis Lectures on Mobile and Pervasive Computing 2008, San Rafael, CA: Morgan & Claypool Publishers.

Lapine, L. 2010. *A Brief History of Satellite Geodesy - October 4, 1957 to Present* National Research Council Workshop on New Research Directions for the National Geospatial-Intelligence Agency White Papers May 17-19, 2010, Workshop National Academies

Li, Y. and Briggs, R. 2006. *Automated Georeferencing Based on Topological Point Pattern Matching.* AutoCarto Proceedings Papers.

Li, Y. and Briggs R. 2009. *Automatic Extraction of Roads from High Resolution Aerial and Satellite Images with Heavy Noise* World Academy of Science, Engineering and Technology 54 2009

Li, Y. 2009. *An Automated System For Image-To-Vector Georeferencing*. Doctor of Philosophy in Geospatial Information Sciences The University Of Texas At Dallas.

Lihong, S.L. and, Y. Huang 2009. *Support Vector Machine (SVM) Classification: Comparison of Linkage Techniques Using a Clustering-Based Method for Training Data Selection* GIScience & Remote Sensing, 2009, 46, No. 4, p. 411–423.

Lo, C. 2008. *Population Estimation Using Geographically Weighted Regression* GIScience & Remote Sensing, 2008, 45, No. 2, p. 131–148.

Luo, J., Danlin Yu, D and M. Xin 2008. *Modeling Urban Growth Using GIS and Remote Sensing* GIScience & Remote Sensing, 2008, 45, No. 4, p. 426–442. DOI: 10.2747/1548-1603.45.4.426

Malm, A. 2010. *GPS and Mobile Handsets Fourth Edition.* Gothenburg, Sweden: Berg Insight.

Mayungaa, S., Zhanga, Y., and Coleman, D. 2005. *Semi-Automatic Building Extraction Utilizing Quickbird Imagery* In: Stilla U, Rottensteiner F, Hinz S (Eds) CMRT05. IAPRS, Vol. XXXVI, Part 3/W24, Vienna, Austria, August 29-30.

Mirnalinee ,T. 2007. *An Approach To Cartographic Object Extraction From High Resolution Satellite Imagery*. International Conference on Computational Intelligence and Multimedia Applications 243-248.

Mitishita, E., Habib, A, Centeno, J., Machado, J., Lay, J, Wong, C 2008. *Photogrammetric And Lidar Data Integration Using The Centroid Of A Rectangular Roof As A Control Point*, The Photogrammetric Record 23(121): 19–35,March.

Mitchell, R. 2007. *Change Detection for Geospatial Feature Maintenance Using the Correlated Land Change (CLC) Process*. Rockville, MD: MDA Federal Inc

Mitchell, R. 2010. *Maintaining National Geospatial Datasets with Change Detection*. Web Powerpoint Presentation to the National Geospatial Advisory Committee

Mitchell, R. 2010. Email correspondence. November 11, 2010.

Praing, R. and Schneider, M. 2009. *Topological Feature Vectors For Exploring Topological Relationships*, International Journal of Geographical Information Science, 23: 3, 319 - 353.

National Digital Orthophoto Program, 2010, Minutes of Fall NDOP Steering Committee Meeting November 16-18, 2010

National Research Council, 2007 *National Land Parcel Data: A Vision for the Future,* Washington, DC: National Academy Press.

National Research Council, 2010a *New Research Direction for the National Geospatial – Intelligence Agency* Washington, DC: National Academy Press.

National Research Council, 2010b *Precise Geodetic Infrastructure: National Requirements for a Shared Resource*. Washington, DC: National Academy Press.

O'Connor, A. 2011 Email January 6, 2010

Qiu, F, Woller, K and Briggs, R 2003. *Modeling Urban Population Growth from Remotely Sensed Imagery and TIGER GIS Road Data* Photogrammetric Engineering & Remote Sensing Vol. 69, No. 9, September 2003, pp. 1031–1042.

Renslow, M. 2010. *The Impact of Technology Development, Innovation, and Nontraditional Mapping Applications* ASPRS, San Diego April 26. (PowerPoint)

Rindfuss, R., et al. 2004 *Developing A Science Of Land Change: Challenges And Methodological Issues*, PNAS 2004, September 28, 2004, vol. 101, 39. p 3976–3981.

Rodrigue, K. 2008. *Oblique Imagery and Change Analysis Efficiency, Accuracy and Public Service Above and Beyond the Call* PowerPoint Presentation (http://www.pictometry.com/government/RESOURCES/Oblique%20Imagery%20and%20Change%20Analysis%20%20Efficiency,%20%20Accuracy%20and%20Public%20Service.ppt)

Schrock, G. 2008. *RTN-101: Mapping Part 14.* The American Surveyor, September, 2008. Available online at http://www.amerisurv.com/content/view/5357/153/. Accessed June 21, 2010.

Smith, D. 2007. *The Future of Global Navigation Satellite Systems: Anyone, Anywhere, Anytime, Any Accuracy by 2027* (http://celebrating200years.noaa.gov/visions/gnss/welcome.html)

Surowiecki, James. 1967. *The Wisdom Of Crowds: Why The Many Are Smarter Than The Few And How Collective Wisdom Shapes Business, Economies, Societies And Nations*, New York. NY: Random House.

Tullis, J., Jensen J., Raber G. and A. Filippi 2010. *Spatial Scale Management Experiments Using Optical Aerial Imagery and LIDAR Data Synergy*. GIScience & Remote Sensing, 2010, 47, No. 3, p. 338–359.

Wang, H. et al. 2010. *Building a Dynamic, Large-Scale Spatio-temporal Vector Database to Support a National Spatial Data Infrastructure in China* GIScience & Remote Sensing, 2010, 47, No. 1, p. 135–162.

Wirola, L. 2008. *High-Accuracy Positioning Or The Mass Market.* FIG Working Week, Stockholm, Sweden Technical session 8F.

Yu, D. 2007. *Modeling Owner-Occupied Single-Family House Values in the City of Milwaukee: A Geographically Weighted Regression Approach*. GIScience & Remote Sensing, 2007, 44, No. 3, p. 267–282.