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 Materials from the Offsite remain unpublished and are not available for dissemination.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

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

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

Executive Summary

Discovering and listing addresses along with recording geographic locations for those addresses are the fundamental practices that allow the Geography Division of the Census Bureau to identify the living quarters of residents of the United States. Being able to locate residences and to communicate Census information to the residents of these living quarters by mail or enumerator visits is the foundation for conducting a census. A key component of the Geographic Support System (GSS) initiative is improving the assignment of addresses and map coverage by updating the address inventory in areas not covered by the USPS Delivery Sequence File (DSF). This report describes current practices and future developments in addresses and addressing that may help advance the Geography Division's strategy for the continuous updating of its geographic databases in support of the Bureau's mission.

During the course of our examination, we reviewed five major topics of interest to the Geography Division related to addresses and addressing.

- Identify types of addresses and address components of importance to the Census Bureau in urban and rural settings for the 2020 Census and support operations, including examination of valid location descriptions identifiable through field data collection or other technologies;
- Review local, state, national, and international standards, approved and in progress, including common elements of importance to the Census Bureau;
- Research the future directions of postal addressing, including Highway Contract Routes (HCR) growth;
- Research the progress of E911 addressing activities; and
- Identify sources of addresses, both local and commercial.

In order to highlight issues and opportunities, this report contains a number of illustrations, flow models, and tables that contain significant information about the topics we describe. A number of these graphics were purpose-built to summarize our findings on the topics we explored, as requested by the Geography Division.

Our research was unable to discover any address type that might be of importance to the Census Bureau for operations supporting the 2020 Census that were not already part of the Geography Division's addressing strategy. However, we found that rural delivery points increased by 10 million units since 2001 and that the combined rural and HCR delivery points account for approximately 43 million of the 150 million delivery points serviced by the USPS. Since the

majority of these delivery locations have addresses that are not associated with the location of living quarters, it is our belief that location descriptions will become an increasingly important requirement for inclusion in the Geography Division's Master Address File (MAF).

We examined various technological approaches for creating location descriptions, focusing our efforts on tax parcels, aerial imagery, road-side imagery (similar to Google's Street View), as well as simple photography augmented by location defining software tools. Our conclusions were that parcel data could be a valuable source for addresses, address locations, and location descriptions. However, since not all of the United States is covered by parcel mapping, we believe that the use of aerial imagery can provide significant value. Additionally, road-side imagery might be able to replace many of the Census field-based data gathering activities and should be considered for further investigation.

Our review of addressing standards revealed a worldwide interest in moving beyond postal addressing standards towards standards focused on the use of addresses for identifying the location of the address, characteristics of the living quarters, and exchange formats allowing the address data to be shared by interested parties. Our research focused on the FGDC draft *United States Thoroughfare, Landmark, and Postal Address Data Standard* which includes a profile for postal addressing. We found the draft standard to be one of the most well thought out and comprehensive of the addressing standards we examined. The draft standard is on schedule to be adopted as the official addressing standard for the United States early January, 2011. The maintenance authority responsible for the standard after its adoption is the Census Bureau. While we examined the complexities of the European INSPIRE Address Specification and the efforts of EURADIN, as well as the potential ISO 19160 addressing effort, we were more impressed by the practicality of the United Kingdom's National Land and Property Gazetteer, based on the British Standard for Geographical Referencing, that serves as the authoritative source for every address, as well as address location, in the United Kingdom.

We examined the potential role that crowdsourcing and Volunteered Geographic Information might play in augmenting address information in the MAF. Crowdsourcing offers many benefits and several limitations, not the least of which is the method of data licensing. However after studying the role of crowdsourced data in address compilation, we concluded that crowdsourcing, when combined with existing methods in a hybrid approach to data compilation, might be able to provide substantial benefits to the Census Bureau.

Our research into the progress of E911 revealed that the conversion process in which non-city-style addresses are converted to city-style addresses for purposes of emergency services is fairly complete. However, it appears that many residents who have E911 converted addresses are not using these addresses for mail and continue to receive postal deliveries at their non-city-style address. Further, the E911 conversions that have taken place, with notable exceptions, exhibit a variety of approaches to the conversion process. One aspect of the E911 address conversion that will benefit the field operations of the Census Bureau is that the owners of living quarters with converted city-style addresses must post these in a highly visible format on or near their house for purposes of the provision of emergency services. Collecting the information on this signage for the residences that do not use their E911 address for mail delivery would help identify the location of these rural residences in an unambiguous manner.

Our survey of sources of addresses, commercial and local, was done by categorizing sources of addresses and examining the type and characteristics of the address information that organizations in these categories could provide. We examined the following categories of address suppliers: direct marketers/mass mailers, package delivery companies, parcel mapping companies, automobile navigation companies, white page publishers, utilities and infrastructure providers, and local sources of address information that included city/county, regional, and state sources. It is without question that local sources of address data have the most comprehensive and up-to-date address databases, but collecting and harmonizing these data pose serious challenges in terms of time and effort. Parcel mapping companies are the next best source of information, although their efforts may be matched by the address databases of package delivery companies that, unfortunately, are not interested in licensing their databases.

Purpose of the Study

The census of the United States is conducted principally by mail. The addresses to which the Census questionnaires are delivered (either by mail or by hand in case of a follow up) are extracted from the MAF, a comprehensive list of addresses and related data that are compiled by the Geography Division of the Census Bureau in a process considered one of the most crucial of all Census activities. In addition to its role in connecting the census form with the respondent (both by mail and household visits in cases of non-response), a comprehensive address list is an essential component in the accurate tabulation of census results, since the location of the address is the major link between residences and census geography.

Before 2000, the address lists used by the Census Bureau to help conduct the Census were not maintained once the Census was complete. In 2000 a MAF was created from the Address Control File originally created for purposes of the 1990 Census. The MAF has been expanded, augmented, and continuously maintained since that time.

The goals set for the MAF require that it contain a complete and current set of MAF Unit records to identify all living quarters in the United States[1]. Associated with each MAF Unit is a complete record of the current address information. Further, it is desirable that the record for each living quarter include a spatially accurate set of geographic coordinates called a MAF Structure Point (MSP) to describe the position of the living quarters, or a location description when other location information is not available. In addition, there is a requirement for an address location and tabulation framework that consists of a spatially accurate and complete network of roads and boundaries.

If the address records in the MAF do not provide the required coverage, comprehensiveness, or accuracy, the results of a census based on the delivery of census forms to the addresses or locations in the MAF will be unacceptable levels of address duplication in some areas and coverage gaps in others. Developing an accurate and comprehensive address list, as well as

[1] Shawn Hanks (2010), Addresses 101: A Background Primer, PowerPoint Presentation, The GSS Initiative Offsite, January 19-21, 2010,

supporting this effort through research on the current practices used in addressing, then, are important keys to the success of conducting a census. In addition, location addresses in the MAF (GPS coordinates, location descriptions and other forms of information that can be used to identify and find living quarters while in the field) are important records for distributing the Census documents in update/leave areas, update/enumerate areas, remote update/enumerate areas, remote Alaska areas, group quarters, transitory locations, and the addresses in the mailout/mailback areas from which no form is returned and a follow-up visit is required.

In preparation for the 2020 Census, the Geography Division, through the GSS initiative, has developed a plan to curtail costs by eliminating the need for a complete pre-census address canvassing operation. The full canvas operation for the 2010 Census, for example, was a cost intensive effort requiring approximately 142,000 temporary field workers, equipped with technologically advanced handheld computers. The field workers were tasked with verifying the addresses in the MAF, correcting errors, adding new units not in the MAF or deleting units on the list that did not exist. The ability to eliminate a full canvas and conduct a partial canvassing operation, designed to field verify addresses in areas where the address list may inadequately represent living quarters will rely, in part, on improving the quality and comprehensiveness of the addresses in the MAF. We believe that enhancements to the MAF could be accomplished by exploiting tools, techniques, and standards now available, as well as new methods being developed in the United States and worldwide by communities involved in the use of addresses, address databases, address technology, and addressing standards.

The purpose of this study is to examine the nature and characteristics of addresses and addressing practices and report to the Geography Division of the Census Bureau on developments that might be adopted to improve the quality and coverage of addresses in the MAF. It is our goal that this research will lead to development of a more robust MAF and that doing so will help support the Bureau's goal for partial address canvassing in support of the 2020 Census.[2]

Scope of the Study

Five areas related to the concepts of addresses and addressing were identified for study and analysis including: address forms and components, address standards, future directions of postal addressing, E911 Addresses, and Local and Commercial sources of addresses**.** The scope of our assignment spanned both current and future directions in addressing. In some cases, our focus reflects benefits that could accrue from adopting existing practices in addressing, while in other instances, it is focused on new or emerging techniques and practices that may become standard at some time in the future. We believe that all the research topics are relevant to understanding, improving, and supporting the development of a more accurate MAF and that they could be of benefit in pursuing the goals of the GSS initiative.[3]

---

[2] The scope of our research did not include addressing practices in Puerto Rico, as was noted in the description of our research plans for this topic presented in Deliverable 1.

[3] Our initial research agenda included a focus on address sanitizing and the use of various software provided by the United States Postal Service. Early in the contracting process the Geography Division advised us that it was in sensitive negotiations with the USPS and that several research items related to addresses, addressing and the USPS

The interrelationships among the topics involved in these five research areas are many and there is an inevitable degree of content overlap between this report on Addresses and Addressing and our reports for Task 3 (New and Developing Technologies), Task 4 (Address and Data Interchange) and Task 5 (Measuring Data Quality).

Research Topics

1. Sub-Task 1: Identification of type of address and their components of importance to the Census Bureau in urban and rural settings for the 2020 Census and support operations.

Addresses in the MAF connect the Census Bureau with the locations of living quarters in the United States for purposes the enumeration and tabulation. As such, these records can provide a conduit for the delivery of a census form or a follow-up visit to a non-response address, as well as serving as a location identifier used to "find" a living quarter in the field or for assigning census counts to blocks.

The Geography Division collects addresses that relate to individual or family living quarters (known as housing unit (HU) records), Group Quarter records (institutional living quarters called GQ records), transitory location records (homeless camps, etc., identified as TU records), and non-residential records that mainly reflect businesses or other commercial activities. Address types are recorded as City-Style, PO Box, Rural Route (including HCR addresses), or as a Location Description.

City-style addresses in the MAF must contain a house number, street name and a ZIP Code. Non-city-Style addresses, which include rural route addresses, post office box addresses or location descriptions, have various format requirements. Post office boxes require the box number and a ZIP Code. Rural routes require a rural route descriptor, rural route ID, rural route box id and a ZIP Code, such as contained in the example - RR 8, BOX 98 92751. Rural routes may also include HCR addresses, which require the contract route descriptor, highway contract route number, and box id as shown in this example - HC 68 BOX 23A 12345.

In addition to the two forms of mailing addresses described above, the MAF may include location descriptions in the form of additional information, such as description of the housing unit (e.g. white double-wide trailer with car park on the left) or an E911 address. The location description may describe a location in respect to adjacent streets, a reference to the building materials used in its construction (clapboard, masonry, aluminum siding), color painted, or other salient field-observable characteristics. Finally the location description may include incomplete city-style addresses, incomplete PO Box descriptions, or incomplete rural route descriptions. Location descriptions cannot be used for mailing, but can enable the address to be located in the

---

(including sanitizing addresses) would managed by the Geography Division, and that we should not expend our limited resources pursuing these topics. Further, the potential use of administrative records to identify living quarters and their associated addresses and location is another fertile area of research that is being focused on by the Census Bureau and will not be covered in this report.

field. Some MAF records contain mailing addresses or location descriptions, but these are treated as separate addresses on a MAF record.

Addresses are further categorized as a location address, mailing address, preferred location address, or preferred mailing address. Preferred mailing addresses are used to indicate that while a specific delivery address may matchable to an address list, it known by another preferred description (e.g. in New York, NY, AVENUE OF THE AMERICAS is also known as 6 TH AVE. An inquiry using a delivery address of 55 AVE OF THE AMERICAS is matchable but not preferred). In the same manner, a location address may not be the preferred location address. These differences are important for a variety of reasons. For example, non-city-style addresses that have been converted to city-style addresses for purposes of E911 emergency services location may not always be used as an official mailing address. These non-mailing addresses are considered a form of location address and marked as non-mailable in the MAF. Information on preferred mailing addresses is provided by the USPS through the DSF, which the Geography Division uses to update the MAF twice a year. [4]

Our research did not uncover any classification of address forms not currently covered by the Geography Division that we believe will be of importance to the Census Bureau in urban and rural settings for the 2020 Census and support operations. However, it is our opinion that there are a number of factors that will complicate the notion of addresses and addressing through the next decade and that these issues bear further examination. We suspect that "location descriptions" in various forms will become increasingly important to the Geography Division over the next decade.

1.1 The Location of Addresses

In terms of locating addresses, the use of the DSF to update the MAF is believed beneficial in areas of city-style addresses, since the address represents the location of the living quarters and can be used to find that location.). However, in rural areas with non-city-style mailing addresses, updating using the DSF does not provide a link to location, since the delivery form is to a box address that does not reveal any information about the spatial location of the residence where the recipients of the mail live. It was estimated by the Census Bureau in 2003 that non-city-style addresses included approximately 15 percent of the population and that these addresses are non-uniformly scattered across the majority of the land area of the United States. [5]

Determining the location of all possible housing units is a necessary part of fulfilling the constitutional mandate that the Census count each and every resident of the United States. For example, the specific location or "geocode" of a housing unit must be known in order for that residence to receive a census form, and for that household to be tabulated to the correct jurisdiction. While field activities encourage the discovery of missed housing units, residents at specific housing units may be missed during the field-based activities of the Census Bureau.

[4] See Appendix 1 for a description of the field layout of the DSF delivered to the Geography Division by the USPS.
[5] U.S. Census Bureau, 2003, Census Bureau to test changes in questionnaire, new response technology, Suitland Maryland: Public Information Office.

Although the MTAIP (MAF/TIGER Accuracy Improvement Program), the national Address Canvassing operation in 2009, and the LUCA program resulted in improvement to the MAF inventory of addresses and locations in city-style and non-city-style address areas, it is clear that improving the quality of addresses in the MAF remains one of the top goals of the Geography Division.

In fact, the primary focus of the GSS initiative is "… address coverage with an emphasis on addresses that (1) are not regularly updated for GEO by the U.S. Postal Service's Delivery Sequence File (2) in non-mail out/mail back areas, principally rural and (3) as they affect the address frame for sample surveys."[6]

Between 2001 and 2009 the number of rural delivery points in the United States increased by 10 million delivery points (see Table 1). During the same period, HCR delivery points increased by 600,000 and Post Office Box Service by 300,000. In essence, the addresses with no obvious location link may have increased by approximately 11 million delivery points over the last decade. Since 2005 the increase in Rural and HCR delivery points has been double of that of the City delivery points, although the rate of change was slightly less between 2008 and 2009. We note that converted E911 addresses (city-style addresses) that are used as mailing addresses (and not all are used as mailing addresses by residents) are included under the rural delivery point category, so the exact amount of growth in non-geographic addresses remains imprecise. We believe that greater than 80% of the rural delivery points are represented by non-city style addresses.[7]

Table 1. Number of Delivery Points by Type of Delivery

Delivery Point Type	**2009** Totals	Change from **2008** (000)	Change from **2005** (000)	Change from **2001** (000)
City	87,670,966	+386	+1,900	+4,000
Rural	39,704,212	+612	+3,500	+10,000
Post Office Box Service	20,091,571	-135,	+90	+300
Highway Contract Route	2,649,132	+60	+300	+600
Total	150,115,881	+923	+5,790	14,900

Data based on 2001, 2005, 2008 and 2009 Comprehensive Statement on Postal Operations, published by the United States Postal Service.[8]

However, it is clear that most rural address formats are not tied to geography and that Location Description is becoming an increasingly important method that allows the Geography Division to

[6] U.S. Census Bureau. 2010. "planning for the Geographic Support System Initiative", – Report from the January 19-21, 2010 offsite meeting, Geography Division, Washington, D.C., p2.
[7] See section 4.2 of this report. Our best estimate is that slightly over four million E-911 addresses are represented in the Rural delivery point category.
[8] The change figures are approximate since the USPS rounded the totals in two of the reports.

enter information in the MAF that would allow the targeting of rural residences for the purposes of Census field operations.

1.2 What is a valid location description identifiable through field data collection or other technologies?

In some cases, the address information available in the MAF has an incomplete city-style address that may prevent the mailing of census forms or may limit the ability to locate living quarters for purposes of field operations, such as Nonresponse Follow-up (NRFU). In these cases, the MAF may include a short description about the location or the physical characteristics of a living quarter. Examples of descriptions vary in degree of specificity and can include house color, style, intersecting streets, buildings materials, or any identifying characteristic that the field personnel believes distinguishes this location from other living quarters.

Unfortunately, the utility of the description varies as some function of the observational skills of the field worker and the uniqueness of the location. As a result, the location description may only have relevancy for the person who described the location. In addition, building characteristics can change over time due to modifications, alterations, or new construction.

In the past decade, the Geography Division has shown significant interest in "location description" as a method of providing a surrogate "address" for a location that could be used to identify a residence for purposes of field operations. The data representing the "location description" description in the MAF include the following form of representations: incomplete city-style address, incomplete PO Box, incomplete rural route, or, in some cases, one of the fields above and a physical description of the location.

We consider the MAF Structure Points (MSPs) collected by the Census to be a form of "Location Description". These geographic coordinates, if properly collected (either through GPS coordinates or digitizing the location of residences on aerial imagery) help describe the location in an unequivocal manner, regardless of whether or not the residence was represented by a city-style address or non-city style address.

For the 2009 Canvassing operation, it had been hoped that the field capture of GPS coordinates to identify the location of living quarters would fill the need for producing a unique metric for location description and by all accounts the collection program was a great success. However, various procedural and technical issues (e.g. lack of access to property, user error, multipath signal scattering, etc.,) that can and often do impede the accuracy of GPS signals may have limited the overall utility of using the collected GPS coordinates for identifying living quarters. We recommend that the Geography Davison examine their MPSs to develop an accuracy profile that would allow them to analyze the methods and procedures that were most effective in gathering MSPs. In the future the Geography Division should continue to evaluate accuracies of the MSPs and develop strategies and procedures leading to more accurate collection of these coordinates.

We researched a range of methods the Census Bureau could use to provide location descriptions that would satisfy the "validity requirement of a location description," which we define as the task of identifying the location of a living quarter in an accurate and unambiguous manner.

The alternative techniques we examined for "describing" locations included recording coordinates directly from a referential quality aerial image and associating them with living quarters either before or during field operations. In addition, we examined the value that information such as the use of parcel and building footprint mapping could provide in creating unambiguous location descriptions. Rooftop digitizing on high-resolution imagery is becoming a preferred method to determine residential location coordinates for purposes of geocoding. It is our belief that parcel mapping systems, which combine accurate location geocoding and properly formatted addresses could be used to provide the Census Bureau with more accurate sources of addressing and location descriptions than those currently used by the Bureau. From a procedural standpoint, we will discuss the use of parcels and then the use of imagery in situations where parcels and address points are not available.

It is our belief that redundant data systems may be required to solve the types of spatial search and discovery required to support census activities and we also examined methods for "describing" locations including road-side imagery comprised of photography, scanning and other recorded signatures (e.g. Wi-Fi) that could be used to provide alternate, but unambiguous forms of "location descriptions". Finally, we will examine commercially available software that can record the focal point, azimuth, bearing, elevation angle and horizon angle during the imaging process. Combining these variables with GPS allows reconstruction of the location where the camera was pointed (its pose) which should provide improved accuracy while the determining location of photographic images of living quarters. We describe these topics in more detail in the following sections of this report.

1.2.1 Tax Parcels as a source of information on the location of living quarters

Based on the requirements for the GSS initiative it will be important to determine the best sources of information about the location of dwelling units and quality of the associated address data. In this section of our report we discuss both of these issues. The first task will focus on the representation of dwelling units.

Within the GIS environment it is possible to represent the location of property in different ways such as a tax parcel, a building foot print, or a single address point. Land parcels are the fundamental operational unit of local government activities.[9] Most local governments are required to maintain a set of tax maps with parcel boundaries that depict the relative shape and location of legally defined property boundaries. These data provide the basis for a host of applications concerning the use, value, and ownership of land. The assemblage of the comprehensive set of parcels creates a mutually exclusive and exhaustive representation of the private and publically owned land.[10]

It is a common practice to mathematically generate a parcel centroid that serves as a point level representation of the containing polygon. The parcel ID associated with parcel can be used as a unique key to link to other information about the parcel including the site address. Several

[9] The National Academies, National Land Parcel Data: A Vision for the Future, National Academies Press, 2007.
[10] It should be noted that the parcel depicted on a tax map is not the legal definition of the property. Even though many communities now generate parcel boundaries from legal descriptions (metes and bounds) any legal transactions concerning the property must be certified by a licensed surveyor.

companies are actively assembling and standardizing parcel data. A list of parcel providers includes:

CoreLogic (http://www.corelogic.com/)
GIS Data Resources - GDR (http://www.gdr.com)
Digital Map Products - (http://www.digmap.com)
Boundary Solutions (http://www.boundarysolutions.com)
Data Quick - (http://www.dataquick.com)
Google - (http://maps.google.com/help/maps/basemap)[11]
Pictometry (http://www.pictometry.com/home/home.shtml)[12]

CoreLogic® (formerly First American CoreLogic®), is the leading supplier of parcel data in the United States and licenses their data to several federal agencies, including the FBI. Its database includes details on more than 124 million parcels covering over 90% of the U.S. population (see Figure 1)[13].

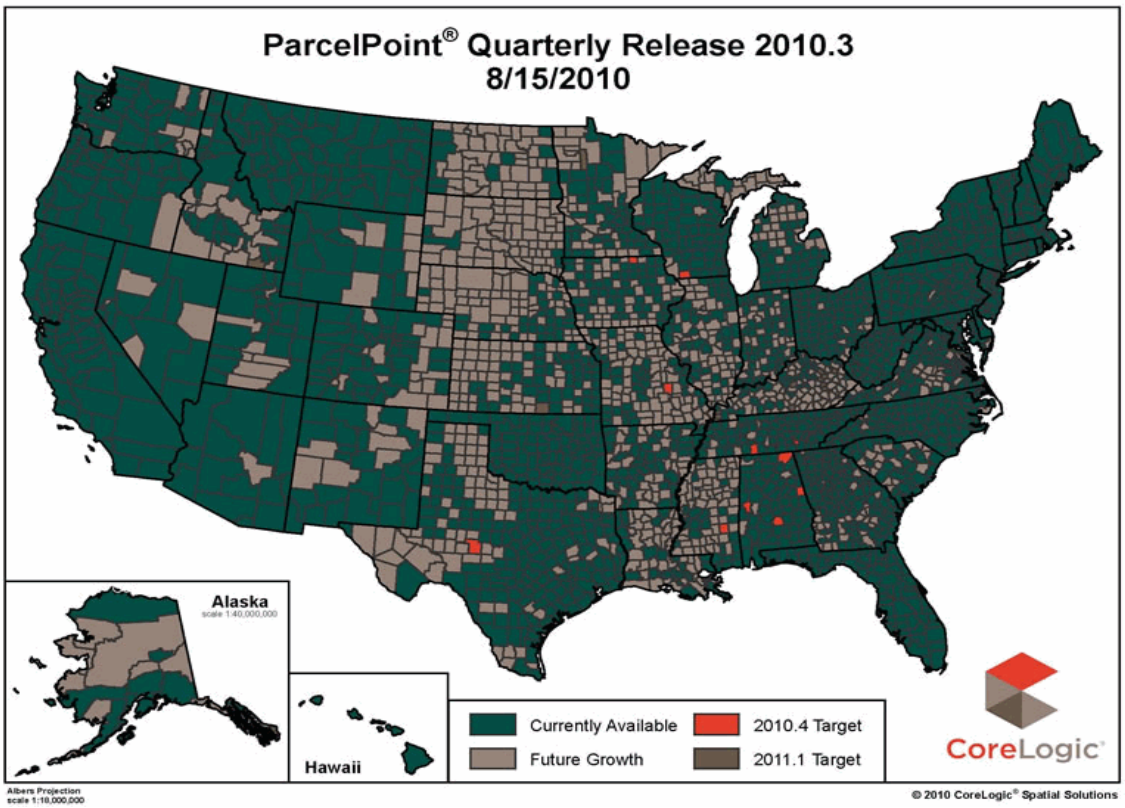


Figure 1. CoreLogic's Parcel Point coverage as of August, 2010

Within a GIS environment parcel polygons and centroids can serve as a valuable source of property reference and related location information. However, as illustrated in Figure 1 the utility of the parcel centroid as a MSP varies with the size of the parcel and the number of

[11] Google collects parcel data from others. It displays, but does not resell, the parcel data that it has collected.
[12] Pictometry resells parcel data they relicense from CoreLogic.
[13] For more details on ParcelPoint covergage, see http://www.corelogic.com/Products/ParcelPoint-Technology.aspx

structures. For example, the parcel centroid would serve as an excellent MSP for any of the small parcels with a single dwelling unit (e.g. 1306 Witter St.). On the other hand, it would be problematic for a large parcel with multiple structures (e.g. 1543 Fort Johnson Rd) or an apartment complex (e.g. 1530 Fort Johnson Rd). From an operational perspective, it would be possible to utilize parcel centroids for only those residential parcels that are less than a certain area (e.g. ½ acre) and contain only a single address. An advantage of parcel centroids is that the coordinates are mathematically generated rather than field collected GPS points. Therefore, they can be maintained based on any update cycle implemented by the provider. They also ensure that points fall within a containing block and would maximize the spacing between dwelling units in areas with very small parcels such as row houses.

From a quality control perspective the geometrically generated parcel centroids, would ensure that that MSPs are contained by the correct block. Geometrically, contiguous parcels can be dissolved to create blocks and the remaining non parcel areas represent "right of ways" that provide a basis for alignment of street centerlines (Figure 2). While these parcel based "blocks" would not represent all census blocks, they should represent all blocks that contain residential dwellings.

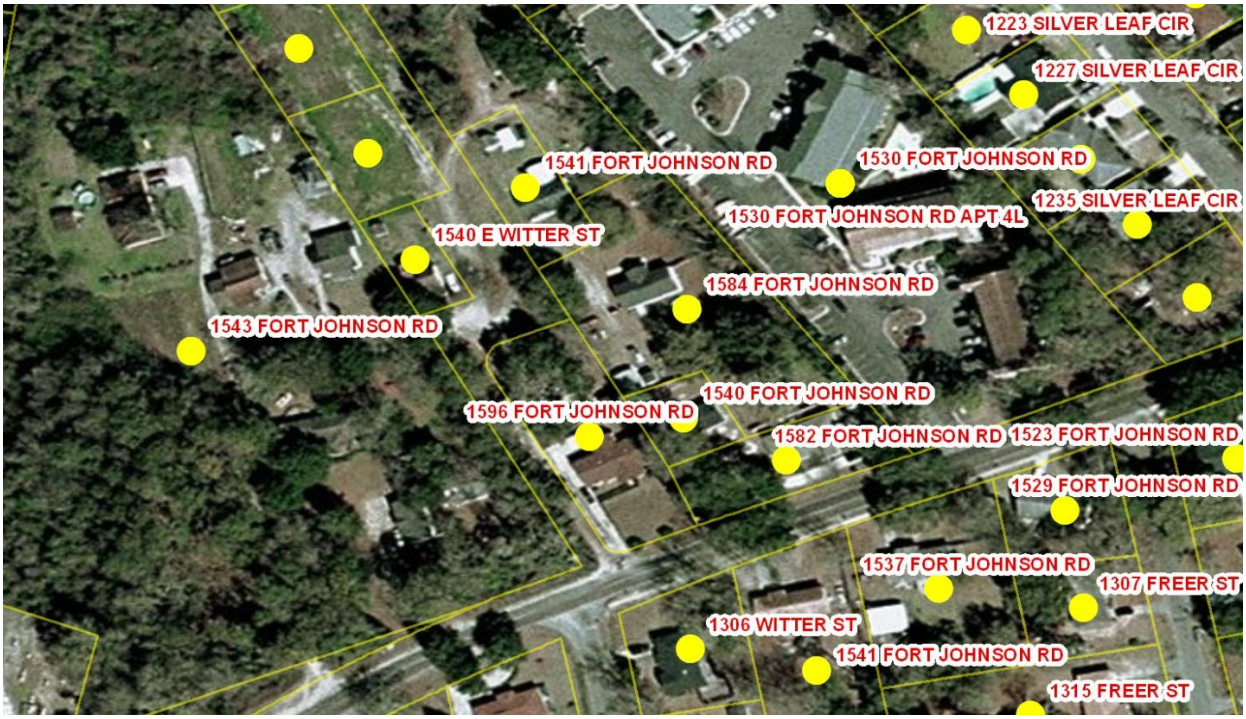


Figure 1a. Core Logic parcel boundaries and centroids for a part of Charleston County, SC

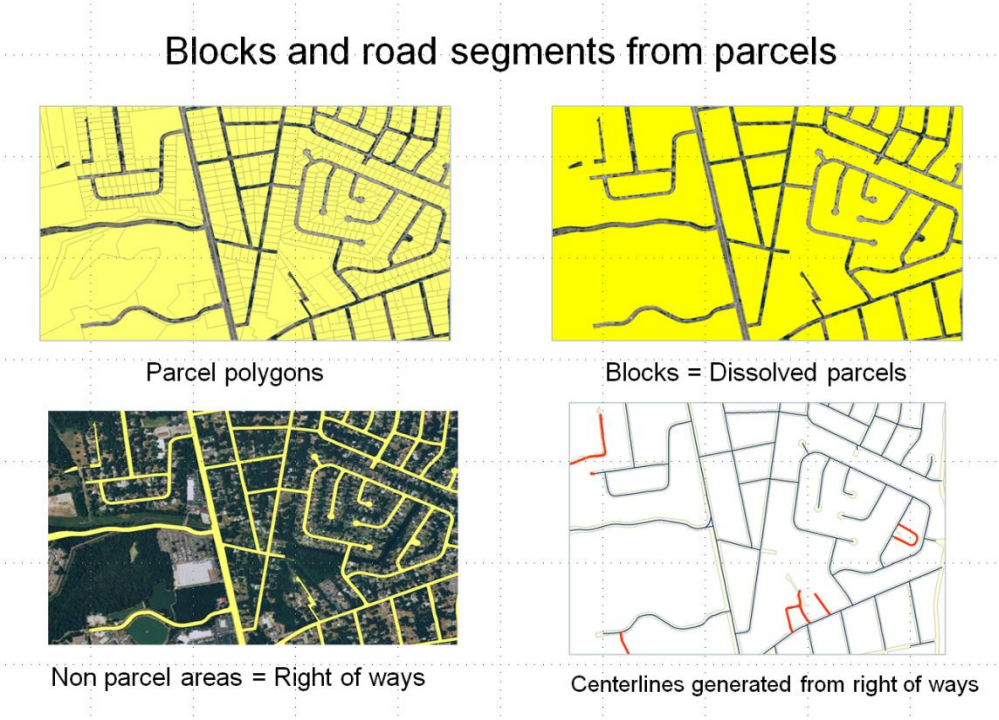


Figure 2. Extraction of blocks and centerlines from parcel data

It should also be noted that parcel data typically exists prior to the actual construction of the structures and provide records of future development. In fact, many communities such as Richland county S.C. now require include digital submission of new subdivisions (Figure 3). In a well-coordinated local government, the street segments, names and addresses are assigned as part of the approval process. These address data are shared with the E911 dispatch system so that ambulances can be sent to construction sites, as required.

Figure 3. Overlay of digital version of approved subdivision in Richland County SC. (http://www.richlandmaps.com/pdf/smac06_dds.pdf)

1.2.2 Limitations of parcel data for collecting MAF Structure Points (MSP)

We believe that it is feasible to implement a multiple tiered system for recording and maintaining MSPs in which parcel centroids are only used for typical, single family residential areas in urban or suburban settings. The following section describes some of the limitations of parcel centroids in other settings. One potential disadvantage is that the centroid is a virtual location that is not *required* to coexist with a dwelling unit under the current data collection standards used in the parcel mapping industry. In fact, neither the parcel boundaries nor centroids represent visible features. Therefore, the use of parcel centroids to identify living quarters could cause confusion during field verification, especially if the field representative refers to commercial online maps. For example, NAVTEQ's address points are geometrically moved from the centroid to the appropriate position on their street centerline file and may be ambiguously located in reference to the location of a specific residence (See Figure 4).

However, since the focus of GSS initiative is unambiguous representation of dwelling units, the fact that the centroid is not associated with a visible feature might not be a problem except for rural areas with large parcels. In these cases, using an MSP that is associated with the dwelling would be a superior solution. A good example of this is the practice in San Juan County NM where laser ranger finders are being utilized to capture coordinates for front doors.[14]

A more serious problem with parcel centroids relates to the relationship between the parcel polygon and structures. Several apartment buildings, condominiums and group quarters can be located on one legal parcel. This parcel may be owned by a single entity or shared among many owners. Within a single parcel there can be multiple dwellings with separate addresses contained in one or more different structures. In these cases there will be a "many to one" relationship between addresses and the parcel. An illustration of this is provided by the Core Logic representation of 49 apartments that are represented at a single point in Charleston County SC (Figure 5). While an apartment complex will have a single owner who is responsible for property taxes, a condominium consists of multiple dwellings with separate ownership and tax bills. Therefore, the local assessor may represent the condominium complex as a single parcel polygon that will contain many units with different tax IDs. Furthermore, there is little standardization in terms of the location of the point or polygon representation for the separate condominiums. If the complex is a single story it is possible to assign a tax ID to each unit and represent it as either a point or polygon feature. However, for multi storied units the floors of the complex may be cartographically distributed for representation without any correspondence to exact geographic location.

Additional concerns relate to the address information. It is important that the address correspond to the site address and not the billing address for the unit. Assessors tend to be more concerned with the correct format for the mailing address of the owner than the site address. In many cases the site address is missing or may be included in a non-structured manner.

[14] Cowen, D. et. al, Deliverable 4, Task 005. *Reporting on the Use of Handheld Computers and the Display/Capture of Geospatial Data). 2010.*

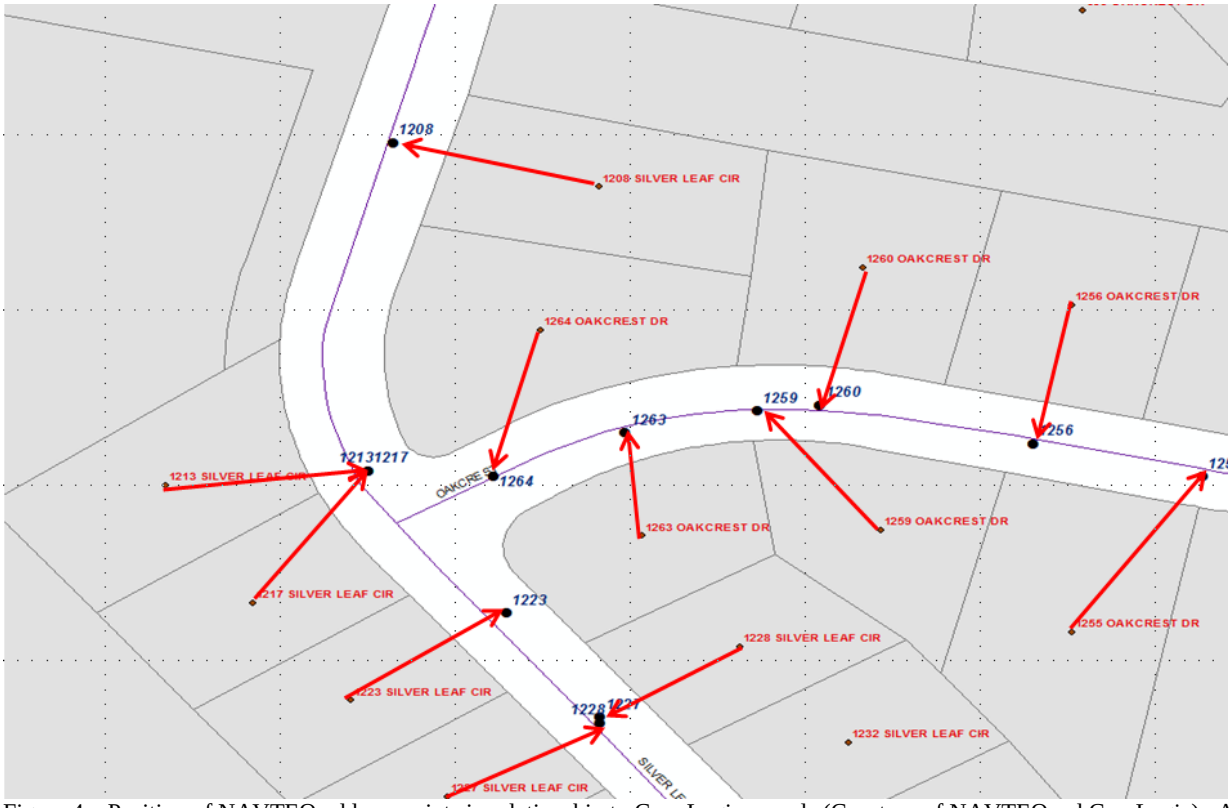


Figure 4. Position of NAVTEQ address points in relationship to Core Logic parcels (Courtesy of NAVTEQand CoreLogic). As noted in the text, some the the centroid locations misrepresent the property location.

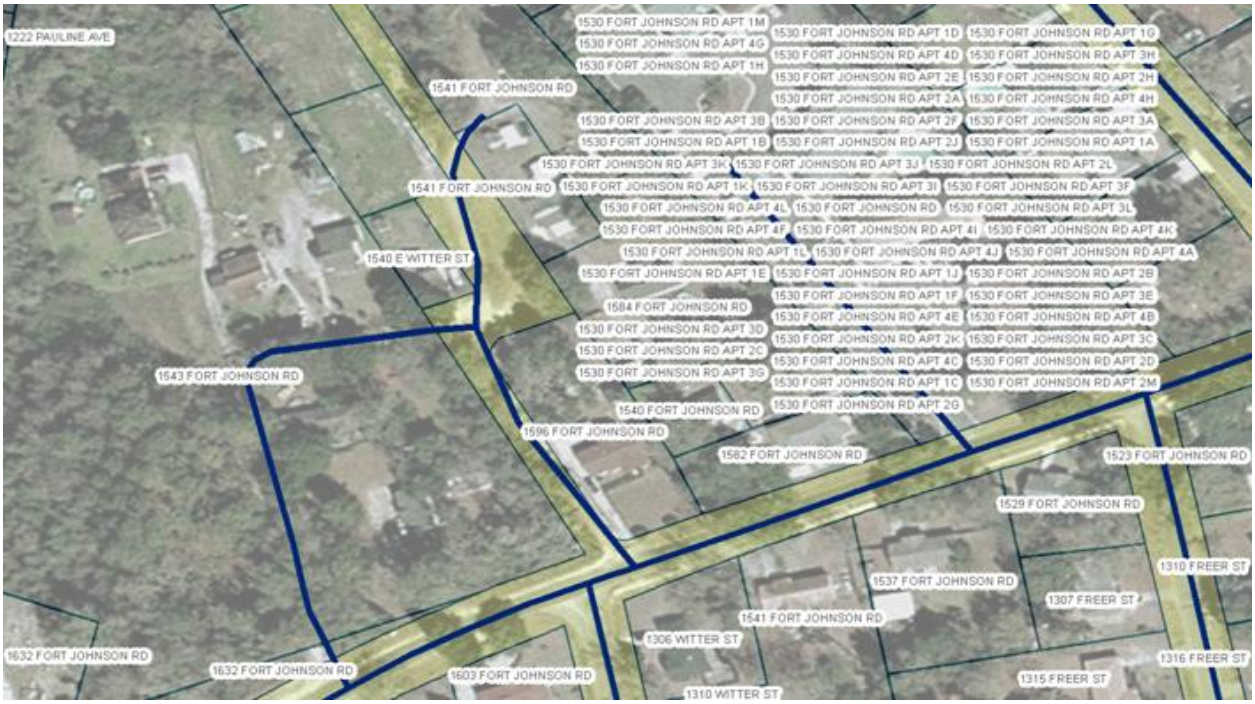


Figure 5. List of 49 apartment numbers associated with one parcel (Courtesy Core Logic)

Since parcel centroids represent a geometrically determined location they are not a good solution for many forms of geocoding or navigation. As noted in the NRC report *National Land Parcel Data: A Vision for the Future*,[15] William Gail, the director of Microsoft's Virtual Earth program indicated:

> While geocoding based on interpolation of address ranges has been adequate for initial needs, it is far from satisfactory in the long run. In fact, even parcel identification is insufficient for many of the emerging internet needs, as a parcel may contain several buildings for which separate identification is desired. Microsoft Virtual Earth recently introduced rooftop geocoding to relate addresses to building locations, ensuring that addresses correspond properly to the relevant buildings. This is the most accurate geocoding currently available for applications such as route-finding, and we continue to improve our capability in this area. . . . Microsoft depends primarily on vendors to supply the most accurate possible information. At the same time we recognize that local governments often have the best first-hand knowledge of the current situation on the ground. Geocoding is likely to represent an ongoing challenge, both in the collection of raw data and in the algorithms that extract information from the data in response to user queries.

The importance of accurate geocoding was illustrated by a comparison of various geocoding services for the residential address and is a prime example of why parcel centroids or MSPs are required for the identification of the location of living quarters (See Figure 6).

[15] The National Academies, National Land Parcel Data: A Vision for the Future, National Academies Press, 2007.

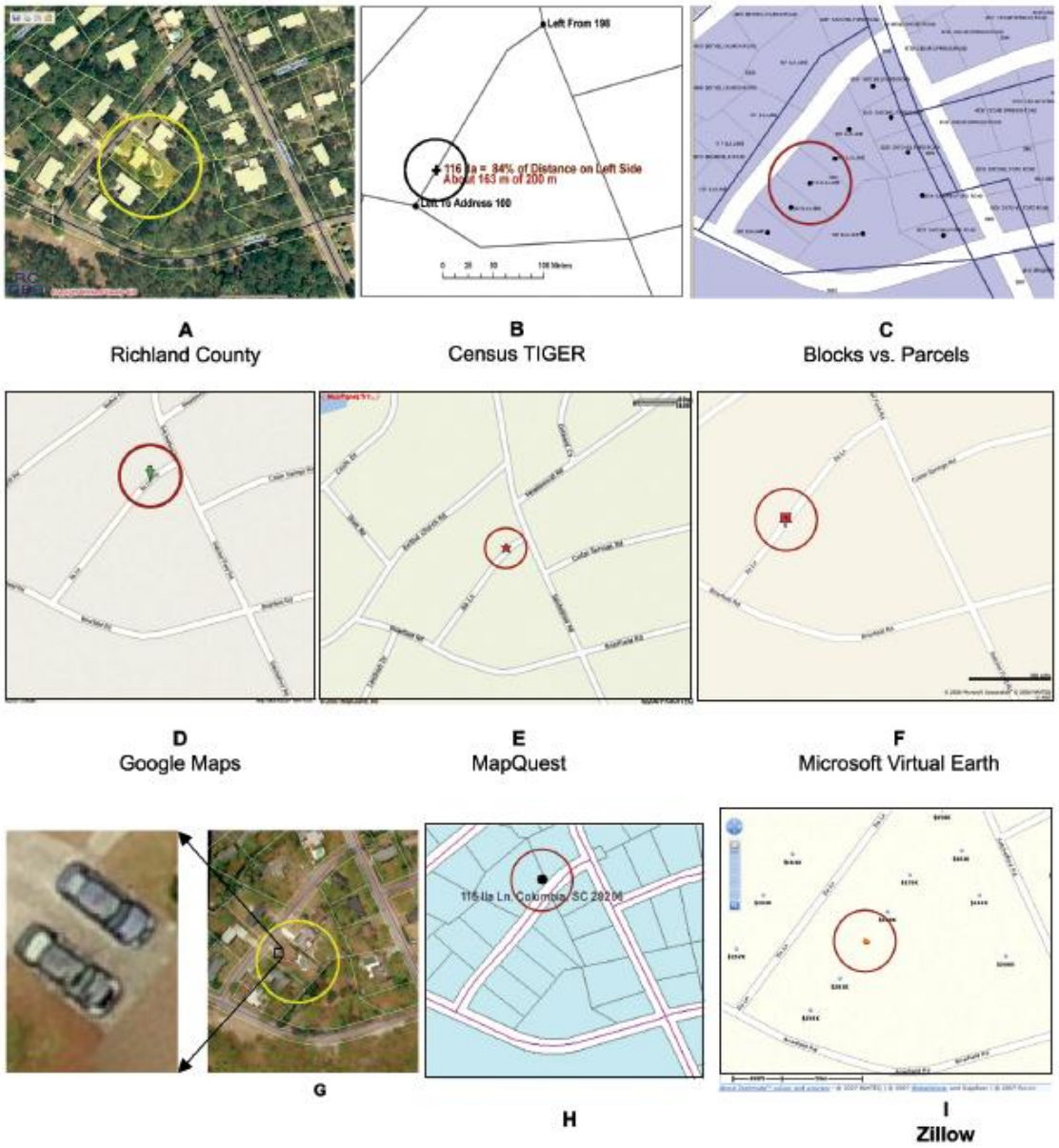


FIGURE 2.10 Various locations of a sample address through different automated address matching applications. SOURCES: (A) Image courtesy of Richland County, SC, www.richlandmaps.com; (B) data courtesy Census Bureau; (C) data courtesy of City of Columbia, SC, Census Bureau; (D) image courtesy Google Maps, copyright NAVTEQ; (E) Map content (c)2007 by MapQuest, Inc., MapQuest and the MapQuest logo are registered trademarks of MapQuest, Inc., used with permission; (F) image courtesy Microsoft Virtual Earth; (G) images courtesy of Pictometry, Inc.; (H) parcel data courtesy of City of Columbia, SC; (I) courtesy Zillow, map data copyright 2007 NAVTEQ, GlobeXplorer and Suppliers, and Proxix. Parts (D), (E), (F), and (I) (c)2007 NAVTEQ. NAVTEQ is a registered trademark of NAVTEQ Corporation. NAVTEQ map content is used with permission.

Figure 6. Alternative Geocoding locations of the same address. (From the National Academies, *National Land Parcel Data: A Vision for the Future*, National Academies Press, 2007**)**

Given the GSS initiative requirement to unambiguously associate the correct response to a specific dwelling unit, the adoption of rooftop geocoding rather than front door should be considered. Excellent examples of rooftop address points are provided by the city of Charlotte NC (See Figure 7) and the state of Indiana (See Figure 8). In these examples, address points have been determined by placement with respect to image or a building footprint – clearly not by direct capture of a GPS point in the field.

The Charlotte system provides a particularly rich environment that includes a current high resolution image, the parcel boundary with tax ID, and a rooftop address point. The Indiana map example illustrates how multiple dwelling units can be positioned on the same structure (See Figure 8). It should also be noted that both of these examples could be linked to a normal GIS desktop (or mobile) operation as a web service provided at low cost to the Bureau. Another interesting example of multiple representation of property is provided by Charleston County SC GIS office. In addition to the image, parcels, and buildings they also distinguish between buildings, garages/sheds, and manufactured houses (See Figure 9).

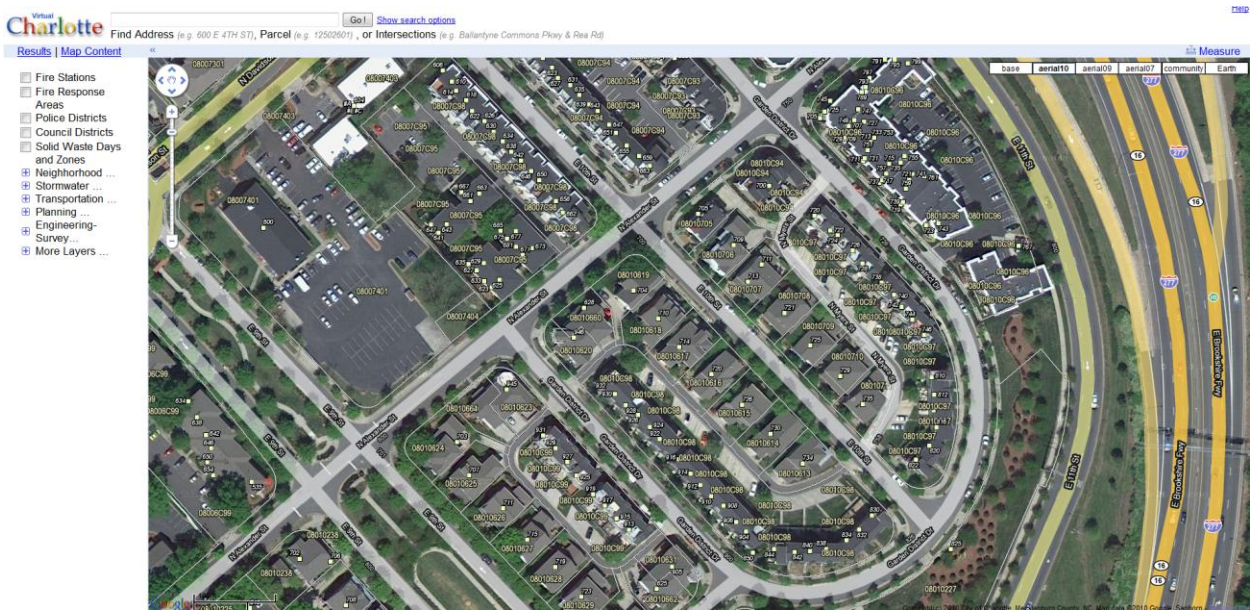


Figure 7. Imagery, parcels and rooftop address points. This example illustrates that multiple address points can exist within the same parcel and building. (Virtual Charlotte http://vc.charmeck.org/)

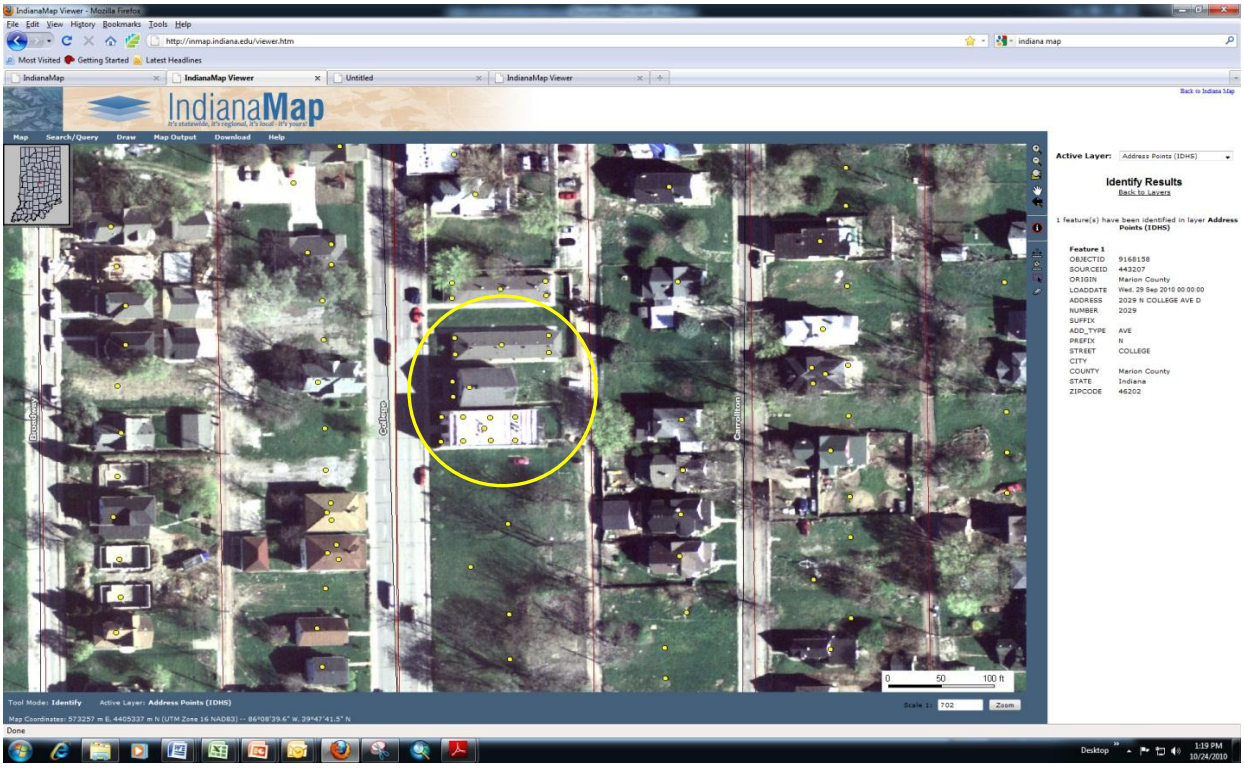


Figure 8. Rooftop address points. (http://www.indianamap.org/)

Figure 9. Alternative representations of parcels, buildings and boundaries in Charleston County SC (http://ccgisweb.charlestoncounty.org/website/Charleston/viewer.htm)

Given the ability to record multiple address points within a parcel, the "rooftop" address point is generally superior to a single parcel centroid for determining the location of a living quarter

within a parcel. However, the placement of the points in rooftop geocoding is usually a manual process rather than the automated procedure typically used to generate parcel centroids.

1.2.3 Imagery as a source of the location of living quarters

In the absence of parcel or purpose-recorded address points, high resolution current imagery can serve as valuable source information for discovering missing street segments and structures. Although the manual image interpretation used to determine the location of living quarters can be time consuming, there are a number of interesting developments in the area of automated image processing that might be of assistance to the Geography Division. A detailed analysis of automated feature extraction and change detection will be covered in Deliverable 6 of this series titled *Identifying the Current State and Anticipated Future Direction of Potentially Useful of Developing Technologies*. However, it should be noted here that current image processing technology exists to automatically extract new roads and buildings from either color imagery or LiDAR and to compare two images to find houses that were not built when the older imagery was captured. (See Figures 10, 11 and 12)

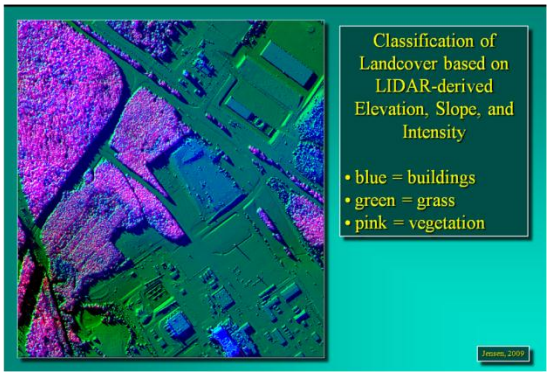
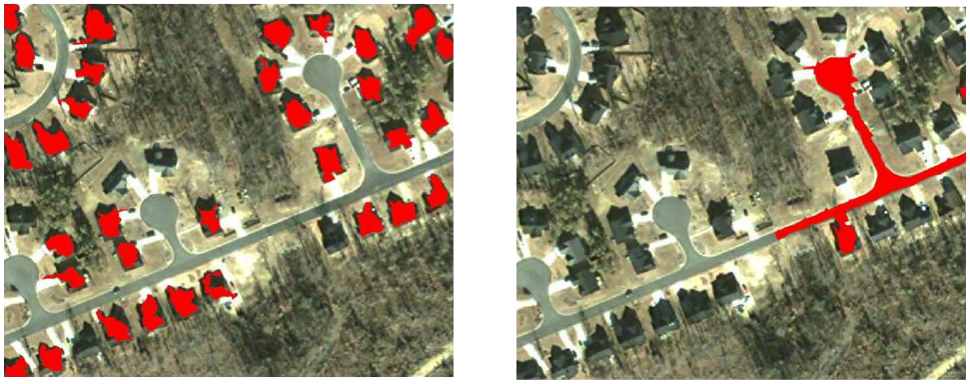


Figure 10. LiDAR derived features



New Buildings – $P^2/A < 70$ New Roads – $P^2/A > 70$

Figure 11. Rule based extraction of new roads and buildings

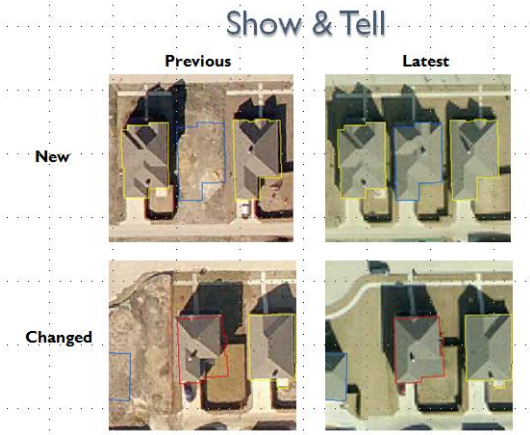


Figure 12. House-finder comparison between two images (Courtesy Pictometry)

Many organizations are actively acquiring high resolution imagery either as conventional orthophotos or oblique photos (See Figure 13). It is common that these images are collected on an annual or biannual basis simply to support internal applications related to governmental issues, particularly property taxation. Over the last decade, the demand from location based services, including the need to monitor the characteristics of real estate, has resulted in a significant increase in the number of companies and agencies willing to invest in high resolution imagery. [16]

Figure 13. Illustration of value of imagery to validate the position of a TIGER road segment (yellow)

[16] The spatial, temporal and spectral requirements of imagery are discussed in detail in Deliverable 6, Task 3 titled "Identifying the Current State and Anticipated Future Direction of Potentially Useful of Developing Technologies."

As a result of this competition between major corporations such as Microsoft, Google and Esri for the best content, the sources of imagery are multiplying. This includes partnerships between Google and GeoEye for commercial satellite data with 50 cm resolution.[17] Of greater interest is a new public-private partnership program that Microsoft has initiated with the goal of providing statewide imagery with a one-foot resolution to be updated on a three year cycle.[18] This initiative is evolving as the Digital Globe/Microsoft Clear 30 program. There is a possibility that this program may become a viable alternative to the Federal Imagery for the Nation program.[19] At the same time the National Agriculture Imagery Program (NAIP) will continue to acquire its one meter data. As a result of commercial interests and NAIP it is likely that there will be detailed imagery available for most areas of the country, but specifically, we believe that there will be high quality imagery available for the large sections of rural America that are not covered by postal addresses or choose not to use E911 addresses as mailing addresses.

It is also significant that many of these image resources are available as web services provided either by the local government or through a commercial mapping application (e.g. Virtual Charlotte).[20] From the viewpoint of the GSS initiative, access to oblique photography from sources such as Pictometry may prove to be extremely valuable resource.[21] An example of this is provided in figure 14 for a part of Charleston County SC where there may be some confusion regarding the association of streets and buildings. In this example, four oblique views of the same area are accessed through the "Birdseye View" option on Bing Maps. Access to the same imagery can also be obtained through Pictometry's on line service (Figure 15), which can also be incorporated into a desktop GIS environment.

Figure 14. Four oblique views of neighborhood in Charleston County SC (Bing Map)

[17] See http://www.geoeye.com/CorpSite/assets/docs/investor-relations/Key_Financial_Stories_Cnet_09_08_08.pdf
[18] See http://www.michigan.gov/documents/cgi/CSSTP_pleased_to_announce_acceptance_of_BME_298755_7.pdf
[19] See http://www.digitalglobe.com/downloads/products/AerialProgram-DS-AP.pdf
[20] See Virtual Charlotte at http://vc.charmeck.org/
[21] See the Pictometry website at http://www.pictometry.com/home/home.shtml

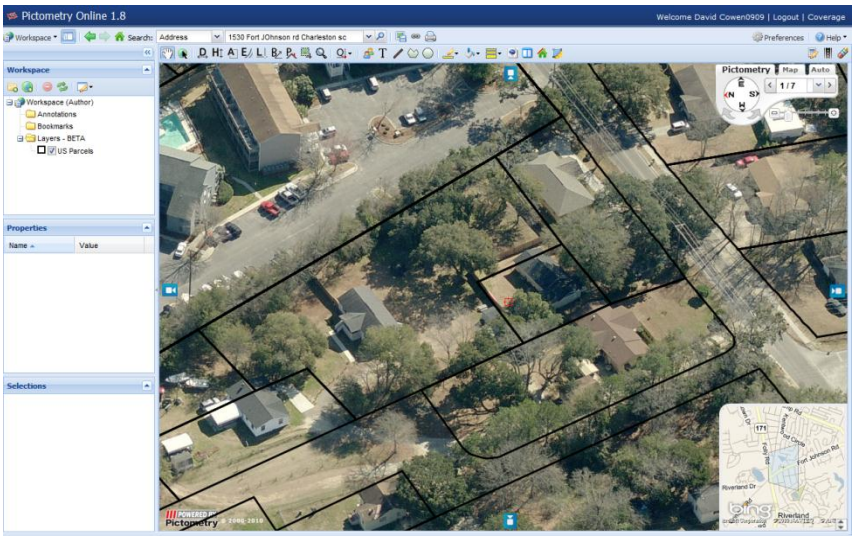


Figure 15. Pictometry on line service with parcel boundaries for a neighborhood in Charleston County SC
This view from west corresponds with the image in figure 14. Parcel boundaries would align with the ortho view.
(Courtesy Pictometry)

Access to these resources is changing the way local governments conduct inspections. Rather than send building inspectors into the field many of the operations are conducted at the desktop. For areas with considerable new development and change, the discovery of new additions, swimming pools, etc., adds significantly to the tax base, which, in turn, justifies the expenditure on new imagery and saves the cost of sending workers into the field. As noted in our report on mobile computing, much of the discovery and actual capture of street and structure points could be performed in office rather than field environments.

This brief review of address-related resources suggests that a hierarchy of approaches and best practices to support the goals of the GSS initiative could evolve based on a considering a fusion of information sources. The crucial issue here is that cost savings minimizing the amount of expensive and time consuming field survey operations. When possible it is desirable to relegate field observations to validation operations. In a coordinated balance between office and mobile operations conducting a field survey should be deployed only as a last resort.

An example of a work flow is included in Figure 16 below. Actual update processes could take different paths through the decision points. The alternatives are based on the availability of authoritative source data. An example of one path is illustrated by the discovery of a new subdivision in Charleston County SC (Figure 17). In this example, the TIGER street segments are compared with the Charleston County E911 street centerline file. Since these street segments are provided by an authoritative source they can be used to update TIGER. While these streets do have address ranges, another source of MSPs was required. In this case parcel centroids provided by Core Logic (they could be from the county or another trusted provider) are used to generate the MSPs. Field operations would consist of validation that could be performed from a vehicle.

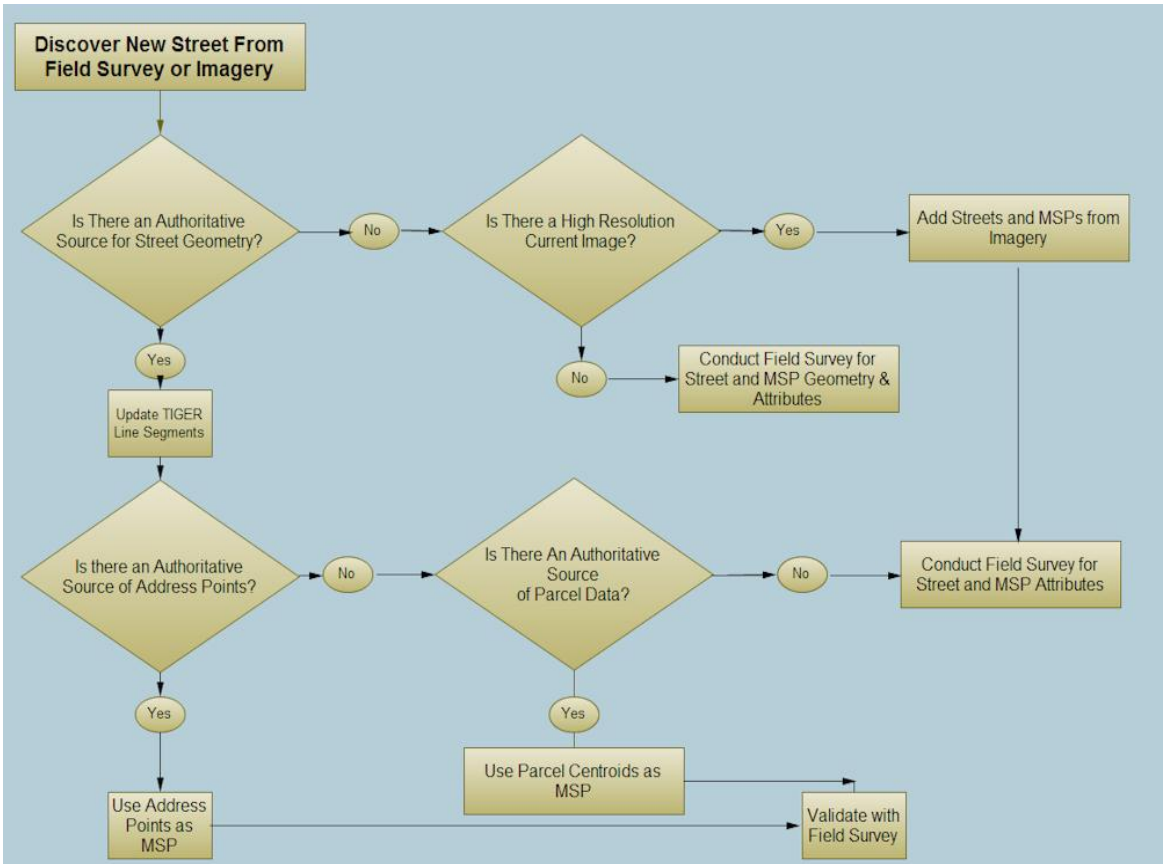


Figure 16. Sample flow chart showing the utilization of alternative data sources in a fusion process for determining MAF Structure Points.

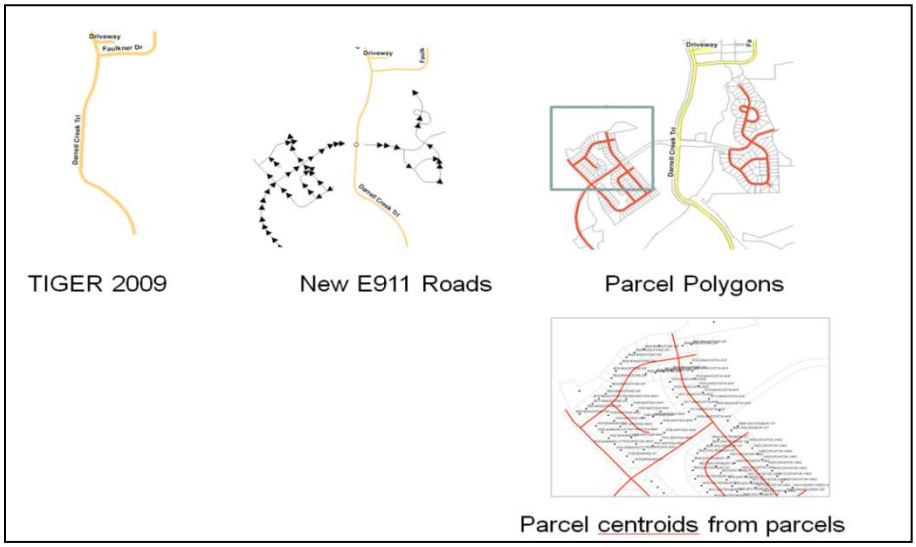


Figure 17. Example of how TIGER roads and MSPs could be updated from authoritative local E911 street files and commercially provided parcel centroids. The new roads were derived by an automated change detection algorithm. The parcel attribute data provides the mailing address. (Data courtesy of State of South Carolina GIS Council and CoreLogic)

1.2.4 Road-Side Imagery

As noted previously, discovering location addresses in non-mailout/mailback portions of the United States remain a difficulty for the Census Bureau. Many rural living quarters may not have a mailing address to help identify them. In addition, numerous rural residences that have experienced E911 conversions and have a city-style address posted on their property for emergency services, continue to receive mail at mailing addresses that do not reveal detailed information about the spatial location of the living quarter (e.g. rural route addresses, HCR addresses or Post Office boxes). Since these types of situations can only be resolved by direct observation, we recommend that the Geography Division consider an automated solution for this data collection problem. Using road-side imagery as a method of compiling and inventorying location descriptions by capturing the required signage to mark a living quarter with its E911 city-style address could be a very helpful procedure. We recommend that the Geography Division, consider road-side imagery for use in capturing addresses for locations in TEAs (Type of Enumeration Area) where address coverage and MSPs are thought to be incomplete or in areas where their data quality needs to be enhanced.

We believe that road-side imaging systems could provide an image-based environment capable of capturing information on "Location Description" of unparalleled accuracy and at a reasonable cost. In essence, these systems can be used to create a complete photographic inventory of the area they image, including detailed photography of all living quarters, business and other buildings along the path of the device. Images can be zoomed and queried for: distance from roads, geographic coordinates, and other relevant information. It is possible that addresses on building or visible mailboxes could be imaged processed to aid in assigning information to addressed locations. In fact, these images could be overlaid with details including Census geography, as well existing location description information from the MTdb database, or even from the DSF. For example, the image could include an overlay of the block and tract information for the living quarters, or the preferred mailing address of living quarters occurring along the path of the device. Since the device would provide very accurate coordinates for the location of living quarters, the locations, in theory, could be linked to the MSP, which could be used as a pointer to other data in the DSF about the location.

Modern road-side imaging systems are capable of capturing highly precise geographic coordinates in virtually any environment, which is not true of handheld GNSS recorders. The hybrid imagery database created by road-side imaging systems produces a queryable image database for all locations of interest and a 3D model of the landscape that could be used to compute locations, distance, and direction where such information was required. Further, the model could be queried at later points in time for data that was not required during the initial survey of the area.

Of course, imagery-based systems are not new. In the early 1900's Andrew McNally II, of Rand McNally & Company, created an auto-photo-guide of the road segments between Chicago, IL and Milwaukee, WI. Although he was driving to Milwaukee on his honeymoon, he stopped at every recognizable intersection where a maneuver was required along the route and snapped a photograph showing the roads and the buildings on each of the corners. Meanwhile, his wife created annotations on which way to turn and for all of the notable buildings at the intersection. Of course, roadways do not look exactly the same when traveling in the opposite direction, so the

McNally's photographed the same intersections on the return route and annotated what you would see on the journey between Milwaukee and Chicago. The auto-photo-guide was a best seller in its day (1905), because McNally realized that imagery by itself was not the key to success and that annotations were required to provide a geospatial reference of use to his intended audience. Of course, the photographic method was then burdened by cost of compilation, annotation, and manufacture and was soon replaced by line maps, a lower cost alternative that were easier to update, but less informative.

It was not until the last decade that low cost, highly efficient technology was combined in an instrument pod that could provide high resolution imagery, a LiDAR point cloud and hemispherical photography coupled with inertial navigation and GPS to create mobile, automated systems that were capable of imaging roadside buildings while traveling at a normal driving speed. The advent of these systems and their potential importance was established by the introduction of the Google Street View product in May, 2007. (See Figure 18.)

Figure 18. Example of a road-side image from Google. Resolution does not reflect true capabilities of the imaging system.

Google's Street View imaging platform was created by Sebastian Thrun of Stanford University who had developed the technology to serve as a sensor platform for his interest in robotics. In fact, Dr. Thrun had developed a miniature version of the platform that was flown on toy helicopters to serve as an image observations deck to provide an overview of the landscape that could be interpreted by machine navigators.[22] Google and Thrun modified the system to meet the requirements of Street View and others have improved on the concept since then.

Modern road-side or street-side imagery capture is the result of highly instrumented vehicles equipped with: an inertial measurement unit (IMU) to measure the vehicles velocity, orientation, and attitude, GPS and Glonass sensors to capture position, and LiDAR scanners to capture a 3D

[22] See Dr. Thrun's homepage at Stanford http://robots.stanford.edu/ , for information on this system as well as the system that he originated to win the DARPA challenge involving autonomous vehicle navigation. Both have relevance to road-side imaging.

point cloud of roadside features The combination of LiDAR and 360° spherical imagery enhance data analysis, navigation and content inventory information. Along the sides and paths of streets and roads these imaging systems can "look" up, down, sideward, forward, and backward. In some cases the sensor platform contains receivers to capture cell tower signal measurement, or signals from Wi-Fi access point locations. These signal sources can be used individually or combined into RF fingerprints for locations that can be used to identify their position when it is required to augment or substitute for weak or missing signals from Global Navigation Satellite Systems (e.g. GPS, Glonass (and eventually Galileo)).

The instrument pod usually contains three LiDAR Scanners and the ensemble of instruments are used to compute all position imposed information on a 3-D model of the environment and take precise positional measurements anywhere through the model. The hemispherical imaging system often contains eight cameras pointing horizontally around the compass, while one camera shoots up. Other instrument pods add odometry by tying into the CAN Bus (Controller Area Network) on vehicles or by adding wheel encoders.[23] (See Figure 19.)

Figure 19. Topcon IP-S2 3D Mobile Mapping System

[23] Odometry is the use of data from sensors to estimate change in position over time. It is often used as a component in map matching when GPS is not available to in-vehicle navigation systems.

The road-side imaging platforms create an analytical model that generally provides high quality resolution and imagery. Earthmine, for example, provides a queryable 3D point for every pixel and 8 million 3D points per image, which are captured at 24 million points per second in 360° (H) by 180° (V) stereo panoramic imagery.[24] Similar systems are offered by Topcon[25] and Trimble[26], while others have been developed by NAVTEQ[27] and Tele Atlas.[28] In terms of accuracy of location, most of these systems advertise better than one meter globally (within 50 meters of the vehicle) and 2 to 50 centimeters are possible depending on the range. Data (images) are usually collected every 10 meters and are described by standardized metadata. (See Figure 20.)

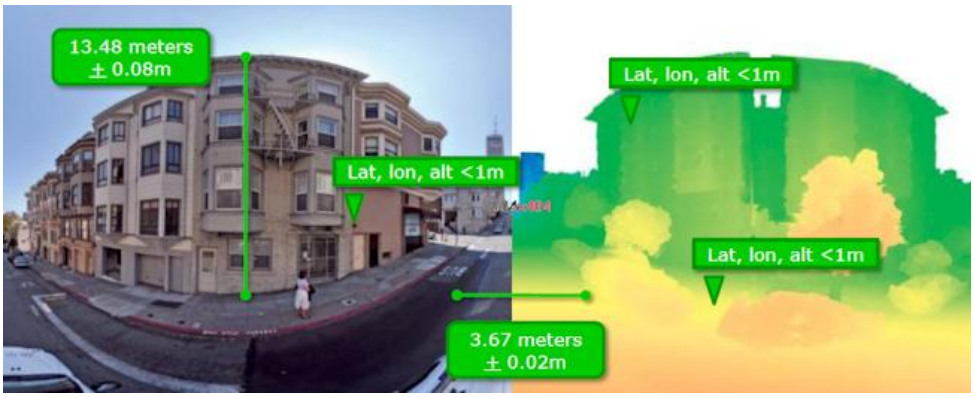


Figure 20. Example of point measurement accuracy from road-side imaging. Illustration courtesy of Earthmine.

Most of the road-side imaging solutions are highly automated, require little in the way of operator training (many companies claim that if you can drive a car, you can operate their system), and the data are inexpensive compared with other alternatives, such as field crews. Data are available through a cloud or online locally, if the client prefers. Earthmine claims to offer data at one-fifth to one-tenth the cost of other companies in the market, if the customer is willing to license, rather than own the data. In addition, the co-CEO of Earthmine claims that the cost of their systems is considerably "...cheaper than walking around with GPS receivers."[29]

Whether the road-side imaging systems could be used to extract address information from visible cues in the imagery for urban areas is unclear since there is a significant variability in address signage. A national standard for address signage has not been adopted. There are local standards for addresses on buildings (planning commission based and related to Uniform Fire Codes). E911 addressing standards vary but generally consist of reflective characters, at least 3 high, on a contrasting, usually reflective background. It is generally required that the address number to be affixed to the building or post visible and readable from the street.

While we see little problem for these systems capturing E911 addresses (both roadside and house attached) due the size requirement for posting the address, capturing smaller city-style address

[24] For more information on Earthmine, see their website at http://www.earthmine.com/index .
[25] See http://www.topconpositioning.com/products/mapping-and-gis/mobile-mapping/ip-s2.html for details
[26] See http://www.trimble.com/news/release.aspx?id=071310a for details on this system.
[27] See http://www.newscientist.com/article/dn19454-nokia-developing-3d-rival-to-google-street-view.html for more information, as well as http://navteq.com .
[28] See http://www.teleatlas.com
[29] Conversation with Anthony Fasero, founder and co-CEO Earthmine, September 15, 2010.

notations from urban mailboxes or as posted on living quarters could be more of a challenge. While street centerline providers have mapped areas and extracted street sign text using the techniques of computer vision, it is important to understand that this task is easier than extracting addresses, which tend to be posted using smaller sized text and colors with lessened contrast.

While some may note that these instrument pods appear fragile, many of them now are being mounted on pickup trucks and subjected to difficult environmental conditions. Others, like Google, have adapted the technology to pedal-powered tricycle-like conveyances that contain the gantry and are capable of imaging areas where vehicle traffic is not allowed, including parks and monuments and narrow residential ways that do not allow for vehicle traffic.

If the Census Bureau has interest in pursuing address acquisition in urban environments, we suggest a test study that utilized computer vision techniques to look for address objects in complex and challenging scenes, coupled with a self-tuning module that would allow improvements based on machine learning. This is the technique that Google has taken to improve face blurring across crowds. With training, these types of algorithms improve and may prove to be capable of address extraction in numerous demanding environments (see Figure 21). Further, an overall test of road-side imagery should be configured to compare and contrast the types of information that can be gathered by imagery and other sensors with the types of information gathered during the 2009 Address Canvassing by field representatives. Doing so would be important for determining whether road-side imagery could be a replacement for canvassing or only suitable as a supplemental technique.

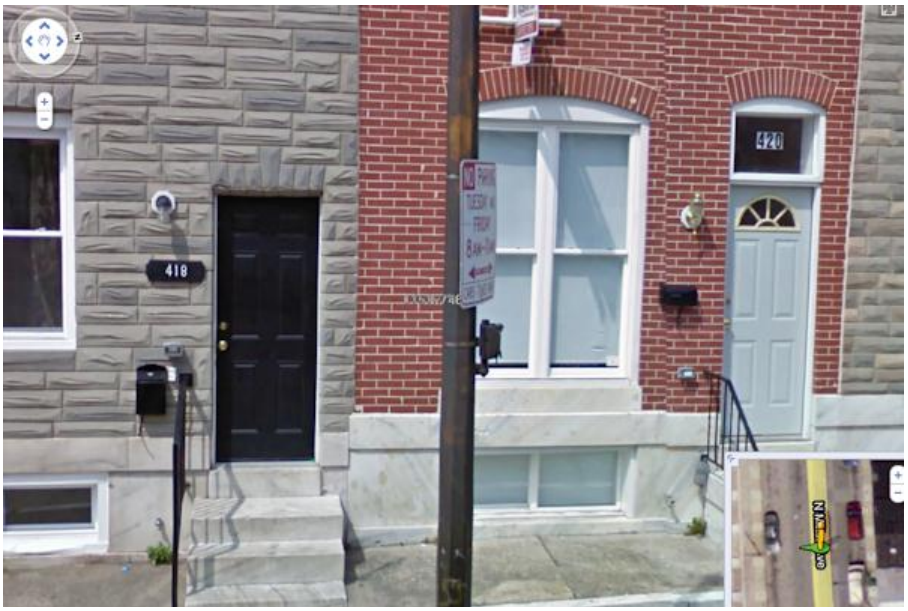


Figure 21. Addresses visible from Google Street View would not present image extraction problems when the addresses are as clearly posted as those shown in the above image from Baltimore. Current systems would exceed the clarity and utility of the image shown.

1.2.5 More Possibilities

Google's vehicles that capture road-side imagery also are equipped with sensors that measure cell tower strength and Wi-Fi access point signals. Triangulating these signals on an individual basis provide examples of a coarse-grained approach to location determination.[30] However, when combined with each other and degraded GPS signals (when these are all that are available) the fingerprint that results from the combination is unique and can produce 2D coordinates with a median accuracy of 1-3 meters. Skyhook Wireless is an example of company that provides location services based on Wi-Fi.[31] Verizon Wireless is an example of service provider that uses cell tower triangulation, Wi-Fi triangulation and fingerprinting to assist in the location of phones that temporarily cannot find a GPS signal.

We raise these systems because they can be of use in identifying the location of photographs that are not properly geotagged or perhaps not located at all. Dobson, in a five-part series on "Navigation, Pedestrians, Landmarks and Local Search" included a speculation on why Google was collecting such diverse data on its Street View platform and developed the following model to explain the effort.[32] (See Figure 22.)

[30] See Anthony LaMarca and Eyal deLara. Location Systems, An Introduction to the Technology Behind Location Awareness. Morgan and Claypool. 2008 p. 102
[31] See Skyhook Wireless at http://www.skyhookwireless.com/4
[32] See Exploring Local http://blog.telemapics.com/?p=140

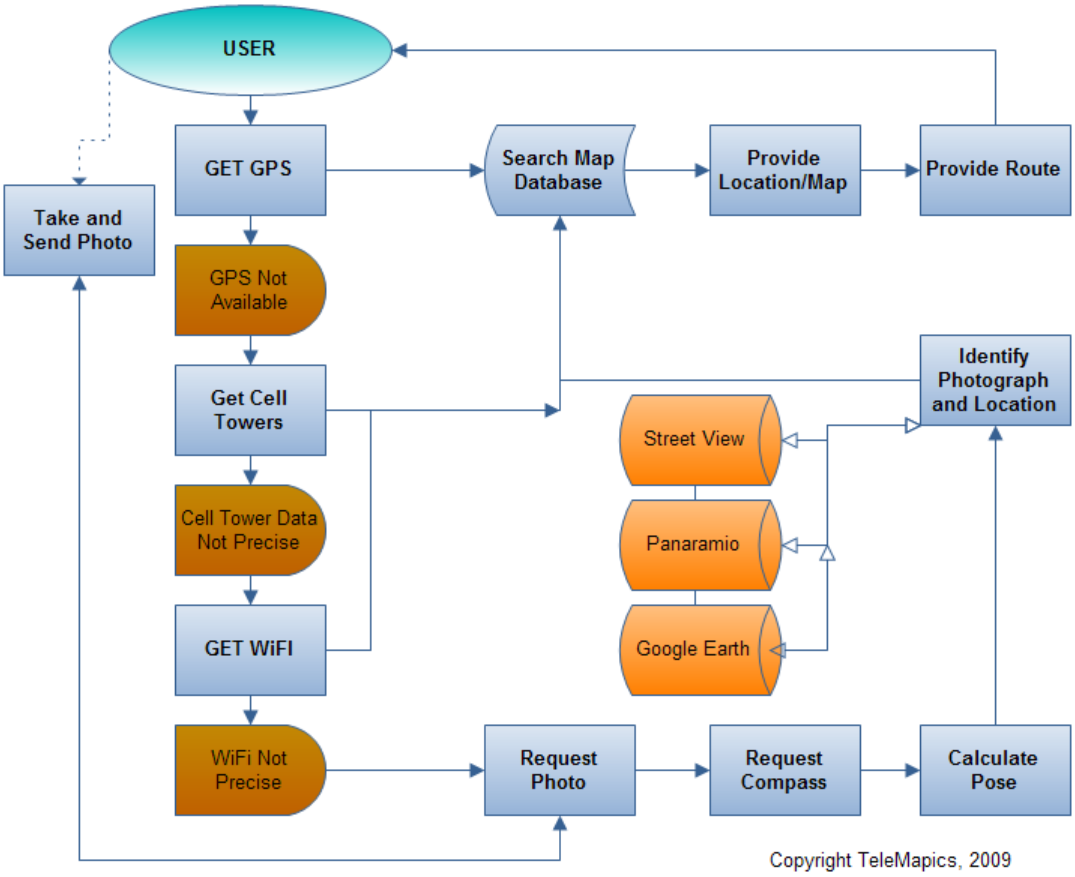


Figure 22. Example of potential use of photography to determine location during pedestrian navigation. If the user takes a photograph when lost, if the pose when the photograph was taken is known, the photograph can be used to with other location identifiers determine both the location of the user and the direction they are facing.

1.2.5.1 Photography and Location Defining Software Tools

In the same way that a user of Google Maps could take a photograph of their location to determine where they were located, Census Field Representatives (FRs) could do the same thing to describe the location of any location they photographed, even using low resolution imaging. The image would need to stamped or attributed with geographic coordinates and additional information about the pose of the camera. Although the geographic coordinate will reveal the location where the photograph was taken, it will not reveal any information about where the camera was pointed in the area surrounding the coordinate. In essence, information on where the camera was being pointed at the moment the photograph was taken could be used to determine which house the camera was aimed at during photography.

An example of a simple, but comprehensive solution to calculating the pose of a camera can be found in Theodolite Pro for the iPhone. The Theodolite Pro software is a multi-function, augmented reality (AR) application that serves as a compass, GPS, map, zoom camera,

rangefinder, and two-axis inclinometer. The program overlays real time information about position, altitude, bearing, and horizontal/vertical inclination on the iPhone's live camera image.[33] Theodolite Pro includes reference angle mode, A-B calculator for height, distance heading, position, triangulation and relative angles, email export with KML data, percent grade display, mil compass readout and an optical rangefinder. Examples of the types of data output from this $3.99 application are shown below (See Figures 23, 24, and 25). The important issue is that systems that functioned in this manner could allow photographs of living quarters to be precisely located and the image could be added to the MAF as a point-in-time location description.

Figure 23. Example of Theodolite Pro image showing readings related to a specific photograph that includes the address of a residence.

Figure 24. Specific output data for the scene is shown on the upper left of the image.

[33] See Hunter Labs at http://hunter.pairsite.com/theodolite/index.html

Figure 25. Theodolite Pro outputs the location of the image as a KML file that can be used with Google Earth.

1.2.6 Issues with the use of Road-side Imagery and Location Photography

Road-side imagery and other methods of location photography, while potentially valuable tools for the identification of the location of living quarters, rely on technology that is relatively new and based on proprietary designs. In addition, with the exception of Google's Street View, the technology has not been tested in large scale implementations. Next, these methods may produce data that is difficult associate with the appropriate records in the MAF, since most would rely on matching location coordinates with the MSPs collected by the Census Bureau and depend on the validity of both sets of recorded coordinates. In other cases, road-side imagery capture might collect street names and addresses not yet in the MAF, raising question about the QC procedures that would be used to accept, reject or further research these data. Next, it is possible that the street centerline coordinates collected by vehicles equipped with road-side imagery platforms might be more accurate than the centerlines captured in TIGER, provoking matching and snapping issues, as well as requiring new methods for merging these data in a manner that advantages the Geography Division, such as by reducing costs while increasing the accuracy and coverage of the MAF.

Another potential issue is whether or not the data from road-side imagery capture systems could be image processed to reveal the addresses on buildings and mailboxes that contain these data. While it is likely that the capture and interpretation of the address signs required by E-911 systems could easily be converted to address data, it is clear that research and development would need to be undertaken in order to determine the image parameters and processes that would enhance the ability to capture all address information available in the road-side images.

Another potential hurdle is researching the manner in which road-side imagery databases could be linked to the MTdb. Given the accuracy of TIGER's centerlines, it should be possible to match the centerlines collected by road-side imagery vans with their representations in TIGER. Addresses revealed through interpreting the road-side imagery and assigned location coordinates based on their positions would have to be matched to the MAF based on searches of neighboring MSPs. In locations where there were no neighboring MSPs, the locations could be accepted or rejected based on measures of the positional integrity of the sensing platform at the time of the data capture.

1.2.7 Summary

In a fundamental manner, road-imaging brings the field environment to the office and takes field crews off streets, roads, and otherwise removes personnel from dangerous locations. In addition, it provides a detailed representation of the physical environment that could be interrogated to provide an unambiguous location description involving coordinates, photography, and other measurements, if desired (distance from a corner, distance from a landmark, distance from a driveway, etc.), to assist field examination should it be required. If the Geography Division determines that road-side imaging is not appropriate for the tasks facing Census, then photographs attributed with data revealing the pose of the camera could be used with a variety of online sources to determine the location of the image and provide a location description that might be superior to those now stored in the MAF.

2. Sub-Task 2: Local and International Address Standards Approved and in Progress

Addresses are primarily created and maintained locally and, yet, their standardization for various use-cases is often managed as a national or international issue. In the past, standardization was often linked to issues of "mailability" or to formatting mail for automated processing. However, the development of GIS and various forms of land and property information systems have witnessed an expansion of interest in the exchange of data on the location of a residence or group living quarters. While the notion of addresses connotes an association with postal services and mail, the trend in address standards has been focused on finding a more effective method to describe the physical location of an address, whether it be a parcel centroid, geographic coordinate or some other form of unambiguous location description. Although postal addresses may never become passé, they are now, most often, referenced as a profile that is part of a larger standardization effort focused on the use of addresses and addressing for applications beyond receiving mail.

2.1 The United States Thoroughfare, Landmark, and Postal Address Data Standard

In 2010 the United States, the Federal Geographic Data Committee (FGDG) published the draft *United States Thoroughfare, Landmark, and Postal Address Data Standard* for public review and the standard is on track for a January, 2011 adoption. Publication 28 of the U. S. Postal Service is regarded as the current standard for addressing mail to be processed and delivered within the United States and its island possessions. A version of the USPS Publication 28 addressing standard has been proposed for use as a profile in the new standard and has the additional benefit of being harmonized with the FGDC Cadastral Standard for parcel information. Harmonization with the FGDC Transportation Standard will be completed at the Transportation Research Board meeting in January, 2011.

The FGDC coordinates the development of the National Spatial Data Infrastructure (NSDI), which is comprised of the policies, standards, and procedures for organizations to cooperatively produce and share geospatial data. Federal agencies that make up the FGDC develop the NSDI in

cooperation with organizations from State, local and tribal governments, the academic community, and the private sector.

The "Street Address Standard", known alternatively as the "United States Thoroughfare, Landmark, and Postal Address Standard", and "the United States Street, Landmark, and Postal Address Standard", is a draft data processing standard for United States address information. While addresses are used for governance and to provide government services, they are also used by businesses and individual for various purposes. Accurate, high quality, exchangeable address are required by all users, but address sources usually are local, address forms usually are diverse, and a uniform standard for exchanging the unique address forms does not currently exist. In addition, although Publication 28 from the USPS provides the official mailing address standard for the United States, many mailable addresses are unrelated to the location of the living quarters for those receiving mail. Post office boxes, rural route boxes and HCR box address forms provide no indication of the spatial location of the residence and make it difficult for government or businesses to provide services that require being onsite (e.g. emergency services, package delivery, or following up on a the status of a Census form).

While addresses in the United States are largely created by local address authorities (although sometimes under the authority of counties or more rarely, states), there is no official standard for address data in the United States,[34] Nor is there a standard that ties addresses to physical location or any feedback loop for correcting addresses on lists transferred between various users of address data. The Street Address Standard is an attempt to resolve the quandary surrounding address and addressing practice in the United States. [35]

The maintenance authority for the standard is the Census Bureau under the auspices of its duties as lead for the FGDC Subcommittee on Cultural and Demographic Data. The Subcommittee is chaired by Randy Fusaro, a staff member of the Census Bureau's Geography Division. The Street Address Standard, after adoption, will be reviewed on a five-year schedule as stipulated for required updating.

The goals set by the organizers of the draft standard are as follows:

- Provide a statement of best practices for street address data content and classification,
- Define tests of street address data quality,
- Facilitate exchange of address information,
- Offer a migration path from legacy formats to standards-compliant ones,
- Provide for different levels of standardization,
- Build on previous FGDC address standard efforts, and

[34] The situation is compounded by the fact that there appears to be no listing by area of the agencies responsible for addresses or the procedures and update practices of these organizations.
[35] FGDC. Geographic Information Framework Data Content Standard: FGDC-STD-014.7-2008. See http://www.fgdc.gov/standards/projects/FGDC-standards-projects/framework-data-standard/GI_FrameworkDataStandard_Part7_Transportation_Base.pdf

- Make this standard compatible with, and supportive of the FGDC framework standard for geospatial data, and with the NENA and USPS address standards.[36]

The draft standard is over 600 pages in length and provides significant detail on street address content and data specification. Rather than provide a detailed discussion on its comment, we were asked to extract the major points of emphasis in the draft standard and present this information in Table 2 below.

The proposed standard includes "…profiles that provide, for a particular application of the base standard, either a restricted subset of the standard, or a limited extension of a standard that does not contradict the base standard or both."[37] For example, the Postal Addressing Profile, which is under review by the USPS, is designed to establish the relationship between the base Street Address Standard and USPS Publication 28, which is the Postal Addressing Standard for the Nation. Specifically, the Postal Addressing Profile:

"1. Shows which parts of the FGDC standard relate to which parts of USPS Publication 28.
2. Relates USPS Publication 28 postal addressing terms to the address elements and attributes defined in the Content Part of the base standard.
3. Extends and restricts the base standard to conform to USPS addressing specifications.
4. Describes how address records that comply with the base standard may be brought into conformance with USPS addressing specifications, and vice versa."[38]

[36] Martha McCart Wells, et.al. Overview of the Draft U.S Address Data Standard. PowerPoint presentation. 2009
[37] FGDC. Postal Addressing Profile of the United States Thoroughfare, Landmark and Postal Address Data Standard. Provisional Draft (under review by the USPS) p2. (The Postal Addressing Profile is included in Committee draft at http://www.fgdc.gov/standards/projects/FGDC-standards-projects/street-address/?searchterm=address%20standard)
[38] Ibid

Table 2. Overview of the proposed Thoroughfare, Landmark, and Postal Address Data Standard

Street Address Data Content	Street Address Data Classification[39]	Street Address Data Quality and Testing[40]	Street Address Data Exchange
Address Elements	**Thoroughfare Address**	**Elements of Quality**	**Forms of XML-based Exchange Enabled.**
Address number Street name Occupancy Landmark name Larger area elements -place name USPS postal address USPS address line	A Thoroughfare Address specifies a location by reference to a thoroughfare.	Dataset Purpose And use Attribute accuracy Logical consistency Completeness Positional accuracy Lineage Temporal accuracy	1. Monolithic or complete exchange 2. Transactional or Incremental exchange
Address Attributes	**Landmark Address**	**Testing Address Quality**	**Required Elements**
Address ID Address Coordinates Descriptive attributes Attributes describing Specific elements Spatial organization attributes Address lineage attributes	A Landmark Address specifies a location by reference to a landmark	Tests of address elements Test of address attributes Tests of address classes	Address data Metadata
	Postal Address		**Exchange process**
	A postal Delivery Address specifies a point of postal delivery which has no definite relation to the location of the recipient, such as a post office box		1. Normalize data to undo localizations 2. Reparse data into an address class 3. Express data in XML format of the standard 4. Prepare metadata describing the data being exchanged
	General Address		
	The General Address class holds address of any class and is a starting point for parsing and classification.		

[39] The classes of addresses are defined by syntax, that is by their address elements and the order in which these data elements are arranged. See the full standard for more detail. Available online at http://www.fgdc.gov .
[40] Address quality issues are discussed in deliverables 11 and 12 for Task 5: Measuring Data Quality.

While numerous address standards exist across the United States on state and local levels, it is unlikely that the addressing standards for Clay County, Minnesota,[41] or the address standard for the Oregon[42] will produce anything other than local compliance. In fact, local and state standards often are designed to serve only as local interchange standards that will allow organizations within a government unit to use address data from another unit or from an adjoining municipality. In due time it is likely that all of these minor address standards will be replaced by the Thoroughfare, Landmark, and Postal Address Data Standard now moving towards adoption by FGDC.

While many may consider voluntary conversion to a national standard for addressing to be unlikely, we believe that the need to exchange addressing data within local communities, as well as between local communities and higher levels of government (e.g. to support regional or state agencies and projects) will lead to these data providers adopting standards that reduce the complexity of interchanging the information they collect. In addition, as location services evolve, more and more commerce will depend on standardized comprehensive, up-to-date, address databases. A national standard, such as the FGDC standard, which is well thought out, comprehensive and inclusive, will likely become an attractive option to those who deal with addresses and addressing systems. We think the FGDC Street Address Standard will be particularly attractive to organizations that recognize the benefits of writing their data once and exposing it to the network for use by others. Regardless of its attractiveness as an interchange format, a comprehensive adoption of the FGDC Thoroughfare, Landmark, and Postal Address Data Standard may take a number of years.

2.2 International Standards

On the international front, the Universal Postal Union (UPU) and the International Organization for Standardization (ISO) through Technical Committee 211 (TC 211) are the major standards organizations with interest in addresses and addressing. The European Committee for Standardization (CEN – Compté Européen de Normalisation) is responsible for European address standards, but is working through the UPU on this topic. The proposed international standard ISO 11180:1993 'Postal Addressing' was withdrawn in 2004 by committee responsible for it (TC154) and a new proposal on addressing, known as ISO 19160, is now being considered.

2.2.1 INSPIRE

INSPIRE (Infrastructure for Spatial Information in Europe) is a PSI Directive (Public Sector Information) of the European Commission (2007) that set the legal framework for the establishment and operation of a spatial data infrastructure for the 27 member states of the European Community, although its emphasis is on environmental issues. The program was a part of the eContentplus Program (now expired) and its work continues under the Information

[41] Clay County, Minnesota. Property Address Standard. 1998 See http://www.co.clay.mn.us/Depts/GIS/GISAddSt.pdf

[42] Joe Bernert. Oregon Geospatial Data Addressing Standard. 2004. See http://www.oregon.gov/DAS/EISPD/GEO/standards/docs/ORAddressContentStdv1.pdf?ga=t

and Communications Technologies (ICT) Policy Support Program (ICT-PSP) running from 2007 to 2013, as implemented under the Competitiveness and Innovation Program (CIP).[43]

The notion behind INSPIRE is that the spatial information of member states should be discoverable, compatible, exchangeable, and useable by members of the Community in a consistent manner. INSPIRE is focused on 34 spatial data themes (both reference and thematic) needed for environmental applications. The spatial data themes are subdivided and described in the three annexes of the directive.

Annex I is comprised of seven "Reference Data Themes" identified as the seven components of reference data of interest to INSPIRE.[44] INSPIRE Reference data has three principal requirements. It must provide an unambiguous location for a user's information, enable merging data from various sources, and provide a context to allow others to better understand the information that is being presented.[45]

The list of Reference Data Themes includes:

1. Geodetic reference data,
2. Units of administration,
3. Units of property rights,
4. Addresses,
5. Selected Topographic themes (hydrography, transport, height),
6. Orthoimagery, and
7. Geographical names.

An early definition of addresses used by INSPIRE indicated the persuasion that an address was a geographic construct : "Location of properties based on address identifiers, usually by road name, house number, postal code accompanied by a geographical name (city, village, town, suburb/municipality, admin unit)", although the physical location of an address is regarded as its most important component.[46]

INSPIRE's efforts appear to be focused on finding ways to harmonize address systems, believing that all national and local address systems share the same concepts and general properties. Unfortunately, harmonizing these data can be difficult because the manner in which house numbers are assigned and conceptualized throughout the EU, especially in rural areas, differs widely across the member states. For this reason, one of the measures being considered is not using house numbers "…but 2D or 3D geographic coordinates to represent real property."[47]

The 2009 INSPIRE Data Specification on Addresses, which will be published as an implementation guideline, defines the spatial data theme "addresses" as: "Location of properties

[43] Watelet, Marcel. ICT Policy Support Programme in the CIP – WP 2009, Teleregions Network Workshop, PowerPoint Presentation. 2009.
[44] INSPIRE. Drafting Team "Data Specifications" – Definition of Annex Themes and Scope. 2008.
[45] INSPIRE RDM. INSPIRE position paper: Reference Data. October 2002. P. 11
[46] INSPIRE. Drafting Team "Data Specifications" – Definition of Annex Themes and Scope. 2008. P. 28
[47] IBID. p.29

based on address identifiers, usually by road name, house number, postal code."[48] As noted above, the Specification on Addresses is focused on promoting interoperability between addresses and addressing systems of the Member States of the EU.

The Specification defines an address slightly differently than in earlier publications, as an address is now defined as: "An identification of the fixed location of a property, e.g. plot of land, building, part of building, way of access or other construction, by means of a structured composition of geographic names and identifiers."[49] The core of the Specification is focused on addresses and their relationships with attributes and covers the manner in which objects (such as land parcels and buildings) relate to property. In addition, in the Specification, addresses are related to a number of address components and metadata concerning their origin including an emphasis on the lifecycle of addresses.

The UML (Unified Modeling Language) class diagram in Figure 26 provides an overview of the main elements of the Address Specification's Address Application Schema, including the association with a country name and associations with additional address components necessary for the unambiguous identification and location of the address instances. The INSPIRE document provides a thorough discussion of the details of the model and we refer interested parties to this reference,[50] as well as Figure 26.

[48] INSPIRE. INSPIRE Data Specification on Addresses, 2009. P.VI
[49] IBID.
[50] IBID . P16

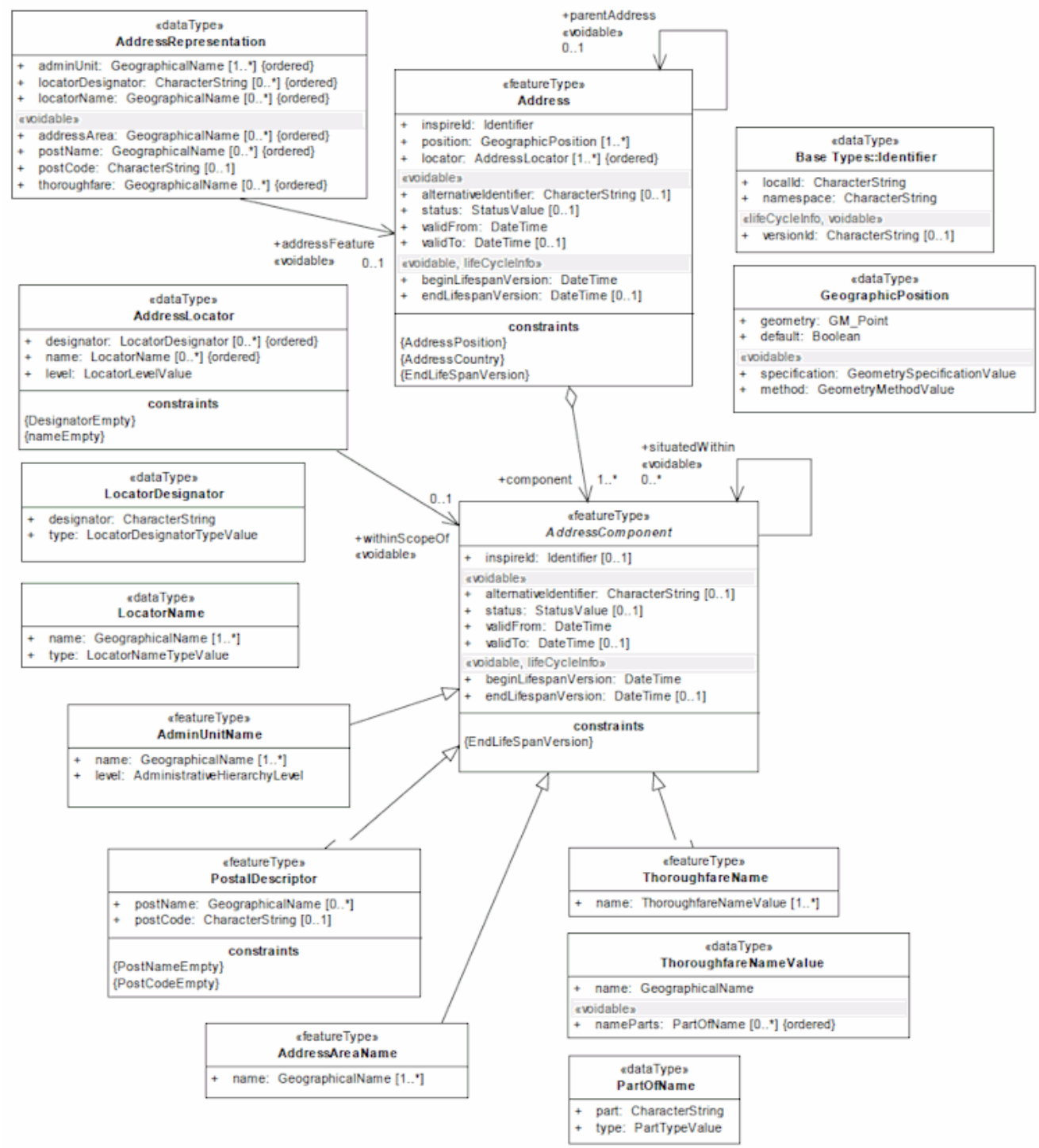


Figure 26. UML class diagram: Overview details of the Addresses Application Schema, in which the address must have an association with the name of the country in which it is located and must have associations to the additional address components necessary to the unambiguous identification and location of the address instance. The source document provides an excellent description of the components and attributes of this diagram,[51]

The notion of a voidable characteristic in the UML is described as follows by the Specification,

[51] INSPIRE. INSPIRE Data Specification on Addresses, 2009. P16

> "If a characteristic of a spatial object is not present in the spatial data set, but may be present or applicable in the real world, the property shall receive this stereotype.
>
> If and only if a property receives this stereotype, the value of *void* may be used as a value of the property. A *void* value shall imply that no corresponding value is contained in the spatial data set maintained by the data provider or no corresponding value can be derived from existing values at reasonable costs, even though the characteristic may be present or applicable in the real world."[52]

Voidable characteristics are a part of INSPIRE because the standard is an effort at promoting harmonization of data across the countries of the European Community. Unfortunately, existing country-based databases may not have collected all of the data characteristics stipulated by INSPIRE. All properties in the standard are required to be reported, if the relevant information is part of the data set. Properties may be reported as "void", if the data set does not include the relevant information.

The INSPIRE Address Specification, which describes the INSPIRE Generic Conceptual Model data specifications, is built around 22 requirements and 16 recommendations that mold the specifications into mandatory and recommended elements. The requirements largely maintain the integrity of the Specification by indicating actions that must be taken to preserve the integrity and interoperability of the Specification and how these actions must be implemented within the overall INSPIRE framework. The recommendations are best practices applied to the specific requirements set for the standard.

The major components of the Address Specification are:

- Data content and structure (basic notions, application schema addresses),
- Reference systems (coordinate reference systems, temporal reference system),
- Data Quality (completeness, positional accuracy, logical consistency, temporal accuracy, thematic accuracy),
- Dataset-level metadata (mandatory and conditional metadata, optional metadata elements, guidelines on using metadata elements),
- Delivery (delivery medium, encodings),
- Data Capture, and
- Portrayal (layer types, default styles, other well-defined styles, layer organization).

It is, perhaps, in its Annexes that the Specification is most informative, particularly with numerous descriptions and graphics describing the challenges of address assignment (Annexes E and F) and address assignment practices in Europe (Annex D). However, Annex B on Examples of Metadata Elements and Annex C on Address Component Life Cycle, although brief, are both informative and interesting. The types of Life Cycle information include: "1) the content specific life–cycle information describing the real world address (when this version of the real world address is valid); and 2) the temporal information on the changes in the database or spatial

[52] IBID. p7.

data set (when the item was inserted, superseded or retired).[53] An abstract test suite, whose requirements must be met by any dataset claiming conformance with the Specification, is in development and will eventually become part of the Address Specification as Annex A.

We believe that the infrastructure being developed to support the INSPIRE addressing standard and its interoperability/harmonization goals may provide useful information for the Geography Division as it contemplates partnerships to exchange address data. In our opinion, the INSPIRE standard is an example of best practices in exchanging address data across a landscape populated with diverse cultures and diverse addressing practices. Although many of the political considerations that have driven the adoption of INSPIRE are not specifically relevant to the operations of the Census Bureau, the INSPIRE standard reflects another perspective on dealing with the types of address and addressing problems that are commonly encountered by the Geography Division.

The INSPIRE Specification on addresses was developed in close cooperation with EURADIN, an organization we describe below.

2.2.2 EURADIN

It appears that the difficulties noted by INSPIRE have helped to create EURADIN (The European Address Infrastructure), aimed at creating a best practices network promoting European address harmonization regarding the definition, registration and access to European address data. The EURADIN effort initiated in 2008 and concluded in 2010.

The general objective of the EURADIN Technical Working Group (TWG) was "To significantly contribute to harmonizing the European Addresses, proposing a solution to achieve their interoperability, and thus facilitating the effective access, reuse and exploitation of that content, which will promote the creation of new added value products and services across Europe."[54] The Technical Working Group authoring the presentation indicated one of their goals was "To make the validation of the proposed harmonization model and Addresses Infrastructures through the development of a Pilot European Gazetteer Service giving access to the addresses of at least 13 European countries and/or regions".[55]

[53] IBID. VII
[54] EURADIN, EURADIN European Address Infrastructure, PowerPoint presentation. 2008. p 5
[55] Ibid. p6.

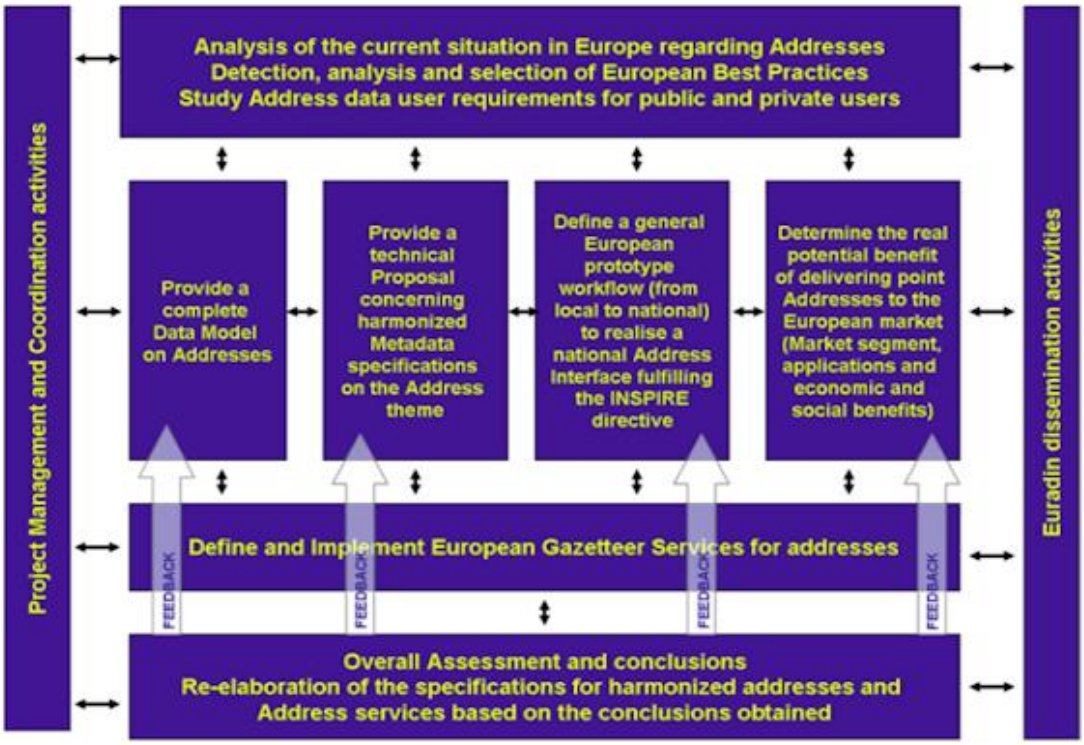


Figure 27. The overall goals of the EURADIN Project

However, EURADIN used the INSPIRE Address Specification described in the previous section on INSPIRE to harmonize their data. Creating and testing EURADIN's Gazetteer Service led its participants to conclude that integration of Pan-European address systems is possible but that no standard software solution exists to resolve many of the difficulties they encountered. Since their evaluation incorporated 20 address systems and 23 gazetteer services, the lack of standard addressing software capable of working with these diverse data is not surprising.[56] (See Figures 27 and 28.)

[56] [56] EURADIN, EURADIN Project 2008 -2010, PowerPoint presentation, 2010. pp. 20-21

Figure 28. The Proposed version of the EURADIN Gazetteer Service

The EURADIN organization promoted "Addressing the Future – First European Address Conference" in June, 2010 and launched the European Address Forum (EAF) at the meeting. EAF, the likely successor of the work carried out by EURADIN, is an open forum aimed at encouraging the development of well-constructed addressing processes and the availability and efficient use of harmonized and good quality address data throughout Europe using the INSPIRE framework.[57]

Among the key recommendations from EURADIN are these:

- There should be a single national "official" address reference database;
- The process should be sponsored and controlled by the Public Sector;
- The data should be owned by the Public Sector and probably the State and fit for the purposes of central and local government;
- The data must be comprehensive, up to date, capable of national integration and of following the INSPIRE standard for subsequent Pan-European integration;

[57] See the announcement at https://www.euradin.eu/Lists/Calendar/DispForm.aspx?ID=35 and the European Address Forum at http://www.europeanaddressforum.eu for more information.

- There should be an integrated creation and maintenance process; and
- The basic data should be free or available at the marginal cost of distribution.[58]

As noted previously, the eContentplus Program that was the source for INSPIRE and EURADIN closed in 2008. EURADIN expired in 2009 and INSPIRE and other geographic themes are now operating under the Competiveness and Innovation Framework Program (CIP) that runs until 2013. Although addresses are an INSPIRE theme and covered by eContentplus and CIP projects, there were no geography themes in the CIP-ICT call for programs in 2010 and it is not yet clear they will be included in the 2011 call for projects.

2.2.3 ISO 19160, Addressing

The ISO 19160, Addressing effort remains as a stage 0 project, that is, it is in a preliminary state. At this point the objectives of ISO for this initiative are to:

- Investigate and formulate requirements in relation to addressing.
- Make recommendations on whether standards should be developed and, if so, how this should be done.
- Not to write an address standard, but to review existing standards and propose how international standardization requirements should be developed.[59]

The addressing standards reviewed for purposes of ISO 19160 are shown in Table 3 below. Section 5 of the ISO 19160 Draft report provides an excellent review of all of the standards for the interested party. Conclusions about the need for a standard for addressing remain unfinished as of the date of the present report.

One conclusion in the report was of potential concern. In Section 7.2 it is stated that "Some postal addresses use Post Office Box numbers. These relate solely to a collection point at a Post Office, and are outside the scope of geographic information, and are therefore excluded from further consideration here."[60] Based on the other categories of addresses described (street or thoroughfare, proximity address, or geographic address), it would appear that the U.S. based address constructs such as rural route boxes and HCR boxes that do not include a specific geographic component might not be considered in the potential development of an ISO addressing standard.

[58] EURADIN, EURADIN Project 2008 -2010, PowerPoint presentation, 2010. P.25
[59] IS0/TC211 Geographic Information/Geomatics, Draft Review summary of project 19160, Addressing, 2010. P6.
[60] Ibid, p34.

Table 3. Standards Reviewed by ISO for Project 19160, Addressing[61]

Standard	Abbreviation used in this document	Publisher	Publication Date	Level	Justification
ISO 19112:2003, Spatial referencing by geographic identifiers	ISO 19112	ISO/TC 211	2003	International	International standard for spatial reference systems using geographic identifiers, such as gazetteers.
ISO/TS 15000-5, Electronic Business Extensible Markup Language (ebXML) -Part 5: ebXML Core Components Technical Specification, Version 2.01 (ebCCTS)	ISO/TC 15000-5	ISO/TC 154	2005	International	International standard for terminology and core components in electronic business and administration. Includes, for example, a number of different address profiles, each specifying a number of address related fields that should be used together on a form such as an invoice.
OASIS CIQ v3.0 Approved Committee Specifications CS02	OASIS CIQ	OASIS	2008	International	International address standard, especially interesting for its XML schema.
UPU S42, International postal address components and templates	UPU S42	UPU	2006	International	International address standard with specific focus on postal mail. A number of countries have already submitted postal address templates that specify how an address should be written on a mail piece.
UPU S53, Exchange of Name and Address Data	UPU S53	UPU	2009	International	International address standard with specific focus on postal mail.
AS/NZS 4819:2003, Geographic information – rural and urban addressing	AS/NZS 4819	AS/NZS	2003	International (Australia and New Zealand)	Regional address standard, especially interesting for its rural addressing scheme.

[61] ISO/TC 211 Geographic Information/Geomatics, Draft Review summary of project 19160, Addressing, 2010, p.8

INSPIRE D2.8.I.5 Data Specification on Addresses – Guidelines.	INSPIRE	INSPIRE TWG Addresses	2010	International (Europe)	Regional address standard that specifies how public authorities of EU member states must make their address data available. Therefore expected to have a significant impact as a data exchange standard within the EU.
AFNOR XP Z10-011, Specifications postales – Adresse postale	AFNOR	AFNOR	1997	National (France)	National postal address standard from a developed country.
SANS 1883-1, Geographic Information - Address	SANS 1883-1	SABS	2009	National (South Africa)	National address standard from a developing country with a wide variety of address types, including free text address types for informal settlement and rural areas.
BS 7666-0:2006, Spatial datasets for geographical referencing-Part 0: General Model for Gazetteers and spatial referencing	BS 7666	BSI	2006	National (UK)	National address standard from a developed country, based on ISO 19112 above.
Draft U.S. Thoroughfare, landmark, and Postal Address Data Standard	US FGDC	US FGDC	2010	National (USA)	National address standard from a developed country with a wide variety of address types and a separate part for address data quality.
Data in this table © ISO 2010					

2.2.4 The National Land & Property Gazetteer

The NLPG is the authoritative national address list that provides unique identification of land and property and conforms to BS 7666, the British Standard for Geographical Referencing. BS7666 was created in 1994 and last revised in 2005/2006 and provides address coverage for England and Wales.

The British Standard BS7666 (Spatial datasets for geographical referencing) is comprised of four parts including:

> "Street Gazetteers, Land and Property Gazetteers, Addresses and Rights of Way... BS7666 specifies a standard format for holding details on every property and street. The standard does not differentiate between commercial or residential properties, between occupied, developed or vacant land, between urban or rural or between addressable properties and non-addressable entities such as communications masts."[62]
>
> The UML for BS 7666, which provides a rapid way to see the main elements of the specification and their relationship, is shown below.[63] (See Figure 29.)

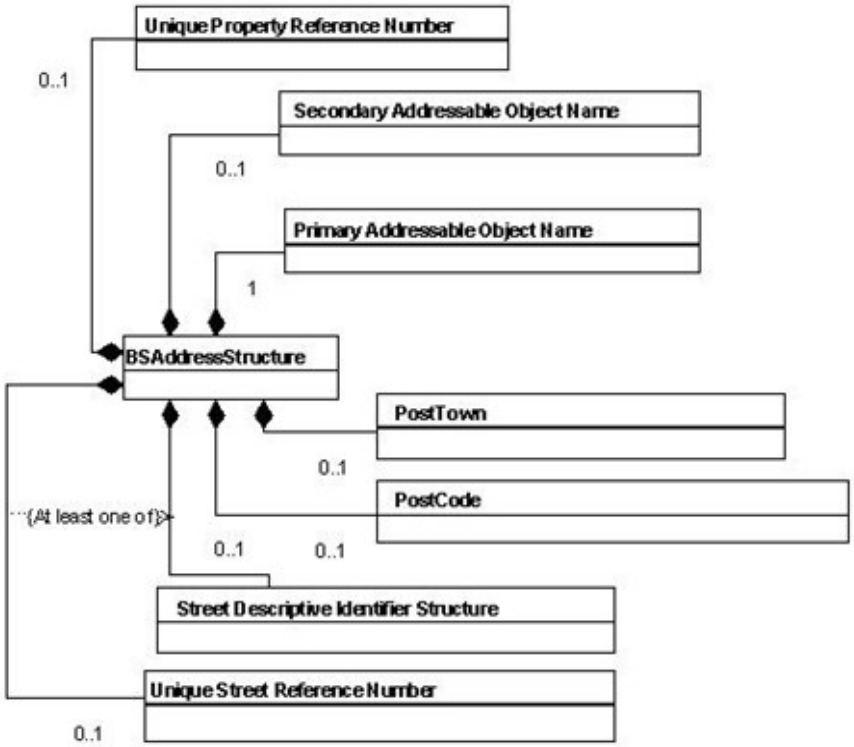


Figure 29. UML diagram for BS 7666 Addressing

[62] See http://xml.coverpages.org/bs7666.html
[63] See UK GOVtalk at http://interim.cabinetoffice.gov.uk/govtalk/schemasstandards/e-gif/datastandards/address/bs7666_address/bs7666_uml.aspx

The NLPG, which is supplied in CSV or XML format, is updated continuously by the local authorities in England and Wales who have statutory responsibility for street naming and numbering. Updates to the system are available on a daily basis. The data in the NLPG are guided by strict format and process standards. Unique persistent IDs (the UPRN – Unique Property Reference Number) identify land and/or property and their associated records (including addresses). Provision has been made for sub-divisions of addressable objects such as flats, and other units within residential and commercial buildings. These Secondary Addressable Objects (SAOs) are addressed by reference to a Primary Addressable Object (PAO), for example a building name or a street number. In addition, the standard covers residences that do not have mail services. Coordinates for the GB National Grid are provided for all referenced locations, as is an Extent Polygon representing each Basic Land Property Unit (BLPU). Metadata, including the official address, alternative addresses, and historic address, is stored as appropriate. [64]

The NLPG is an authoritative and exceptionally thorough implementation of a spatial dataset for geographical referencing. The existence of a similar dataset in the U.S. appears unlikely. We believe that many local governments in the United States do not have computerized land record system, nor any computerized method of updating or communicating this information to others. In fact, CoreLogic employs teams to visit country courthouses and extract these records in order to build their, as of yet incomplete, parcel data base of the United States. Next, some counties (e.g. Orange County, CA) regard their property and address databases proprietary and will not license them without payment of a fee. Finally, the size of the United States and the complexities of its landholding may impede the development of a national parcel or cadastral data base. We regard these hindrances as unfortunate, since a parcel/address database that was updated daily could provide the Geography Division with an authoritative, current and comprehensive source of address information with which to conduct its operations. It should be noted that a controversy surrounds the fee-based NLPG and the desire of local communities and businesses to have a public (free) version of the address register, since most of the components of the address register have "…long resided in the public sector."[65]

2.3 Summary

Standards for addresses and addressing appear to be moving away from a focus on postal addressing standards toward interest in spatial datasets for geographical referencing that include postal addresses as one of many forms of address information that could help locate or describe the location of a living quarter or property. Most geographic referencing standards are focused on establishing geographic coordinates to describe the location of living quarters and property, as well as linking addresses with a persistent ID that remove any definitional ambiguity from the records used to identify them.

We reviewed a variety of standards and believe that nationwide adoption of the FGDC United States Thoroughfare, Landmark, and Postal Address Data Standard could be of significant benefit to the address related activities of the Geography Division. The standard appears to be a well thought out and comprehensive. While the FGDC standard will be adopted by agencies of the Federal Government and may eventually be adopted by the local and statewide agencies

[64] NLPG, BS7666 Explained. Undated. See http://www.iahub.net/docs/1183553456634.pdf
[65] See http://www.ukauthority.com/Headlines/tabid/36/NewsArticle/tabid/64/Default.aspx?id=2864

responsible for addressing throughout the United States, our suspicion is that this conversion will take a number of years, perhaps decades, before it becomes a true national standard.

In respect to international standards, we believe that the infrastructure being developed to support the INSPIRE addressing standard and its interoperability goals may provide useful information for the Geography Division as it contemplates partnerships to exchange address data. INSPIRE spinoffs like the EURADIN effort focused on gazetteers and interoperability may provide valuable information on techniques for exchanging addresses that could be of use to the Geography Division. While we were impressed with the UK's NLPG, we think the nature of governance in the United States precludes such a model from ever being implemented on a national scope. Finally, it appears that the ISO project 19160 is too early in its potential development process to be of significant benefit to the Geography Division.

The standards reviewed are not proscriptive, that is, they do not dictate (or even suggest) the schema for an operational address database such as the MAF. They do provide sets of data elements that encapsulate information that an operational database should produce. The degree to which this data is available in the MAF is unclear. For example, the address elements used in the MTdb are all present in the draft *United States Thoroughfare, Landmark and Postal Data Address Standard*, but there are no explicit fields in MAF that correspond to the data quality measures enumerated in the draft standard. It is quite possible that equivalent data quality information exists in the MAF and could be inferred from the internal processes and rules used to populate, match, certify, and maintain MAF. One way to evaluate the information content of the MAF against the standards would be to determine which address standard data elements could be explicitly valued.

3. Sub-Task 3: Research the future directions of Postal Addressing

In this section of our report we will report on two topics of specific interest to the Geography Division. Our first examination explores crowdsourcing, often called Volunteered Geographic Information (VGI), and whether this form of data compilation might provide additional address information that could be used to improve the MAF. The second topic covered deals with HCR Addresses, their status and possible expansion.

3.1 Crowdsourcing/Volunteered Geographic Information

The "architecture of participation", "crowdsourcing", "User Generated Content" (UGC), "Volunteered Geographic Information" (VGI), "Neogeography", "Participatory Geography", "Citizen Science", and "Citizen Mapping" are just a few of the various terms used to describe the sharing, cooperation, and collective action that have allowed crowds of people with ordinary skills to build and publish geographic databases without the need for coordination, management, or an overarching business structure. As noted by Shirky, it is now possible to achieve large scale cooperation at a low cost based on serious, complex work taken on without institutional direction. [66] In some fundamental sense, the introduction of the Web and the associated collapse in the costs of communication, reproduction, and publishing, as well as the introduction of powerful, low-cost technology, has created a new ecosystem in which some of the functionalities once the domain of the professional classes of map makers, geographers, and GIS practitioners have been transferred to the general public.[67]

The UGC that is produced by users of websites and digital media is contrasted with the efforts of traditional media producers such as broadcasters, production companies, publishing companies, and address and map database companies. The most common names for these types of efforts that are applied to efforts for gathering maps and addresses are crowdsourcing, VGI and Neogeography. Primary examples of relevance to the Census Bureau are OpenStreetMap's attempt to build an attributed street-level crowdsourced map of the world, and OpenAddresses' attempt to build a world-wide database of addresses. We examine these efforts to determine their utility for enhancing the MTdb of the Geography Division. In addition, to the standard approach to crowdsourcing, we will examine hybrid approaches that blend crowdsourcing as a complement to the database compilation efforts of existing organizations. In this document we will focus on crowdsourcing for addresses and addressing, while we will focus on crowdsourcing for map data in Deliverable 6 for Task 3 titled the "Current and Anticipated Future Directions of Potentially Useful Developing Technologies".

[66] See Clay Shirky. Here Comes Everybody. Penguin Press. New York. 2008

[67] See Don Tapscott and Anthony Williams. Wikinomics. Penguin Books, London. 2008 for detailed discussion of crowdsourcing and Wikinomics. The work does not address geographic wikis, but is one of the authorities on crowdsourcing.

3.1.1 What is Crowdsourcing/VGI?

In his landmark book the Wisdom of Crowds[68] James Surowiecki postulated that in many circumstances the crowd is smarter than the trained professional, or even the smartest member of the crowd. His book caused little interest at the time of its publication, but the popularization of the Internet along with its influence on social networking caused renewed interest in his research which prompted a decided focus on the types of accomplishments that might occur if "listening to the crowd" could be transformed into an information source.

The Internet provides everyone who is connected to it with the equal opportunity to publish any information of interest to them, as well as to communicate and collaborate on their own terms when it is convenient for them to do so (usually asynchronously). Perhaps more importantly, the Internet and its services have provided the infrastructure for mass collaboration through social networking (see Figure 30). Online collaboration benefits from collective intelligence, aggregated knowledge from decentralized sources, and low cost. The influence of these benefits are often described as Wikinomics [69]

While Social Networking is a complex phenomenon, three of its most important aspects can be classified as "me sharing", "thing sharing", and "community information sharing". While websites like Facebook are clearly about individuals sharing information about their personal lives (i.e. "me sharing"), other sites, such as Wikipedia, are focused on thing sharing (that is collaborators on the websites contribute their skills, either in the form of content, editing or fact checking). Still other websites like Patch, an AOL product that functions as an online local newspaper, and sites like OpenStreetMap, function as local-social sites focused on community information sharing.

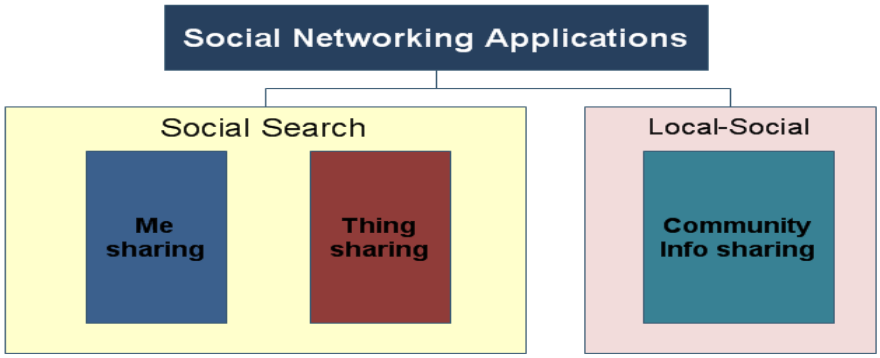


Figure 30. Social Networking Applications and mass collaboration

The desire for community information sharing focused on geography led to the notion of Neogeography and the desire to create free, worldwide databases of geographic information using mass collaboration via the medium of social networks to provide geographic data and attributes. There are a number of potential benefits to crowdsourcing geographical data, but foremost among these is the spatial adaption of Linus's Law (Linus Torvald, the creator of

[68] See James Surowiecki, The Wisdom of Crowds, Random House, New York 1967.
[69] Op. cit. Tapscott and Williams

Linux) from this "Given enough eyeballs, all bugs are shallow"[70] to this "Given enough eyes, all spatial errors are shallow." Linus's law is really about the numerousness of contributors or reviewers and one would assume that there should be more willing contributors interested in researching spatial data than there are professional data compilers focused on gathering addresses and other spatial data. Further, the distribution of contributors to crowdsourced systems aimed at gathering spatial data should be more widely distributed than the professionals, employed by government organizations, or businesses focused on compiling addresses and spatial data.

It is important to note that the contributors to crowdsourcing generally exhibit self-selection (they contribute on a specific theme, because it is of interest to them) and self-benefit (the contribution will help them, or people they are interested in helping, solve a general problem in their location). One of the major criticisms of crowdsourcing is that some geographic situations are too complex to be understood without access to detailed resources that are often not readily available to the public (See Figure 31). In addition, it is likely that contributions to the crowdsourced database may not be uniform in terms of spatial coverage, as they reflect the distribution of population density in general.

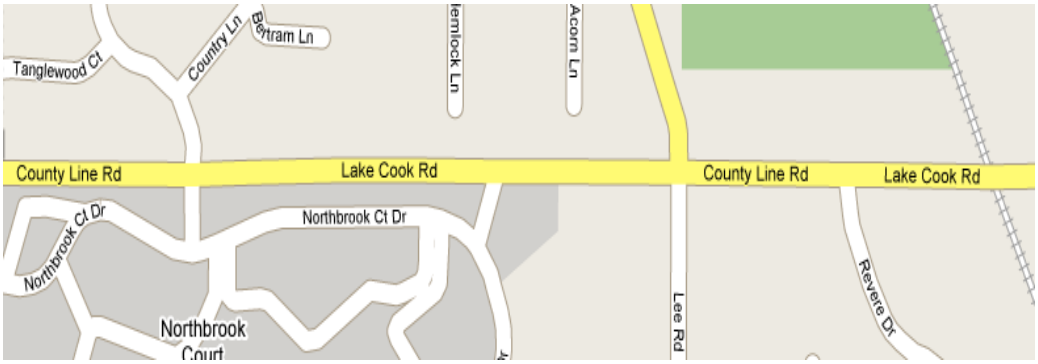


Figure 31. The main road in this image is on the border between Lake and Cook counties in Illinois. Not only do the street name change by jurisdiction, but the address ranges reset every time a jurisdiction changes. This is an area that is not a good candidate for crowdsourcing, since jurisdictional borders can be difficult to verify based on visual inspection while in the field.

Next, it is possible that user priorities may lead to unreliable contributions or prejudice in response (spata is the acronym for spam data). For these reasons, extensive error-checking may be required to validate the crowdsourced data, but accuracy validation is often not a formal part of crowdsourced systems. It is important to note that crowdsourced systems are regarded as "self-healing" over time. Once again, this is the application on Linus's law in that crowdsourced systems generally rely on other users to correct mistakes in the database. Errors in pure crowdsourced approaches to address compilation, at least theoretically, should decrease over time, just as the comprehensiveness of coverage should increase over time.

[70] See Wikipedia http://en.wikipedia.org/wiki/Linus%27_Law

3.1.2 The potential weaknesses of a crowdsourced approach

Crowdsourcing is a potentially powerful method of data collection whose results can be impacted by a variety of method-induced limitations. In the following section of our report we examine several of the salient issues with the use of crowdsourcing to collect data on addresses and other geospatial data.

3.1.2.1 Contributors and contribution patterns

Budhathoki et. al. analyzed the demographic of the contributors to OpenStreetMap (OSM) and noted that predominately they are: male, under forty years of age, have some college education, and that a large percentage of the users have had some prior experience with GIS and maps. See Figure 32 for more details on the specifics of the demographics.[71]

Often, there are not enough data gathers to validate data or a method to redeploy them in areas with inadequate or erroneous coverage. Budhathoki and Nedovic-Budic indicate that while a there are a large number of OSM "participants" (around 180,000), only around 30% of the registered users have ever contributed geographic information and of those who have contributed approximately 40% have contributed only once (See Figure 33).[72]

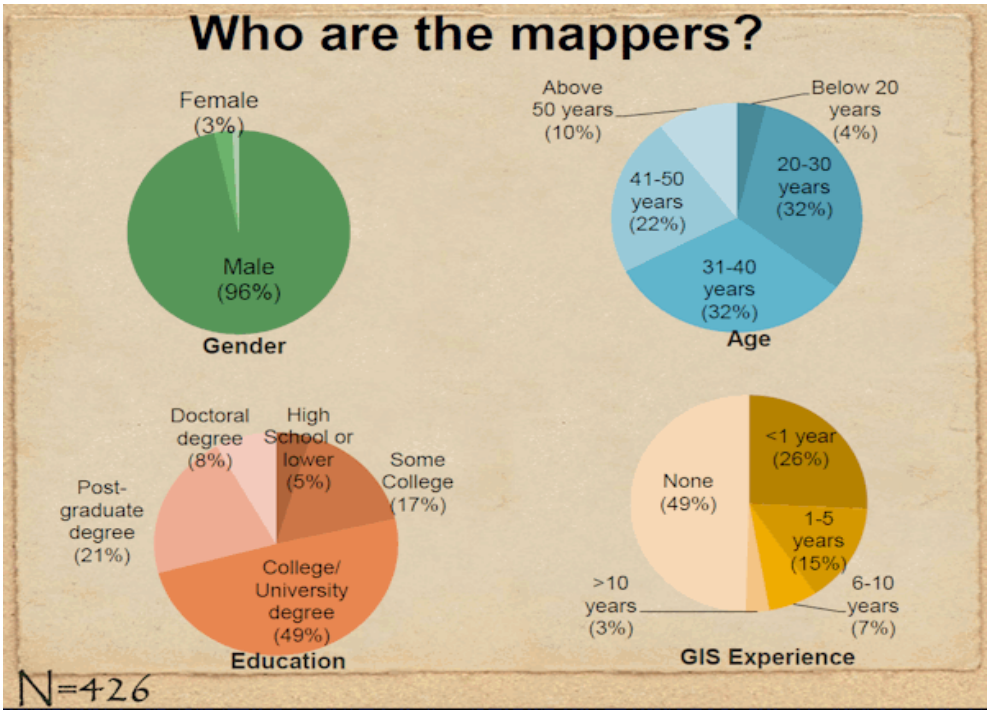


Figure 32. Budhathoki et. al. The Demographics of OSM contributors

[71] Nama R. Budhathoki, Muki Haklay, Zorica Nedovic-Budic, Who Map in OpenStreetMap and Why? State of the Map 2010. Atlanta, Georgia, 2010 (PowerPoint presentation)
[72] Nama R. Budhathoki, Zorica Nedovic-Budic, How to motivate different players in VGI? GI Science, Zurich 2010 GI Science, Zurich 2010

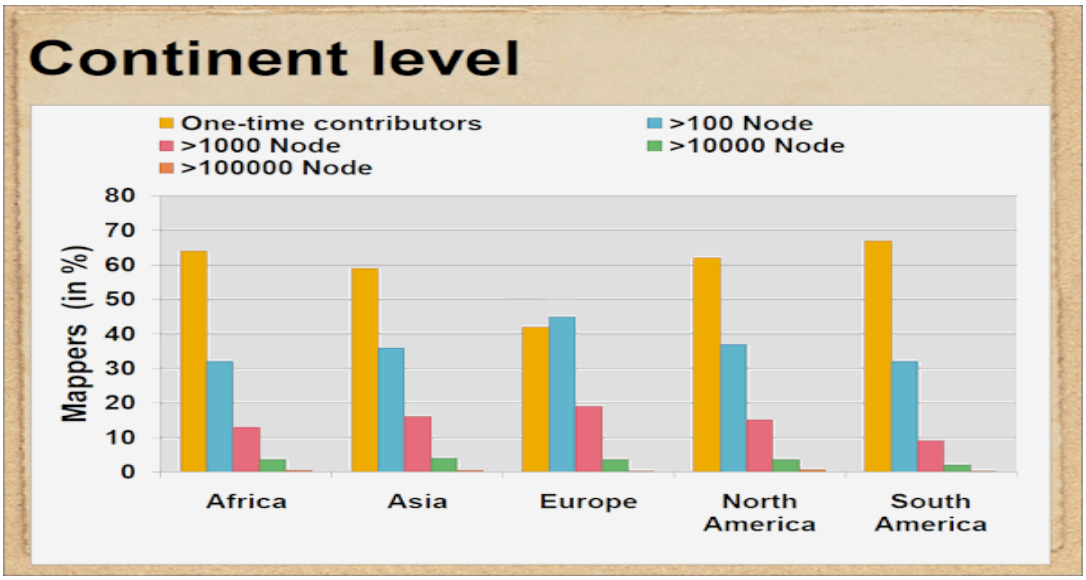


Figure 33. The distribution of contributors to OSM

The distribution of contributions to OSM, as shown by Budhathoki et. al. (above)[73] roughly follows a power curve function that is commonly found in crowdsourcing efforts. For example, Shirky indicates that less than two percent of Wikipedia users ever contribute and that among the contributors no effort is made to even out their contributions and that this pattern of the "long tail" appears to apply to social interaction, as well as commerce.[74] In his 2008 assessment of OSM, Haklay contends that approximately ten users accounted for 45 percent of the nodes collected in the OSM database at that time.[75] Detailed data about the numbers of contributors who added attributes to the nodes is available but has not been analyzed.

3.1.2.2 The consequences of the distribution of contributors

The distribution of those willing to contribute spatial data about their local area often does not provide uniform spatial coverage. Recently, 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 more areas were covered with unattributed (often even unnamed) roads.[76] Perhaps of more importance to the Census Bureau, these authors also indicate that because interest in mapping is likely only to be found in a small percentage of the population, population density is an important factor in determining the comprehensiveness of spatial data and explains a bias in the OSM coverage toward highly populated areas.[77]

---

[73] Nama R. Budhathoki, Muki Haklay, Zorica Nedovic-Budic, Who Map in OpenStreetMap and Why? State of the Map 2010. Atlanta, Georgia, 2010 (PowerPoint presentation)
[74] Shirky, op.cit. 1616-23
[75] Dr. Muki Haklay, Understanding the quality of user generated mapping, PowerPoint Presentation, 2008 http://www.slideshare.net/mukih/osm-quality-assessment-2008-presentation?from=ss_embed
[76] Haklay and Ellul, Completeness is volunteered geographical information, Journal of Spatial Information Science, 2010. p8.
[77] Ibid, p 15

One of the countries with the most developed OSM coverage is Germany. In 2010 Zielstra and Zipf undertook a study that compared the freely available data from OSM for Germany with the latest version of the commercially available Tele Atlas Multinet Database.[78] Our interest in the study is that the results revealed:

> "In all cities studied, the diversity of the freely available data is significantly higher between inner-city and rural areas, which can be explained by the presence of more active members on the project in the larger cities. Further tests showed that the completeness of the data is marked by strong differences between the large and medium-sized cities. By using a circular buffer method, a significant decrease in the data was observed as the distance from the city center increased."[79]

The decrease in OSM coverage with increasing distance from population centers coupled with the relatively limited attribution of the roads (including addresses) that have been collected to date appear to be limiting variables that may not be resolved using standard crowdsourcing procedures.

It is, also, important to note that mapping efforts based on crowdsourcing are often based on users "war-driving" routes while recording their path on a GPS unit. This method-induced limitation is the major reason that many of the streets in the OSM UK database are unattributed. While "probe" data (also called passive crowdsourcing) can be an excellent technique for capturing road geometry and definition, it does not reveal information about addresses or the numerous map attributes that cannot be inferred from GPS traces. In the OSM's United States database, there is comparatively more address information, but only because the database benefits from the import of a version of TIGER from 2005.

Although crowdsourced systems are thought to be self-correcting over time, it is not yet clear that the movement towards crowdsourcing is one that is self-perpetuating or perhaps one that will be successful in every geographical location. The questions regarding what incents a contributor to provide geographic data to OSM remain largely unresolved. Budhathoki et. al. have provided modest evidence that the contributor's desire for free map data and the chance to share knowledge of local areas combine to promote contributions to OSM.[80] Budhathoki, however, indicates that while time is a partial explanation for the large number of contributors to OSM in Europe, it "…does not fully explain the unusual number of contributors (and their contributions) in Europe compared to other continents" and he continues by noting that the high cost of spatial data availability in Europe may be another important factor influencing the number of contributors.[81]

By all accounts, the progress of the OSM in the United States has lagged behind the effort in Europe, possibly because of the availability of TIGER and other spatial data that are in the public domain. OSM has imported a version of TIGER 2005 that was not uploaded until 2007.[82] The OSM model retains data on the provenance of the information used to build its database and

[78] Dennis Zielstra, Alexander Zipf, A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany, 13th AGILE International Conference on Geographic Information Science, 2010
[79] Ibid, p14.
[80] Budhathoki, et. al, Op. cit. slide 17
[81] Nama Budhathoki, personal communication, August 29, 2010.
[82] Wiki OpenStreetMap. See this link http://wiki.openstreetmap.org/wiki/TIGER

OSM has indicated there have been many edits of the TIGER data in some cities in the United Stateas, while they also indicate that many other cities have not been edited at all We believe that the spatial pattern of these edits reflect the distribution of population in the United States and we have concerns that the crowdsourced model will not be able to provide comprehensive map or address coverage due to this limitation. While it would seem logical that OSM consider importing a new version of TIGER to remedy this problem, they have indicated it is unlikely that a more up-to-date version of TIGER will ever be imported. The reason for their concern is that "…it will be difficult to determine if differences between future TIGER and OSM are due to good corrections made by OSM editors or they are from bad TIGER data.[83] While this may be a reasonable strategy for OSM, it may make using OSM data more difficult for the Geography Division, if such an interest exists.

3.1.2.3 Editing and Standards

A crucial problem in crowdsourced systems is that they often deemphasize the importance of structured editing as a compilation standard. For example, a debate is currently raging in OSM related to the use of automated edits. The OSM Wiki takes this stance:

> "Mappers are what makes OSM work. They are the ones spending countless hours out in the open collecting data, or hammering their notes into an editor and uploading data to OpenStreetMap. They take pride in their work and often have a sense of ownership for their contribution. If your script changes something that is obviously wrong, e.g. a typo where someone wrote "hihgway" instead of "highway", then it surely does them a service. However, if you start judging their work and modifying it according to what you think is right then you might actually break something that someone has put in there on purpose. Be very careful with such edits and ask yourself: 'Am I absolutely sure that I'm not cluelessly interfering with something that is well thought out?'"[84]

Finally, a lack of uniform standards and quality control are inherent aspects of crowdsourced systems. For example, OSM allows its participants flexibility and does not have any restrictions on assigning tags to nodes, ways (streets and roads), or areal features. Although though there are recommendations for contributing data, there are no standards. For example, the OSM Wiki states:

> "There is no standard for mapping addresses, house numbers, postal codes, etc. in OpenStreetMap, although most people use the Karlsruhe schema for addresses. This layer shows address information tagged according to the simple model of the Karlsruhe Schema. It does not include the 'relational option'."[85]

While the Karlsruhe schema is now an approved feature[86] (one that was approved through a voting process), it not used by everyone who contributes address data to OSM (and as noted previously, those who do contribute attributes are a significant minority).

---

[83] Ibid. Wiki OpenStreetMap.
[84] Wiki OpenStreetMap. See this link http://wiki.openstreetmap.org/wiki/Automated_Edits/Code_of_Conduct
[85] Wiki OpenStreetMap. See this link http://wiki.openstreetmap.org/wiki/OSM_Inspector/Views/Addresses
[86] Wiki OpenStreetMap. See this link http://wiki.openstreetmap.org/wiki/Approved_features

Without belaboring the point further, it is clear that OSM and similar efforts can be useful for a variety of purposes, but that there are limitations hindering its wholesale use for reference purposes. It is likely that the future of crowdsourced methods, at least in respect to addresses and mapping, is its use in hybrid systems that advantage themselves by combining the best of standards-based compilation systems with crowdsourced approaches to gathering address and spatial data.

3.1.3 Hybrid Address and Mapping Systems

There are at least three approaches for the use of crowdsourcing in the creation of map data.

1. Create a unique map database from scratch based on User Generated Content (see the OpenStreetMap database of the United Kingdom and Germany).

2. Create a "new" derivative map database by using an existing public domain database while updating it through crowdsourcing (see the OSM site for the United States).[87]

3. Augment existing government/commercial database with User Generated Content and supplement these updates with professionally compiled data for the sole purpose of providing a more up-to-date version of the database that meets the original purpose for creating the database.

Our interest is in the third possibility and we will discuss how this technique could be used to create a hybrid database potentially of benefit to the Census for updating, expanding, and augmenting the capture of addresses and other spatial data. It is important to note that the hybrid approach, which seeks to advantage an existing database by fusing it with crowdsourced data, is not what OSM and other groups had in mind when they began their efforts at spatial database building. To a significant degree the crowdsourcing movements want their data to be free and desire the same for any data that is combined with their data in a derivative work. Due to Title 13 restrictions on the use of addresses, this issue could cause difficulties for the Census Bureau and we will address that topic later in this discussion

3.1.3.1 Spatial Database Creation

A standards-based spatial database results from a quasi-waterfall process, with recirculation of data being an important aspect of the system design (see figure 34). The process is designed to optimize spatial data quality, temporal relevance, and metadata. Initially the data are collected according to specifications and quality controlled to meet those specifications. Inaccurate data are re-collected or placed in the update queue for further examination. Ongoing research operations are targeted at change detection, then updating areas of high change at the earliest opportunity. Other changes are discovered by systematically working research teams through the entire database coverage over time. The overall assignment is designed to increase the integrity of the database while maximizing the time value of money spent on updating. It is the ongoing, active attempt to harmonize the database, its elements, attributes and metadata that distinguishes standards-based compilation from crowdsourced compilation efforts. In essence, the standard-

[87] Wiki OpenStreetMap. http://wiki.openstreetmap.org/wiki/WikiProject_United_States

based compilation method imposes a discipline on the data compilers that should be reflected in the quality, coverage and temporal aspects of the data.

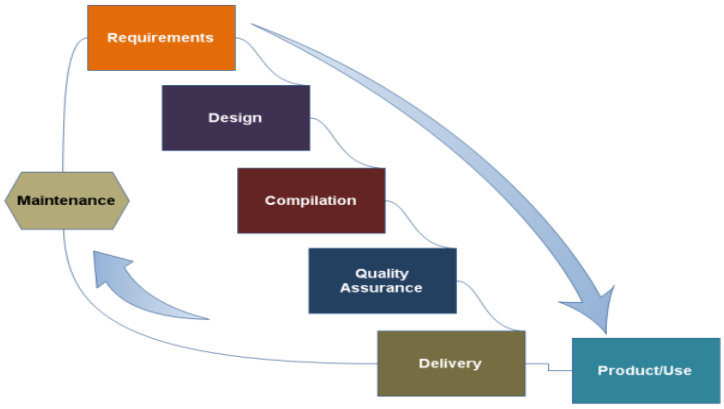


Figure 34. The standards-based spatial database building process

A useful method for envisioning the combination of address and mapping systems based on standards and crowdsourced data gathering systems is the comparison tables (Tables 4 and 5) below.

Government/Commercial	Crowdsourced
Training in compilation	Self-selection
Specialization	Local experience
Staff size limited	Contribution pool potentially unlimited
Research limited	Research hours limited to free time
Salaried	Avocation –contributing is its own reward

Table 4. Comparison of Compilation in Government/Commercial and Crowdsourced environments.

Government/Commercial	Crowdsourced
What are my coverage goals	How many people will contribute?
What are my accuracy goals	Where are they located?
How much can I spend on updating	How long will it take to get good results over large coverages?
What size of staff can I afford and how can I incent them to create the best database possible?	How do you motivate collaborators over long periods?

Table 5. Comparison of potential constraint in traditional compilation and crowdsourcing. The government/commercial approach to compilation places coverage goals and accuracy at the top of the requirements matrix. The Wiki depends on "good will" to build its database and is unable to predict either coverage or accuracy, but knows that both should improve over time.

However, even with the benefits that standards provide, these funded database building efforts often produce data are too expensive to maintain. In addition, these "professionally compiled" databases often are: out-of-date, inaccurate, non-comprehensive, of variable quality, and too expensive to maintain. For example, NAVTEQ's database extension and update costs in 2007 were over $300,000,000 and yet in some locations, it appears that OSM line data are equal or close to the quality provided by NAVTEQ, although lacking in the array of attributes that NAVTEQ provides.[88] It is for this reason that numerous organizations involved in mapping are attempting to include some form of crowdsourcing as a fundamental aspect of their address and map database building efforts.

TomTom/Tele Atlas, for example, uses both active (contributed) and passive crowdsourcing (tracks from GPS units equipped in vehicles) to build their navigation database. NAVTEQ has considerable passive crowdsourcing assets, but little in the way of active corrections contributed by users of their database. NAVTEQ is hoping to use active crowdsourcing, based on use map corrections suggested by users of NOKIA's popular Ovi Maps.

It is a little realized issue that organizations desiring to collect crowdsourced data for address and map corrections benefit from being "customer-facing" or from offering a product or service that people find of interest. Tele Atlas's success with crowdsourced data is through TomTom's Personal Navigation Devices (PNDs) that include MapShare software that makes it possible for users to actively suggest data changes, while the system passively records details of the GPS tracks following their daily activities(see Figure 35).

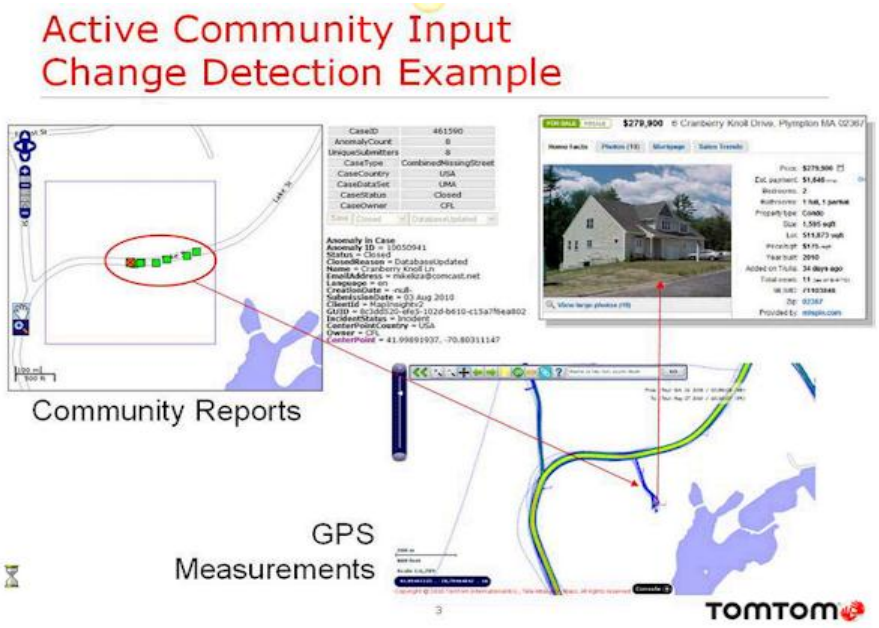


Figure 35. Report of a new street and residence made by active Community input through TomTom's Map Share. The input included the correct street name and the correct address. In addition, the passive community input (GPS tracks) showed the location of the street and its geometry. Image courtesy of TomTom/Tele Atlas.

[88] This calculation was based on details revealed in NAVTEQ's publicly disclosed quarterly financials (10q) for 2007. The data comparison information was based on a comparison of data tiles between OSM and NAVEQ as conducted by the principal author of this report.

The question of importance here is whether a government agency that hopes to successfully use crowdsourcing to collect data would be seen by potential contributors as a customer-facing entity that deserved their support. For example 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. (We discuss the FCC Broadband initiative later in this document under Sources of Addresses.)

We think that this model could be adapted as a form of crowdsourcing address information for the Census Bureau, but are unsure of its potential conflict with Title 13. Having a website, such as the American Factfinder, where you can enter your address and have Census information returned for your area, is of interest to many people. Trapping the user's IP address and then geolocating it and associating it with the address entered might provide a helpful source of address information for validating the MAF. Further, the form presented to the user could request that they confirm the location of their address on an aerial image of their neighborhood that could be zoomed to identify individual houses.

We believe it more likely that harvesting OSM for address information would be more advantageous than the Geography Division or the Census Bureau attempting to establish itself as a crowdsourcing agent. Due to the limitations placed on the Census Bureau related to Title 13, it is unclear that the Census Bureau would be considered a "open source" by the general population. In a related example, the OSM community does not trust Google to continue to provide free maps and in some respects, OSM's success was a response by many of its contributors to Google's license that required any crowdsourced contribution to be assigned a joint copyright protecting the user and Google. The OSM community interpreted the Google license as taking advantage of the data contributors by using their data, but not allowing them a perpetual license to use Google's mapping data in return. A more balanced opinion, it seems to us, is that Google needed the copyright release to use the crowdsourced data in the first place, since it had not generated the data, but was responsible for publishing it.

We think it unlikely that the crowdsourced community would to contribute data to government agencies and many commercial businesses, for that matter, in the volume required to make crowdsourcing a success. As more and more groups solicit crowdsourced data for addresses and maps, they serve to dilute the pool of crowdsourcers willing to contribute to specific system. Simply put, the Census Bureau might have difficulty trying to establish itself as a meaningful repository for the compilation efforts of potential crowdsourcers. For this reason, an association with the OSM community would be beneficial to the Geography Division as a data partnership. In addition, the insights on technology stacks and map compilation methods gained by such a relationship might be of significant value in future data partnerships.

It is our recommendation that the Geography Division consider adopting crowdsourcing as a method of enhancing its data through agreements with commercial and open source providers of these data. We believe the model below makes sense for the Census Bureau, as well as other agencies of the federal government. In this model some entities may only be harvesters of data provided by others, while other entities may be both harvesters and providers (see Figure 36).

For example, the Geography Division may choose to provide TIGER data to others, harvest specific USGS data and decline to share its address data due to Title 13 restrictions.

In addition to the existing links between government "networks" of content sources and users, we believe that User Generated Content (UGC is another term incorporating the concept of crowdsourcing) will become as useful a data source as the data offered by commercial data providers. Further, it is our opinion that UGC will initially enter the network through its adaption by commercial providers, which will eventually lead to direct interaction with "open" data providers.

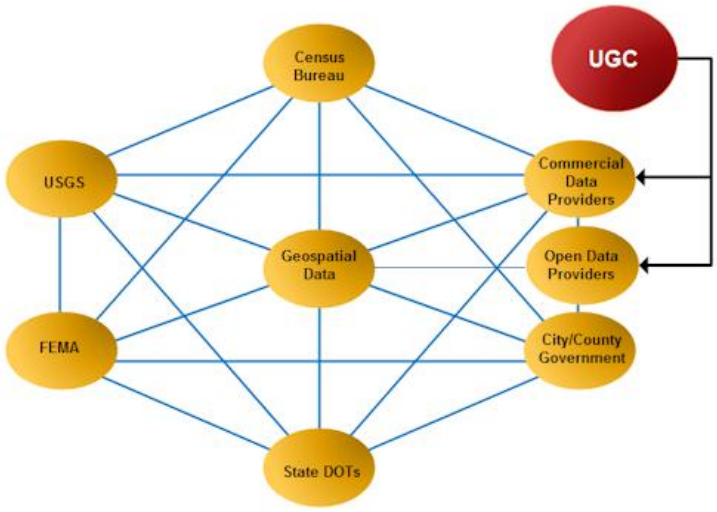


Figure 36. Crowdsourcing as a supplement to data sourcing procedures at various levels of government. User Generated Content (UGC) is another term for crowdsourced data.

As soon as feasible, the Geography Division should consider using crowdsourcing data made available from open data providers and commercial data providers. For example, an association with OpenAddresses.org could, over time, become a valuable source of address information that might not be available for updating the MAF from other sources.[89] OpenAddresses aims to be the OSM of the address community and is described as an Open Source web portal for the management of open, worldwide, postal addresses. The effort is geared at storing only geolocated postal addresses[90] and the addresses are not associated with persons, companies, point of interest, or any other identifiable entity, other than the location of the address. (See Figures 37 and 38.)

To date, donors of United States-based addresses to the organization include: the City of Washington, DC, the State of North Carolina, the State of Arkansas, the State of Indiana, and

[89] See the OpenAddresses website at http://openaddresses.org. In addition, see the Wiki pages at http://code.google.com/p/openaddresses/w/list

[90] Geolocated postal addresses are addresses for which the geographic location is known and usually described in geographic coordinates.

City of Portland, OR. It is too early in the game to predict with any certainty, but our opinion is that OpenAddresses may find it difficult to collect enough data to make them a major source of addresses. While the announcement of the founding of OpenAddresses was greeted with enthusiasm, the site has recorded few new, major data partners since that time. In addition, the system of recording addresses lacks standardization and verifying the actual existence of the contributed addresses appears to be a weak link, at least as presently configured. However, it is difficult to project the success of crowdsourced projects and we recommend that the Geography Division consider establishing an informal relationship with OpenAddresses.

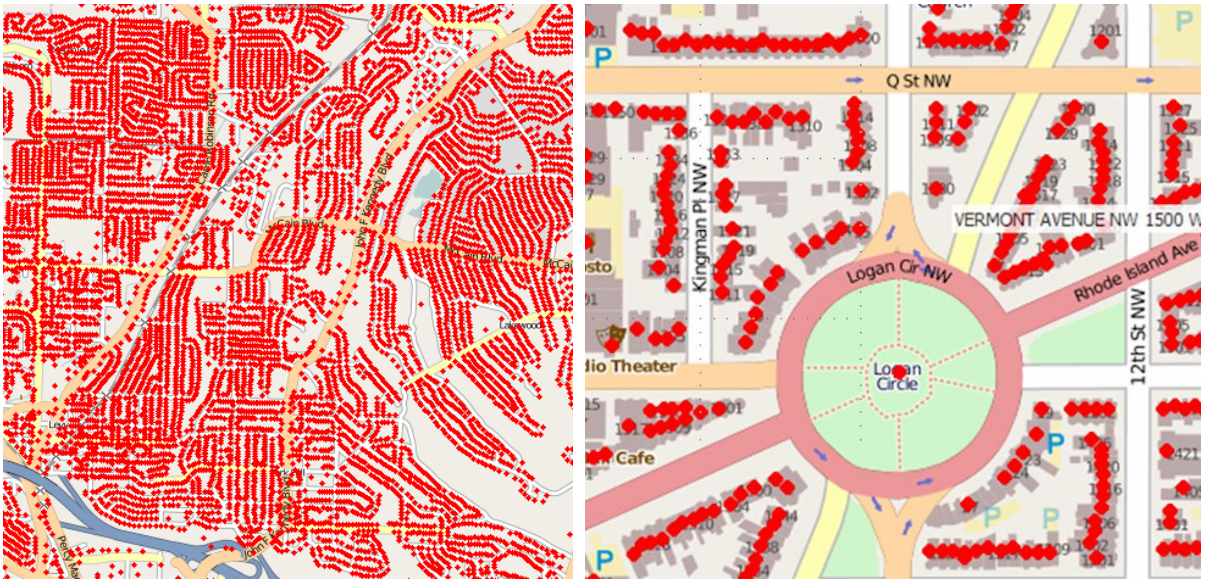


Figure 37. Examples of OpenAddresses.org address coverage in Arkansas and Washington, D.C.

Figure 38. Address Editor used by OpenAddresses Contributors

3.1.4 The potential fly in the ointment

OSM data, until recently, was licensed under a Creative Commons license with a ShareAlike attribution, commonly known as CC by SA.[91] The ShareAlike attribution requires that if you redistribute the work, you must do so under a license similar to the CC by SA. One of the historic problems with using OSM data has been related to the fact that the contributors to OSM did not want to create a database that could be used by commercial organizations (specifically Google) to a create a commercial database while impoverishing the one created by the industrial endeavor of the OSM contributors. As a consequence, the addition of the ShareAlike attribution has somewhat limited the use of the OSM data by organizations that might hope to repurpose it or improve it for their own purposes. In essence, if you change the database and redistribute it under license for a fee, you must also redistribute an identical version available without the fee under the CC by SA license.

However, the OSM is in the process of migrating to the Open Database License (ODbL) due to problems related to the use of the CC by SA 2.0 license applied to databases. Details on the ODbL license are best interpreted by legal experts, but the new license appears to be one that could allow use of these data by the Census Bureau.[92]

The issue of interest here is that the new license is focused on the "public use" of the data. That is, you must: attribute any public use of the data and you must keep the database "open' if you publicly redistribute the database. However, it would appear that the non-public use of these types of data would not violate the ODbL. OpenAddresses, however, continues to use the CC by SA 2.0 generic license mentioned above. In any event, the tension between the ODbL and the requirements placed on the Census Bureau concerning Title 13 and addresses must be examined by the legal division of the Census before publicly available crowdsourced data that is produced under license is used by the Geography Division.

3.1. 5 Summary

Crowdsourcing and activities that harness its abilities, like OSM, are changing the face of the compilation of addresses and other spatial data. MapQuest, for example, is running a version of the OSM database for specific countries, while it continues to rely on NAVTEQ for its standard mapping product in the same countries. However, the formal MapQuest site is provided updated address and map data by NAVTEQ on a quarterly basis. The MapQuest-OSM site takes updates from OSMPlanet every minute and updates its OSM website every ten minutes with map tile updates. Search results are updated in 2 to 24 hours and directions are updated on a daily basis.[93] While Census would not have need of ten-minute updates, a source that was more flexible than the multi-month update required for the DSF might be of considerable interest to the clients of

[91] See this page for the creative commons SA 2.0 generic license http://creativecommons.org/licenses/by-sa/2.0/
[92] See http://www.opendatacommons.org/licenses/odbl/
[93] Michael W. Dobson, Exploring Local http://blog.telemapics.com/?p=334#comments . See the comments section item #4 for details on the updating process.

the Geography Division, although constantly changing, unevaluated street data could play havoc with the geographic boundaries that are maintained by the Geography Division.

Crowdsourcing relies on local knowledge as the engine for its updating. From a practical standpoint, crowdsourcing would appear to be the best source for acquiring detailed local knowledge. Factors related to the distribution and motivation of contributors to crowdsourced data systems, however, place some limits on efficiency of using crowdsourced data for problems that require national coverage. At the present, neither the OSM nor OpenAddresses databases have the required geographic coverage or the attribute accuracy in the rural areas that are of significant interest to the Geography Division under the GSS initiative. Conversely, it is our opinion that many commercial sources suffer similar limitations in rural areas. As a consequence, an association with OSM and OpenAddresses may benefit the Geography Division by providing useful data and knowledge on alternative methods for collecting address information.

3.2 Expanded Use of Highway Contract Routes

Although the term Highway Contact Routes (HCR) covers general transportation contracts to haul mail between locations, it also includes Contract Delivery Services (CDS). The USPS Publication32 – Glossary of Postal Terms,[94] defines HCR as follows:

> **"Highway contract route** (HCR) - A route of travel served by a postal contractor to carry mail over highways between designated points. Some HCRs include mail delivery to addresses along the line of travel. Formerly called star route."

The name "star routes" was not a description of the geometry of the routes, but a reflection of the fact that in 1845 the contractors delivering this mail were tasked with delivering the mail with "celerity, certainty and security". Postmasters soon tired of writing the term out and simply began abbreviating the routes as "***", which soon generated the term "star routes". The use of the term "star routes" was abandoned in 1970 by the USPS and the term Highway Contract Routes is the official replacement term.

According the USPS Handbook P-5 "Highway Contract Routes – Box Delivery Services" October, 2004[95]

> "A box delivery[96] HCR is a contract agreement between the Postal Service and a private individual or firm for the delivery and collection of mail from homes and businesses. The services provided by the HCR box delivery carrier are identical to those provided by postal rural route carriers. The supplier, depending upon the contract requirements, performs specified duties such as maintaining a roster of customers; casing mail; selling stamps, certified mail, collect on delivery (COD), Express Mail, registered mail, Parcel

[94] http://www.usps.com/cpim/ftp/pubs/pub32/pub32h_p.html
[95] http://www.branch38nalc.com/sitebuildercontent/sitebuilderfiles/P5_HIGWAY_CONTRACT_MANUAL.pdf
[96] Box delivery routes are similar to rural delivery service and provide home or business delivery of mail.

> Post, merchandise returns, or money orders; picking up mail from collection boxes; and markup and forwarding of mail. A route may be operated personally by the supplier (owner-operator) or the supplier may elect to hire employees. If hired employees are used regularly, they must be paid the wages and fringes as required by the current wage determination attached to the contract."

HCRs were created to provide mail service to unusually remote areas or areas with relatively low populations densities, at least compared to the delivery points now considered "Rural", by the USPS. It appears that the informal dividing line between Rural Delivery and HCR delivery is that routes with an average of one residence or business delivery per mile are candidates for rural delivery, while those below this number are candidates for HCR delivery. Whether a residence is a candidate for city delivery, rural delivery, or HCR is established through the determination made by local postal managers.

3.2.1 HCR Address Forms

Some of the mail delivered by HCR contractors is addressed using a format that includes the USPS contract number of the route along with the box number of the delivery address, as follows: "HC 68 BOX 23a." However, HCR delivery points also include rural addresses that have been converted to E911 mail, formatted with a city-style address (note that these addresses are often called "street-style" addresses by USPS personnel). According to USPS Publication 28 on addressing[97] there are variations in the format of HCR delivery addresses (including addresses that use the old Star Route designation) and these variations can be found in Appendix 2 on Highway Carrier Route Addressing Formats.

Although HCR may have a unique address form (see below) they are accounted for in the DSF, provided to the Geography Division by the USPS, under the "R" Record Type Code and the Detailed Value code "F" is used to denote that the address is a "residential contract box."[98]

3.2.2 Numbers and Growth

Between 2001 and 2009 HCR delivery points increased by approximately 600,000 units, trending somewhere near 75,000 additional delivery points a year, although the increase between 2008 and 2009 was sixty-thousand units.[99]

In 2003, the USPS changed the official designation of the HCR routes in their Domestic Mail Manual to include Highway Contract Routes. Although the increase in the number HCR that include Contract Delivery Services appears modest, apparently the USPS has considered expanding these types of routes (although limiting the expansion to new delivery routes) as their

[97] See specifically http://pe.usps.com/text/pub28/28c2_005.html
[98] This information is in a Geography Division Excel Spreadsheet titled "New DSF Layout".

[99] See Table 1 Number of Delivery Points by Type of Delivery

cost is significantly less than routes where mail is delivered by a USPS direct employee.[100] Anecdotal evidence gathered from newsletters and website of interested parties[101] indicates that HCR are being used to provide service in towns such as Puyallup, WA (population 35,690), Beaverton, Oregon (population 86,000), and cities in southern Florida and elsewhere, rather than being restricted to rural communities. Whether or not the USPS continues to expand the use of HCRs is a topic that we recommend the Geography Division pursue directly with the USPS since it is unclear whether the potential expansion of HCRs also indicates an expansion of HCR forms of addressing in to non-rural locations.

3.2.3 Conclusions

HCR are Contract Delivery Services (CDS) that deliver mail to over 2.6 million delivery points and appear to be adding delivery points at the rate of 75,000 units a year. The HCR delivery addresses are may include a box designation, or addresses that have been converted to city-style E911 addresses. I t may be possible that the Census Bureau could work with the USPS in an attempt to mine route data (e.g. locations of delivery points) from the drivers who service the HCRs, presuming that this narrowing of the search window for rural living quarters would be of interest to either the Geography Division or the Field Division.

4. Sub-Task 4: Research the progress of E911 Addressing

Enhanced 9-1-1 (E911) is a telephone, landline-based system that routes an emergency call to the appropriate 9-1-1 Public Safety Answering Point (PSAP), which relies on a query of the Automatic Location Identification database (ALI) to identify the caller's residential address and phone number, if the call is from a landline phone. The conversion of addresses to emergency response systems (E911) has required each residence in the area being converted to be assigned a city-style address, at least for purposes of emergency response services.

Many of these E911 addressed locations were previously identified only by a rural route address, box number, or Contract Highway address used for postal identification. However, these forms of addresses did not indicate the location of the residence of the mail patron. Ideally, the transformation of rural addresses to city-style addresses in the E911 conversion process, especially when the E911 address is used as a postal address, could result in an expansion of the comprehensiveness and accuracy of address listings and possibly location descriptions for inclusion in the MAF.

E911 addresses, once assigned, are inventoried in the ALI database of the PSAP and immediately used for the provision of emergency services. Unfortunately the new E911 city style addresses may not always be used for mail delivery. Whether the E911 address also serves as a postal address can be a requirement based on a local or statewide ordinance in some areas

[100] Kevin. R, Kosar, 2008 The U.S. Postal Service's Use of Contractors to Deliver Mail, CRS Report for Congress, See http://www.nrlca.org/userfiles/File/USPS_Use_of_Contractors.pdf
[101] E.g. "postalmag.com http://www.postalmag.com/usps-contracting-out-mail-delivery.htm

and at the option of the household resident in other jurisdictions. E911 addresses that are submitted for mail delivery and subsequently approved can then be used for mailing purposes, although the replaced rural address remains active with the USPS for a one-year period.

In rural areas where houses were identified by non-city-style addresses, the implementation of E911 required assigning permanent city-style addresses to residences and signage at the location of the residence. The signage is used to identify the location of the address and the preferred point of access to the residence (usually a driveway), thus allowing emergency responders to find the physical location of the residence and render the required emergency services.

In order properly identify residences in E911 conversions so they can be found by emergency services, the owners of residences are required to place highly visible signage with the new E911 address on their mailbox, or if the mailbox is not located in front of their house, to place a signpost with their address next to their driveway. If the house is visible from the road, the signage can be placed on their house near the door or closest section of the residence to the road. The E911 address signage could be used by Census personnel during field operations as a reasonably accurate and unique "location description" for identification of residences who owners have opted to maintain their non-city-style postal addresses.

Our research has indicated that E911 address conversion is a complex and variable process. In some communities the process involved remapping street and road centerlines at high accuracy levels, renaming all streets and roads with unambiguous, easy to differentiate names, and providing address numbers for each residence based on a GPS coordinate recorded for the house and a distance measured from a coordinate origin set for addressing purposes within the area being converted. In other areas the locations of the residences were digitized from USGS quadrangles, using existing street names and new E911 addresses were conformed to the residences that had pre-conversion city-style addresses.

4.1 Detail

The major steps of the "recommended E911 address conversion process" often included:

- Establishing an Addressing Committee.
- Establishing addressing standards and enacting an addressing ordinance.
- Naming all roads with unique names and updating a base map.
- Providing GPS coordinates for all residences within with the area, including a plan for adding new residences and accommodating future growth.
- Assigning address number to each house using the position of the GPS coordinate on map grid whose origin and interval accommodate local geography.
- Submitting the proposed addresses to the local branch of the USPS for reconciliation with existing addresses and review.
- Providing an old to new address conversion lists.
- Providing a road name and number range list.
- Providing a detailed map of new address locations.
- Reviewing and approving address maps.

- Designating an Addressing Officer to maintain accurate address information once created[102]

In order to assign addresses to each residence, a map grid origin was set that allowed capture of all roads and residences in the area being converted. In some cases, the local authority set the origin to the southwest of the county. In other cases, the origin was set at the center of the town or another location that has historically been considered as the center of the town for addressing purposes. Many localities adopted flexible rules and it is not unusual to find multiple grids within an area (even within a city), especially if different areas had some residences with city-style addresses prior to the conversion. In addition, different grids are often found in adjacent areas that are served by different PSAPs or E911 service centers.

In some conversions, street or roads in the area were demarcated using an interval that may run unchanged for the length of the thoroughfare. A common interval used in E911 conversion is to divide a mile into segments of approximately 52 feet for the assignment of an address whose numeric value is intended to indicate the distance of an address from the street's origin. NENA however recommends using a 5.28' increment that allows for 1,000 uniquely numbered addresses per mile. During its E911 conversion, the State of Maine indicated that intervals of over 50 feet should be used only in sparsely populated rural areas and that cities might be better served with 25 foot increments. Other communities simply did not follow the recommended standards.

Across the country, the increments, if used at all, vary and cannot be directly compared between systems without knowing the origin of the grid and the segmentation distances. To make matters somewhat more complicated, the interval may change along roads that cross between urban centers and rural, agricultural areas where residences are far apart. Conversely, some authorities worked with systems in which the address was simply the number of feet that the residence was located from the origin of the road. Note that there is little uniformity in terms of address numbering or address parity in E911 address conversion areas.

In more affluent counties, GPS surveys were used to establish street centerlines and a GPS coordinate was captured for each residence and the coordinate was usually represented on the street centerline. In these cases, the E911 address was assigned relative to the position of the GPS coordinates on the map grid and presented as a numeric address based on the interval set for the street. As a consequence, the address, in most interval based systems, serves as an approximate surrogate for a location. In a small number of areas, the E911 conversion was accomplished using a laser range finder to accurately locate the front door of each residence. In other cases, the GPS coordinate reflects the location of a driveway or, in some cases, the preferred entrance to the residence. Finally, in some situations the coordinate was captured from the presumed location of a residence on a USGS quadrangle and may not exhibit the locational accuracy of other methods.

The implementation of city style addresses does not indicate that Postal Rural Delivery, Rural Route and Highway Contract Delivery route addresses and box numbers have been eliminated in the converted areas. While some counties have opted to use the post-E911 address as mailing

[102] . The Addressing Officer is usually a local jurisdiction employee, but in some cases the local postmasters serve as the local Addressing Officer.

addresses, other communities, however, have allowed their residents' the option of deciding whether to apply for and adopt the E911 address as their mailing address. Several states, such as Maine, West Virginia, and Texas, standardized the E911 addressing process statewide and the addresses conversions from these states should be in a form that would be of significant benefit to the Census

4.2 How many conversions? How far to go?

Perhaps the most common source of E911 addresses that also serve as mailing addresses, is the LACS^{link} software (Locatable Address Conversion) introduced by the USPS in September, 2004 that was designed to provide a technology for address conversions other than those in response to a customer move. LACS^{link} allows business mailers to update their rural-style addresses with new, locatable city-style addresses in areas that are experiencing 9-1-1 emergency response address conversions. LACS^{link} is available as an option in CASS Certified Software (Coding Accuracy Support System) that is used by the USPS to evaluate the accuracy of address-matching software programs used by mass mailers.

The 2004 LACS^{link} replaced the original LACS program. The Locatable Address Conversion System (LACS) allowed mailers to update their databases when an address was converted to another form. The LACS database was implemented by the USPS in 1988 and the original impetus for this effort was to allow mass mailers to accommodate E911 conversions.

The LACS^{Link} database consists of several types of address conversions:

1) PO Box renumbering (initiated by USPS)

2) Secondary address renumbering (initiated by building owner)

3) Street style to street style (initiated by local planning authority)

4) Rural style (including HCR Addresses) to street style (initiated by local planning authority).

In the current version of the LACS^{link}, (November 2010) there are 4,125,151 rural style addresses that have been converted to city-style addresses out of a total of approximately 6,000,000 addresses in the LACS database.[103]

Addresses are never removed from the LACS database unless they are causing an operational issue that is usually a result of the address being reassigned by a local authority.

In the 2001 USPS Comprehensive Statement on Postal Operations, it is stated that there were approximately five million addresses in the LACS database. In the 2005 Comprehensive Statement on Postal Operations it was stated that the number of addresses in the LACS database (officially LACS$^{link)}$ by this date) the number was "more than 5 million". Presuming that these numbers are true would lead to the belief that the number of E911 conversions have plateaued. However, the fact that not all E911 addresses are used as mailing addresses confounds further

[103] Personal communication, Ruth Jones. Address Management. United States Postal Service.

analysis of the trend using postal data. Curiously, in Table 1 of this report, we indicated, based on USPS statistics, that there were approximately 42 million rural delivery points being serviced by the USPS and that these included the converted E911 addresses being used as mailing addresses. Rough subtraction would indicate that there are approximately 38 million rural delivery points that continue to use non-city-style addresses.

A study titled "*Geocoding rural addresses in a community contaminated by PFOA: a comparison of methods*" by Vieira et. al. indicated that LACS^{link} was an inferior substitute when compared to the Wood County, West Virginia official E911 data table (these tables are often referred to as a MSAG – Master Street Address Guide).[104] Although the sample sizes were small, the Wood County E911 data table was able to successfully geocode more street addresses and more rural route box numbers than using LACS^{link}. This outcome appears to be further evidence that E911 address conversions are not being used for postal addressing

Further research indicated that the LACS^{link} database may be incomplete, as its currentness depends on the local addressing authority providing the USPS the post-E911 conversion addresses and updates for new addresses on a timely basis and, then, the postmaster forwarding these data to regional Address Management Systems (AMS) that process and forward them to the National Center. However, our conversations with companies in the E911 conversion industry indicated that in their experience in numerous communities, the residents are allowed to opt out of using the E911 address as a mailing address if the local Post Office agrees. When this option is elected, the post E911 address will not appear in the LACS^{link} database or the DSF that is used by the Census Bureau.

It appears that the most comprehensive sources of post E911 address conversion data are available from the jurisdictional authorities that managed the address conversion process. In some states the conversion process was managed by a statewide authority, while in others it was the result of a county-based process. In many cases, the E911address conversion data (usually in the form of a data table or a Master Street Address Guide (MSAG)) has been merged with GIS systems to leverage the data quality associated with the new addresses and may be more accessible to the Census Bureau because of this transition. Unfortunately, a number of local communities partnered with commercial GIS and 9-1-1 conversion firms and in a minority of these localities, the conversion data is privately owned and not available to the public without the payment of a fee (see Appendix 3 for concise information on the MSAG and related terms).

It remains unclear what percent of the United States has successfully completed the E911 address conversion process, although all indications are that E911 conversion has been completed by the majority of counties across the United States. During the course of our research, we spoke to numerous businesses that provide E911 address conversion and mapping services to communities across the United States and were told that this segment of their business slowed several years ago and that RFPs requesting such services were now issued very infrequently. Officials at the FCC (the lead agency for Phase 2 of 911 (Cellular) and the agency concerned

[104]Veronica M Vieira, Gregory J Howard, Lisa G Gallagher, and Tony Fletcher, Geocoding rural addresses in a community contaminated by PFOA: a comparison of methods. Environmental Health 2010, 9:18doi:10.1186/1476-069X-9-18 http://www.ehjournal.net/content/9/1/18. PFOA is Perfluorooctanoic Acid, an agent that persists indefinitely in the environment. It is toxic to and a carcinogen in animals, as well as dangerous to people.

with the responsibility for ensuring that 911 systems work during emergencies) indicated that the agency did not have statistics on the coverage or percent of addresses that had been converted to E911). Similarly, the same questions remained unanswered when posed, to the staff of the RIBBS National Customer Support Center (Rapid Information Bulletin Board System) of the USPS.

Perhaps the best source of information is the National Emergency Number Association (NENA) which indicates that: "At the end of the 20th century, nearly 93% of the population of the United States was covered by some type of 911 services. Ninety-five percent of that coverage was Enhanced 9-1-1 service. Approximately 96% of the United States is covered by some form of 9-1-1." [105]

NENA also provides a map of the United States E911 deployment[106] shown in Figure 39, below. On the map the following legend applies: White =No data available, Red = No 9-1-1/ or Basic, Yellow = Enhanced 9-1-1(requires address conversion to city style addresses), Light Blue = Started Phase I Wireless (cellular telephone numbers are provided to emergency services during the call), Dark Blue = Completed Phase I, Light Green = Started Phase II (both the cellular telephone number and location of cell phone are automatically reported to emergency services during a call). Dark Green = Completed Phase II

Note that counties that have begun to deploy or have completed wireless Phase 1 or Phase 2 emergency services should have completed Enhanced 9-1-1[107] services and converted addresses to a city-style E911 addressing system, as the infrastructure to support Phase 1 and Phase 2 wireless is based on the conversion to E911.Areas shown in red on the map are the locations where E911 address conversion has not yet occurred (and in some cases where basic 911 services are not available).

Our research with companies in the business of providing assistance with E911, confirmed the NENA map, as these companies indicated that their conversion business now is infrequent, although not all communities use a commercial firm to manage the conversion process.

4.3 Summary

E911 address conversions should prove to be of benefit to the Geography Division for purposes of expanding and augmenting addresses and, possibly, address descriptions in the MAF. Address conversions that are used for postal addresses will eventually be available from the USPS through the database that supports the LACS[link] addressing system, as well as from the DSF. Since addresses converted for E911 purposes in some communities do not necessarily need to be used as mailing addresses, it is likely that the Geography Division would benefit from contacting the local or state authorities responsible for managing the E911 conversion process for access to the MSAG or local variant of this database.

---

[105] See the NENA website at http://www.nena.org/911-overview-facts.

[106] See the NENA website at http://nena.ddti.net/Documents/NENA%20Wireless%20E911%20deployment%20map.pdf

[107] Phase 1 and Phase 2 are 911 services focused on receiving the cellular caller's telephone number (Phase 1) and eventually their location information from either cell towers or cell phone-based GPS readings. As noted in the text, all counties that have Phase 1 or Phase 2 services already have E911 services.

Several states, such as Maine, West Virginia and Texas, standardized the E911 addressing process statewide and the addresses conversions from these states should be in a form that would be of significant benefit to the Census.

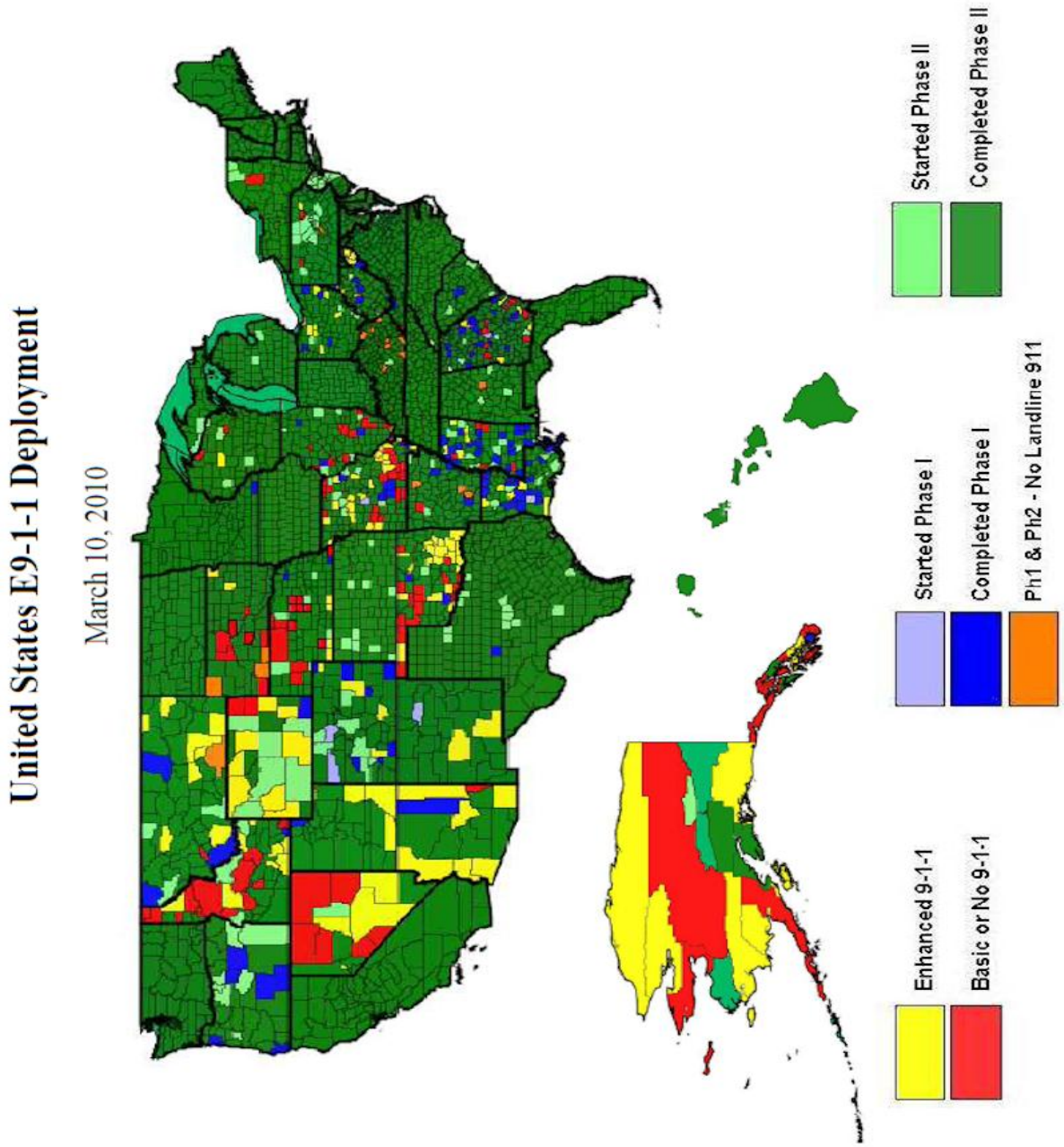


Figure 39. Status of E911 Deployment in the United States.

Sub-Task 5: Identify source of addresses, commercial and local

The Census Bureau spends considerable time, effort, resources, and expense maintaining the MAF to support the next census. While the MAF benefits from a number of internal and externally focused efforts designed to improve the quality and comprehensiveness of its address list, we researched potential providers of addresses that could prove to be significant sources of quality address information for the Census Bureau. Our purpose here is not to be exhaustive, but to focus on the types of address sources that could be of benefit to the census for updating and expanding the MAF.

We note here that it is possible that the Title 13 restrictions that hinder the Census Bureau from exchanging or publishing address information might actually be a benefit in terms of attempting to license commercial address data. If the data that would be provided by vendors for use in the MAF could not be publicly distributed, it would be possible for these vendors to maintain the competitive advantages of their address data while benefitting the Census Bureau in the process.

Our survey of the commercial providers included direct mailers/national mailers, package delivery companies, parcel and land development companies, navigation database companies, white pages companies and public service companies. Our survey of local sources of address information was focused on government offices and agencies that were tasked with collecting accurate address information for purposes of local governance, although we, also, examine regional and statewide agencies tasked with address management.

Based on our examination of commercial sources of addresses, we have concluded that the majority of these companies rely solely on the USPS as the source of their address data, that is, they rely on the DSF[2]. While some companies claim to collect additional data that may make their address list more robust, it is unlikely that the majority of the companies in the classes of companies surveyed have access to address information that is superior to DSF that is used to update the MAF twice a year

5.1 Direct Marketers/Mass Mailers

Direct marketers and mass mailers prepare address lists that can be used by their customers for purpose of marketing goods and services. Mass mailers usually practice advanced address and addressing hygiene in order to qualify for the lowest U.S. Postal Service mailing rates, as well as to increase the efficiency and targetability of their lists over their competitors. It is possible that some of these “scrubbed” address lists might be of benefit to the Census Bureau in its goal of improving the quality and comprehensiveness of addresses in the MAF.

The fundamental weakness of mass mailers and direct marketers as sources of address data is that these companies are concerned only with mail-related use of addresses and do not require location information beyond that provided by the address. The fact that they do not know the specific street address or the precise location of a Rural Route Box or a HCR box does not significantly impact their business, even though the majority of these companies use LACS^{link} to help them convert rural address points to city-style address, in order to help sharpen the focus of direct marketing campaigns. Since the aim of the GSS initiative is to support continuous

updating of the Bureau's address inventory in areas not covered by the DSF, it is unlikely that working with commercial mailers and direct mailing enterprises would be of significant advantage in pursuing this goal.

However, these professional mailing houses have developed an understanding of the DSF2 (Delivery Sequence File –Second Generation[108])and associated software for address hygiene that might be of interest to the Census in enhancing their processing of the DSF or perhaps by importing the monthly DSF2 data after it has been processed by these companies. Indeed, these companies must be CASS certified in order to achieve discount rates from the USPS.[109] If the mass marketers fail the Cass certification their postage costs increase creating a price disadvantage compared to their competitors. Most mass mailers update their mailing lists on a monthly basis, which could make their list more current in respect to address changes than the MAF.

Although the MAF is updated only twice a year, it is updated from a source that it thought to be superior to the DSF2. We were not able to determine the specific differences between the address information in the DSF (available only to the Census) and the publicly available DSF2. We were told that updating the MAF using the DSF takes four months to complete. We note that the Geography Division should ensure that the value added by the DSF cannot be produced by the DSF2, which commercial firms use to update their address lists on a monthly basis.
Listed below are four commercial mailers that provide address data. The first two, Valassis and Harte Hanks, offer a variety of additional services for purposes of direct marketing. The second two are more traditional address lists providers that also offer other services.

Valassis (formerly ADVO and known for its Red Plum branded advertising) is regarded by many as having the most comprehensive address list in the direct marketing industry and the company claims to reach virtually all households in the United States.[110] Valassis uses proprietary address systems to monitor the data they ingest from the USPS (primarily the DSF^2) and provides targeted mailing lists that are CASS-certified (Coding Accuracy Support System).[111]

Harte Hanks, known for its Penny Saver advertising publications, was founded in 1923 and has grown into a direct marketing company of worldwide scope.[112] In addition to its integrated multi-channel direct marketing solutions, Harte Hanks offers Advanced Data Quality software (ADQ) that provides CASS processed addresses enhanced with delivery points (ZIP + 4, carrier route codes), Address Change Services (ACS) and Address Element Correction (AEC and AEC^2).[113] See the Harte Hanks Advanced Data Quality Dictionary for more details.[114]

---

[108] The version of the file available to the public from the USPS is the DSF2 –Delivery Sequence File Second generation – see http://www.usps.com/ncsc/addressmgmt/dsf2.htm
[109] USPS. Cass Technical Guide 2009-2010 Cycle. United States Postal Service. Washington D.C. 2009.
[110] Jennifer M. Staab and Vincent G. Iannacchione. Evaluating the use of residential mailing addresses in a national household survey. Joint Statistical Meeting. 2003.
[111]See http://www.valassis.com/1024/Home/10home.aspx and http://www.valassis.com/1024/Services/DMplanning.aspx .
[112] For more information about Harte Hanks see http://www.harte-hanks.com/
[113] Except for the ADQ software, the other modules mentioned here are USPS software provided to CASS certified mailers.

InfoUSA was established in 1972 and focused initially on supplying a national database of business listings. Since then, the company has expanded into compiling additional databases including one that contains data on 210 million consumers and their mailing addresses, as well as14 million business listings including addresses. The company indicates that its consumer data is continually updated and processed against both the USPS National Change of Address (NCOA) and DSF, producing a database that is 93% deliverable. The information is also carrier route and ZIP+4 coded using CASS.

The company's 14 million business listings are one-hundred percent phone verified based on the company's staff placing over 20 million calls each year to verify the accuracy of their business listings data. As a point of reference, Google recently re-signed with InfoUSA allowing InfoUSA to continue as their major supplier of business addresses for the Google Local Search index. Google was considering replacing InfoUSA with a combination of its own efforts and crowdsourcing. However, after visiting the infoUSA facility in Omaha, Nebraska and witnessing the work of 600 staff actively calling business locations to authenticate listing information Google, decided that infoUSA was reasonably authoritative.[115]

Melissa Data, founded in 1985, provides U.S. and Canadian address, phone and email verification solutions, mailing lists, and data hygiene and enhancement services. The company bills itself as a developer of high-performance contract data verification and mail management solutions. Melissa Data has a number of software packages for address list development and maintenance. Its Postal Services Enhancements included NCOA[link,] DSF, UAA processing and CASS.[116] The company appears to have a number of unique software products that might have some interest to the Geography Division.

5.2 Package Delivery Companies

The Geography Division has expressed interest in how FedEx and UPS acquire their address information. Both companies use the DSF[2], as well as CASS, NCOA, and other software services provided by the USPS. In addition, both companies use navigation databases from NAVTEQ in their package delivery and logistics support businesses.

FedEx and UPS have invested substantial sums in operations research and logistics software used to analyze the parcel flow, delivery time requirements, and geographic location requirements to compute desired routes for their drivers. These routes, based on the optimized package delivery sequence, are broadcast to the handhelds used by the drivers of the companies' delivery vehicles.

---

[114] Harte Hanks, Advanced Data Quality Dictionary, Harte Hanks. Undated. See https://www.adqdirect.com/HDS.SSP.WebFiles/Documents/Data%20Dictionary/ADQ_Direct_Data_Dictionary_Standard.pdf

[115] For additional information on infoUSA see http://www.infousa.com/Home/home/83552

[116] For more information on Melissa Data see http://www.melissadata.com/

In some cases, the map or address files that they use are erroneous or not up-to-date. In these cases the address database that each company creates is supplemented through the electronic pads that every driver carries throughout their route. These devices are capable of recording GPS coordinates at any delivery point including addresses that are not on their maps or included in their address databases. In addition, these devices send the location coordinates of the driver to the logistics center where they are used to indicate, in near real-time, the location of the vehicle for purposes of forecasting the time of delivery and scheduling package pickups. Further, using a larger, purpose built electronic slate, drivers can sketch in streets, new developments, and other data that are not in the databases and enter them along with relevant postal addresses and the location coordinates of the delivery point.[117] Entry of the data into these slates automatically synchronizes it with the master database used in logistics planning.

Rob Carter, CIO of FedEx, indicated that their corporate address database is a prime corporate asset that they do not willingly discuss in public. However, on further probing, it was revealed that in addition to standard address data, the FedEx database includes preferred entrance for delivery, preferred time for delivery, alternative delivery doors and specific information about each location to which parcels are delivered.[118] Carter indicated that keeping their address database up-to-date was one of FedEx's most vexing problems.

It is likely that both UPS and FedEx have the best available databases of E911 conversions, since one of the many benefits of converting from a rural delivery point address to a city-style address is the ability to receive packages at a residence. However, since many rural residents have chosen not to use the E911 address for mail delivery, the city-style address conversion will not show up in the LACS^{link} database. UPS and FedEx drivers have been capturing the actual addresses of rural residences based on the signage that must be posted at a house that has been provided an E911 address, whether or not the address is used for mail delivery. The unfortunate issue is that neither FedEx nor UPS offer their corporate databases as part of their logistics offerings.

FedEx offer a publicly available tool (you must login) available at MyFedEx called Address Checker that provides street level address matches in the United States and Puerto Rico. Its database is updated on a monthly basis from the DSF^2. The FedEx Address Checker includes these functionalities: checks to see if the street exists in the city and state, or zip code entered, checks to see if the street number is within a valid range for the street entered, provides possible alternatives when an exact match cannot be found based upon the street number, city, state, and zip code entered, and checks to see if the address is commercial or residential, if the address is found in the database.

While both companies provide logistics services, they do not make their corporate address databases available through these businesses. Instead, both companies are relicensors of the NAVTEQ navigation database, which is incorporated into their logistics offerings. The logistics

[117] See Episode 2 of the Geospatial Revolution Project for a video showing some of the addressing technology used by UPS. http://geospatialrevolution.psu.edu/episode2/complete (advance time to 5:50)
[118] Personal conversation with Rob Carter, Executive Vice President of FedEx Information Services & CIO of FedEx Corporation. October 2010

services are focused on optimized route planning, real-time wireless dispatch, strategic territory planning and vehicle telematics to ensure driver and citizen safety.

5.3 Parcel Mapping Companies

An emerging and potentially fertile area of address information is from firms, such as CoreLogic, that create parcel (property) databases intended to cover the United States. These databases include parcel identification numbers, parcel/ lot dimension, street addresses and street names, as well as other data typically used in property location and identification. CoreLogic describes their ParcelPoint product as a database comprised of more than 124 million parcels that cover over 2,200 counties and townships, representing approximately 90 percent of the population of the United States and 86 percent of the total land base.

During a conversation with representatives of CoreLogic at the 2010 ESRI USER Conference, it was revealed that CoreLogic has a network of stringers (sometimes employees) that frequent county courthouses throughout the United States to gather property information, as well as maps and any imagery that are publicly available. This field work is the source of the Parcel ID as well as the address and ownership information that help define the property boundaries outlined through the use of aerial imagery. CoreLogic provides the ensemble of data collected in a variety of products in addition to Parcel Point.[119]

MetroScan® provides a comprehensive database of residential, commercial, industrial and vacant property. CoreLogic indicates that this online solution is updated through the collection and management of raw real estate related data and imagery from more than 3,600 government and proprietary data sources.[120] The company indicates that their county databases have 65 different fields to search, including property ownership and characteristics (including absentee owners), current and past sales figures, telephone numbers, parcel and street maps, census tract data, assessed value, transfer dates, land use and other property details.[121]

Census personnel may also be interested in examining the CoreLogic REiSource product that provides title companies with marketing and farming tools via access to homeowner data in more than 3,050 counties. The database includes assessor maps and recorded document images that define the property's location and history.[122]

Finally, the CoreLogic RealQuest® service provides what the company calls the undisputed source for locating real estate information across 3,085 counties on 145 million properties (including residential, commercial, industrial, agricultural and vacant land) representing 97 percent of all real estate transactions in the United States. The database is updated daily and is gathered from more than 10,000 government and proprietary sources. What is interesting about this database is that in includes:

---

[119] For information on ParcelPoint see http://www.corelogic.com/Products/ParcelPoint-Technology.aspx
[120] For information on MetroScan see http://www.corelogic.com/Products/MetroScan-Online.aspx
[121] The user manual for Metroscan can be found at http://www.corelogic.com/uploadedFiles/Pages/Products/20100302%20Metroscan%20Manualsm.pdf
[122] For information on REiSource see http://www.corelogic.com/Products/REiSource.aspx

> "… property records, tax assessments, property characteristics and parcel maps from tax assessors and county recorders offices across the nation. This information is combined with flood, demographics, crime, maps and *property visuals*, document images, trending data and other information from proprietary sources provides a complete picture."[123]

We suspect that CoreLogic, through its various products may have data on the location of rural living quarters that could be of significant use to the Geography Division in finding the residences associated with rural delivery points, as well as those locations that do not receive mail service.

Datamap is a small business that provides exclusive data on all new residential developments across North America.[124] The company claims that its DataMapGeo database provides the best solution for finding the location of new residential developments. The streets in the database, boast sub-meter accuracy and additional data includes: name of development, city, state, ZIP Code, county, point addressing, Census Block & Tract, county parcel PIN and/or block & lot, parcel area, perimeter and acreage, and centroid points of parcels or rooftop geocoding. In addition, and perhaps of interest to other Divisions of the Census Bureau, the company maintains metadata on the developer, development, and community including the home type, average square feet, and average sales prices among other information.[125]

The data provided by the company is obtained through what they describe as "exclusive" relationships with builders, as well as non-exclusive relationships with field crews and other data aggregators. The company indicates that its data is for developments that are in the "move in" stage, indicating that the street name and address information should be complete and authoritative.

5.4 Automobile Navigation/GPS databases

The developers of navigation and GIS databases such as NAVTEQ, Tele Atlas and Google, represent a sizeable segment of the addressing market. These companies require accurate addresses to be able to provide navigation services to their customers. In addition, these companies are now advantaging their address quality by embracing crowd sourcing and allowing their user populations to provide address update information, when their maps are incorrect, out of date, or otherwise need augmentation. All three companies have developed their own proprietary database covering streets and addresses in the United States. In turn, these navigation databases are used by audiences that have a stake in helping the companies keep the map and address data up to date.

It is our understanding that NAVTEQ, Tele Atlas and Google update their addresses with the DSF^2, CASS, LACS^{link} and other products available from the USPS. In addition, NAVTEQ

[123] For information on RealQuest see http://www.corelogic.com/Products/RealQuest.aspx
[124] For more information on DataMap, see http://www.datamapintel.com/
[125] More information and sample data can be found at http://www.datamapintel.com/streets.asp .

licenses ParcelPoint® data and the geocoding product PxPoint from CoreLogic.[126] It is unlikely that CoreLogic provides data to Google, as the mapping products of CoreLogic are based on Bing Maps from Microsoft (which are based on NAVTEQ data). However, neither Google nor NAVTEQ currently licenses the address information in their databases.

Tele Atlas, now a division of TomTom, does license its address data through the company's Address Points product. However, the database currently is sized at only 80 million address points *globally*. We had presumed that the passive crowdsourcing utilized by TomTom/Tele Atlas, in which consumers agree to provide the GPS traces for their device to TomTom, would allow the company to develop accurate and potentially comprehensive positional coordinates for addresses. To date TomTom/Tele Atlas has received over 80 million updates from users including over two trillion GPS points, which would lead one to conclude that some of these points must be where trips start (at an address) or at trip ends (destination addresses). However, for purposes of privacy, TomTom peels off the first two minutes and last two minutes of coordinates contained in the GPS paths provided to them. As a consequence, the company captures, but does not keep, the data associated with the GPS coordinates of the start and end addresses of these potentially valuable tracks.[127]

We regard navigation database companies as one of the weaker commercial sources of address information. However, NAVTEQ, in conjunction with Bing, has fielded vans equipped with LiDAR, Inertial Measurement Units, GPS, and hemispherical imaging to capture information that would allow Bing and NAVTEQ to provide a product to compete with Google's Street View. We believe these technologies bear watching as potential sources of address information in the areas they cover. Unfortunately, the vehicles from Google and NAVTEQ are focused on collecting data in urban commercial centers and not in rural areas.

5.5 White Pages Publishers

White pages publishers are most often local telephone companies who create these directories of their clients' addresses and phone numbers. The white pages, while familiar to a large audience, are declining in importance due to the availability of address information on the Internet.

Telephone service in the United States changed dramatically in 1974 with the Bell System Divestitures which mandated the breakup of AT&T. The nationwide holdings of AT&T, known as the Baby Bells, such as Ameritech, Bell Atlantic, Bell South, etc., became the regional telephone operating companies and began to produce their own white and yellow page publications to carry on business after the breakup. In turn, independent local exchange carriers began to provide service in areas not favored by the larger Baby Bells. This proliferation of white page publishers served the needs of the phone companies and their customers, but with the rise of the Internet, there was a need for a comprehensive national database of phone numbers

[126] See NAVTEQ's website for details on this collaboration http://www.nn4d.com/site/global/market/connections/enterprise_americas/customer_collaboration/corelogic/p_corelogic.jsp

[127] Personal communication with Christopher Wilson, Director of ADAS, TomTom/ Tele Atlas. 2010.

and addresses. Unfortunately, there was no central source for these data except for local telephone directories. In turn, this led to "scraping", which is the practice of scanning local telephone books, applying OCR technology, and using the data to create regional or national digital telephone directories.

The problem with this approach, however, was that the phone books were regarded by the local phone companies as their intellectual property and protected by copyright. This issue was resolved by the Supreme Court in the landmark 1991 case of *Feist Publications v. the Rural Telephone Service of Lenora, Kanas*. In the case, the Supreme Court, in an opinion by Justice O'Connor, ruled that names, telephone numbers and addresses were observable facts and not capable of being protected by copyright. The court continued that at least some modicum of selection, classification and arrangement were required to qualify for copyright protection and that a phone book did not rise to the level of these requirements.[128] Further, the court advised all concerned that the copyright protection that could be assigned to a compendium of factual data applied only to the organization of the data (to the classification), but not the data if the data were of a factual nature

With this decision by the Supreme Court, numerous publishers began agglomerating the local listings from telephone directories in an attempt to create directories of names, addresses and phone numbers that were national in scope. In fact, today's common practice of screen scraping by using web crawlers to discover factual information from pages displayed on the Internet was one of many outcomes prompted by the Feist Decision.

The reason why we have included White Pages directories in our list of address sources is that these directories often contain up-to-date information on the address and phone numbers of living quarters throughout the United States. Since these directories have no copyright protection, the Geography Division could consider using these publications as sources of information to augment the MAF. Alternatively, the Census Bureau consider using web spidering (web crawling) to discover address information on the web, a concept we will return to later in this document. However, if the Census Bureau did not feel like expending this effort or was deterred by the nature of the compilation process, it could license the same data from InfoUSA or other businesses (e.g. Acxiom) that routinely scan phone directories for purposes of mailing list generation.[129]

5.6 Utilities and infrastructure providers

Address data from utility companies, particularly electric suppliers, could be of significant benefit to the Geography Division since these suppliers, in the aggregate, provide a service that is delivered to the majority of the residences in the United States. In addition, most utilities are regulated in geographic units that are well-defined. In addition, the services are often delivered to individual apartments, offering the potential of locating within building addresses.

[128] See Feist Publications v. Rural Telephone Services Co http://www.oyez.org/cases/1990-1999/1990/1990_89_1909 .
[129] For more details, see the infoUSA video on data quality at http://www.viddler.com/explore/infousadotcom/videos/3/

We pursued the potential use of address information from public utility firms such as South Carolina Electric and Gas that provides gas and/or electric services for much of South Carolina. SCE&G has advanced GIS capabilities including the use of ArcGIS 10 Mobile for field data collection. The company constantly maintains its data and licenses parcel data from CoreLogic, liaises with local governments to acquire digital submissions for planned developments, and supplements these sources with field efforts and the use of maps and imagery. In addition to capturing address information, SCG&E attempts to capture a single geographic coordinate as part of a RFID-based automated Meter Reading System. The RFID chips are "self-aware" and can passively broadcast the volume of gas or electricity used, as well as the GPS coordinates and information about the hardware incorporated into the meter to which it is attached.[130]

Unfortunately, the SCG&E address data are encumbered by their privacy agreement with customers and cannot be shared. In addition, during our discussion, it became clear that their GIS database was considered a key corporate asset and one that they are determined to keep out of the hands of competitors or potential competitors. We heard the same sentiment during conversations with other utility companies, such as San Diego Gas and Electric (SDG&E),

We thought that the class of collaborative utility exchanges might be more amenable to sharing their data. We examined the Johnson County, Kansas Collaborative Utility Exchange, a coordinated data exchange system involving KCP&L, the Kansas Gas Service, Johnson County Stormwater Management, and the Johnson County Rural Water District No. 7 that is aimed at the efficiencies to be gained by sharing managed data. The program has led to one-stop shopping for most of the local infrastructure players and reduced the cost of data collection.[131] Unfortunately, the Collaborative Utility Exchange data cannot be distributed outside of the consortium.

The privacy rights of customers as well as other security concerns appear to preclude the sharing of infrastructure data (both streets and addresses) with third-parties. It may be that direct conversations between the Census Bureau and utilities would provoke a response different than the ones they were willing to discuss with consultants. In addition, it is possible that Title 13 limitations related to the Census Bureau sharing address information might provide utility companies an avenue that would protect their sharing of customer address information.

There appear to be few infrastructure service providers who have not embraced GIS and do not operate or are pursuing the development of an intelligent infrastructure. We urge the Census to consider pursing data exchange agreements with these promising sources of address data.

5.7 Local Sources of Address Information

Our research on address sources did not change our initial belief that the most useful sources for discovering address information, including location descriptions and alternatives to postal

[130] Conversation with Wayne Meyer, Manager of GIS. South Carolina Gas and Electric.
[131] Hanes Zacharias. County Manager, Johnson County, Kansas. Fostering collaboration with Spatial Data. 2010 See http://www.magicgis.org/magic/symposiums/2010/presentations/zacharias_ses.pdf

addresses, would be found at the local level. Although in some situations, the local data can be most easily harvested from regional or state authorities.

Our contacts with the National States Geographic Council revealed a 2009 NSGIC study that concluded that twenty-three states have a statewide or multi-jurisdictional address database. In addition, fourteen of these states indicated that their address database included individual addresses; while nine states indicated that their data were comprised of address ranges (some states collected both individual addresses and ranges - see Table 6). In many cases, the decision to provide statewide support for E911 has caused states to establish a coordination center for address and street centerline data.

State	Does your state have a statewide (or multijurisdictional Address database?	Is this database based on individual addresses or address ranges?
Arkansas	Yes	Both
Connecticut	Yes	Address ranges
Delaware	Yes	Both
District of Columbia	Yes	Both
Illinois	Yes	Address ranges
Indiana	Yes	Individual addresses
Kansas	Yes	Both
Kentucky	Yes	Both
Maine	Yes	Address ranges
Maryland	Yes	Both
Massachusetts	Yes	Address ranges
New Hampshire	Yes	Both
New Mexico	Yes	Address Ranges
New York	Yes	Both
North Carolina	Yes	Individual addresses
Ohio	Yes	Both
Oklahoma	Yes	Both
Oregon	Yes	Address ranges
South Carolina	Yes	Individual addresses
South Dakota	Yes	Address ranges
Tennessee	Yes	Both
Utah	Yes	Address ranges
Vermont	Yes	Individual addresses
Virginia	Yes	Address ranges

Table 6. Statewide address databases. (The data in this table were provided, in part, by William Burgess, Washington Liaison, National States Geographic Council.)

The National Association of Counties (NACo)[132] is another potential source of leads for information about addressing at the county level in states that do not have statewide address inventories. NACo provides an online list of elected county officials such as county clerks and assessors who might be in a position to provide information about county-wide GIS and other activities focused on addresses and addressing.

In addition to the need to provide E911 support, 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).[133] 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.[134]

The NTIA's State Broadband Data and Development Program is designed to help states gather data twice a year on the availability, speed, and location of broadband services. The data will be used by the NTIA to update a public, searchable, interactive national broadband map that is due to be completed in February 2011. The NTIA specifications require that broadband availability be aggregated to census blocks in areas where census blocks are less than 2 square miles in area and to street segments where census blocks are greater than 2 square miles in area.[135] Many of the coverage datasets are produced by the companies providing broadband services in areas across the country.

However, the following states have received awards from the NTIA that specifically included funds for the development of address files and mapping addresses in relation to broadband availability.

Arizona	Maine	New Hampshire	Utah
Arkansas	Massachusetts	North Carolina	Virginia
Colorado	Mississippi	Ohio	Virgin Islands
Indiana	Missouri	Rhode Island	Wisconsin
Louisiana	Montana	South Dakota	

Table 7. States that have carved-out NTIA funds for address file development

Each of the states or territories has taken a slightly different approach on the mapping that will result from the funds. New Hampshire is proposing to use the funds to develop the first publicly available master address file for the state. Other states are using the funds for E911 addressing

[132] See http://www.naco.org/Pages/default.aspx
[133] 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/ .
[134] See this page at Broadband.gov for more information http://www.broadband.gov/qualitytest/about/ .
[135] See http://www.broadbandmap.ny.gov/content/a-deeper-look.html

systems in counties that have not been able to perform the conversion due to lack of funds (e.g. Ohio and Arizona).

We recommend that the Geography Division follow-up on the NTIA program as it funds a new and potentially authoritative source of addresses, particularly rural address, that are of significant importance to the GSS initiative. See Appendix 4 on Address File Development/Mapping that provides an indication of each of these state's plans for the use of the funds for address file development.

Within states we found a variety of sources for addresses and GIS data. For example, the South Carolina GIS Council has made considerable progress in assembling address points from 27 of 46 counties with another four counties soon to join (Figure 40).

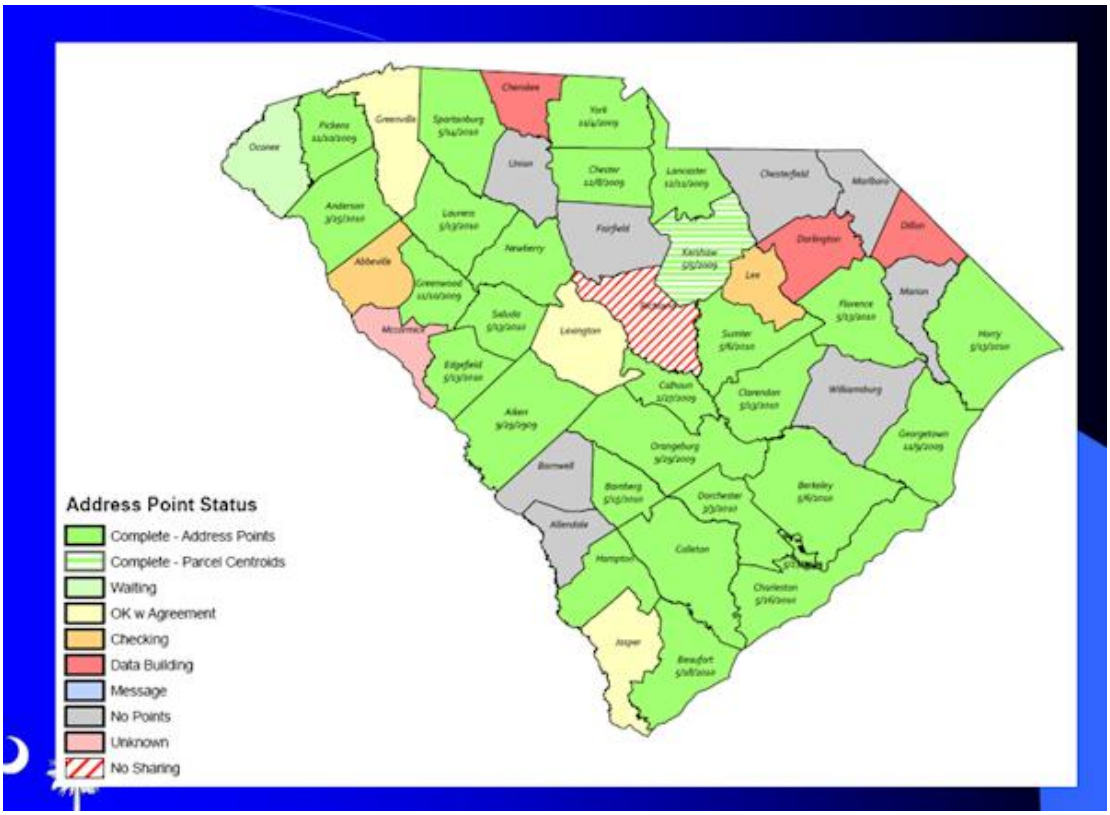


Figure 40. Status of address points acquired by the SC GIS Coordination Office (Image provided by Tim De Troye, GIS Coordinator).

While there are efforts to create county and state collections of existing address points, there are also a number of ongoing programs to monitor building permits. These programs would be particularly valuable during a targeted address canvass. Examples of these are provided by the Central Midlands SC Council of Governments (Figure 41) which aggregates the building permit data for a four county area and the Lexington County SC Planning Department (Table 8) that maintains an online **C**ounty **A**ddress **R**ange **D**irectory – "CARD" providing a simple listing of existing street names and address ranges.

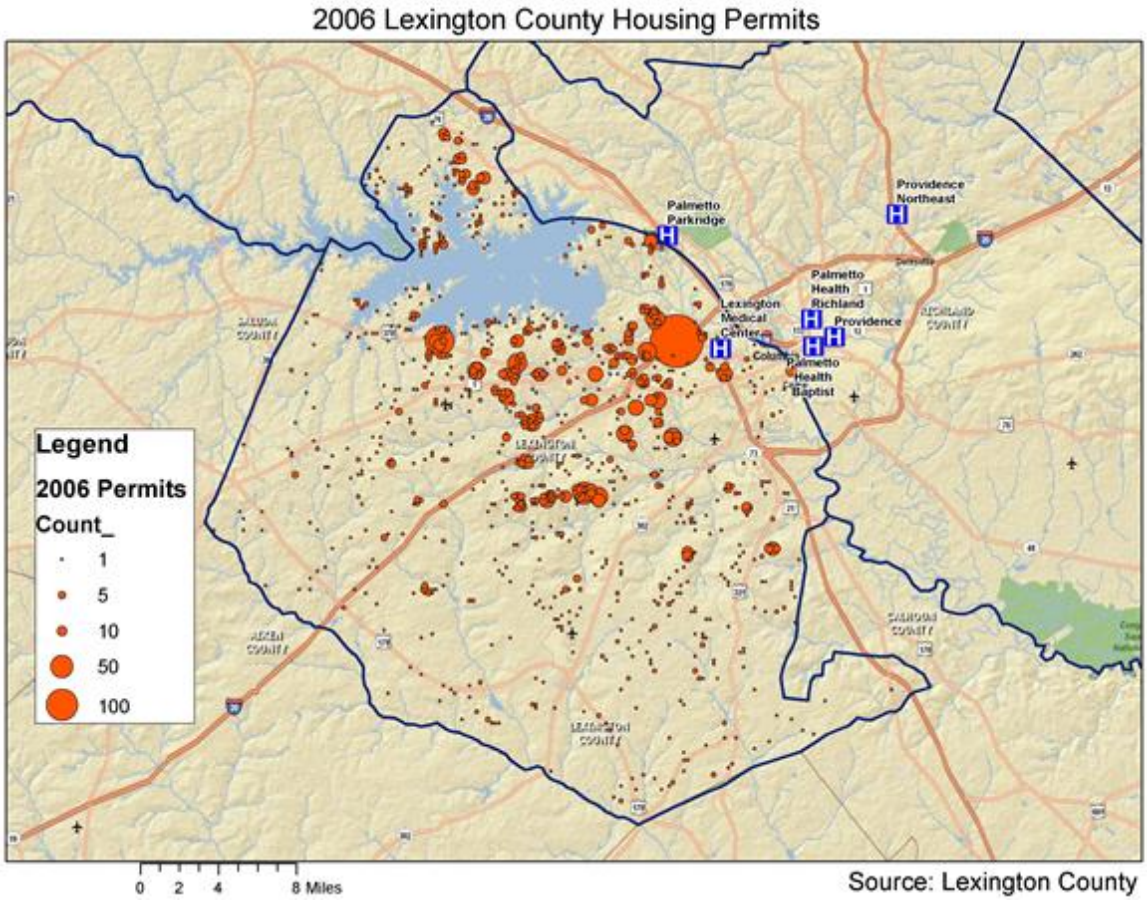


Figure 41. Map of 2006 building permits - generated from point level address file acquired from the Lexington County Property, Mapping and Data service

DIR	STREET NAME	LO#	HI#	ODD/ EVEN	POST OFFICE	ZIP CODE	MUNICIPALITY	SCHOOL DISTRICT	COUNTY COUNCIL DISTRICT
	1ST ST	1800	2015	BOTH	CAYCE	29033	CAYCE	2	9
E	1ST ST	100	499	BOTH	SWANSEA	29160	SWANSEA	4	1
E	1ST ST	500	598	EVEN	SWANSEA	29160	SWANSEA	4	1
W	1ST ST	100	399	BOTH	SWANSEA	29160	SWANSEA	4	1
	2ND AVE	100	199	BOTH	LEXINGTON	29072	LEXINGTON	1	3
	2ND AVE	1800	1899	BOTH	WEST COLUMBIA	29169	WEST COLA.	2	8
	2ND ST	1825	2017	BOTH	CAYCE	29033	CAYCE	2	9
E	2ND ST	100	199	BOTH	SWANSEA	29160	SWANSEA	4	1
W	2ND ST	100	999	BOTH	SWANSEA	29160	SWANSEA	4	1
	3RD AVE	100	399	BOTH	LEXINGTON	29072	LEXINGTON	1	3
	3RD ST	1800	1999	BOTH	CAYCE	29033	CAYCE	2	9
E	3RD ST	100	499	BOTH	SWANSEA	29160	SWANSEA	4	1
W	3RD ST	100	499	BOTH	SWANSEA	29160	SWANSEA	4	1
E	4TH ST	100	199	BOTH	SWANSEA	29160	SWANSEA	4	1
W	4TH ST	100	399	BOTH	SWANSEA	29160	SWANSEA	4	1
	5TH ST	500	599	BOTH	WEST COLUMBIA	29169	WEST COLA.	2	8
E	5TH ST	100	499	BOTH	SWANSEA	29160	SWANSEA	4	1

Table 8. A sample of the street names and address ranges from Lexington County, South Carolina Card Address Range Directory.

A compelling example of the availability of local address data from regional sources is MetroGIS, a regional geographic information systems initiative, founded in 1995, representing the seven-county Minneapolis-St. Paul, Minnesota metropolitan area.[136] The collaboration is a voluntary association of local and regional governments that involves partners in the state and federal government, as well as universities and other organizations both commercial and non-profit. MetroGIS has two databases that deal with address (parcels and Addressable Street Centerlines) and is considering developing a database of Addresses of Occupiable Units (although no progress has been made on this task). In addition, the consortium is developing a database comprised of address points (see Figures 42 and 43).

Figure 42. Address points showing specific store location (Provided by Randall L. Johnson, MetroGIS Staff Coordinator).

Figure 43. MetroGIS and reasons for collecting address points (Provided by Randall L. Johnson, MetroGIS Staff Coordinator).

[136] MetroGIS. Overview MetroGIS Spatial Data Infrastructure (SDI) – Creating Public Value, MetroGIS, .2009.

5.8 Local Governments

As mentioned above, we believe that local governments are usually the most authoritative, up-to-date and comprehensive sources of address information. For that reason, we provide a more in-depth examination the use of local sources of address information than provided for the other address source categories that we have described previously.

Local governmental entities that administer property records are a potential authoritative public source for address information. This is because of two conditions: it is typical practice to associate an address with a parcel (in addition to other identifying information); and most, if not all, states have laws that designate property records as public information. Further, an increasing number of county and city governments have digitized their property records and are making that digital information available to the public, freely through web interfaces, or by purchase.

We examined, Fairfax County Virginia, and Fredrick County Maryland as examples of the benefits and problems that might arise from using county governments as sources of address information.

Fairfax County Virginia (2009 est. population 1,037,605) maintains a large collection of over 30 layers of publicly available digital data for the county including the following items germane to the GSS initiative: addresses and address points, parcel boundaries, major transportation (edge of road pavement), minor transportation (edge of parking lot pavement), street centerlines, sidewalks, and building outlines.[137] These data can be purchased for $2850, or viewed online at: http://www.fairfaxcounty.gov/gis/VirtualFairfax/VirtualFairfax.aspx#appTop.

The information provided is quite detailed and accurate (with a reported positional accuracy of 2-3 feet for planimetric features), as can be seen in Figure 44, where individual townhouse units are easily identifiable. It is this level of detail that causes to recommend local sources as the best, least expensive, source of addresses, address points and location descriptions of the living quarters of interest to the Geography Division.

At first glance, these types of data would seem to be an excellent source for addresses and address location information about dwelling units/living quarters.

[137] Fairfax County, Virginia. See http://www.fairfaxcounty.gov/maps/county/countywide_prop_topo_2009.pdf

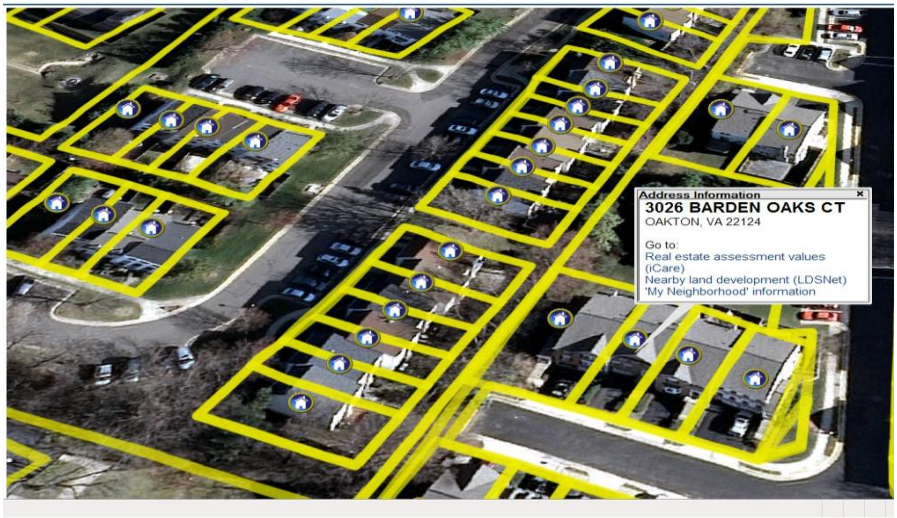


Figure 44. Addresses, address points, and parcels for a townhouse development in Fairfax County
Source: http://www.fairfaxcounty.gov/gis/VirtualFairfax/VirtualFairfax.aspx#appTop.

However, multi-unit developments, such as condominium complexes and apartment buildings present some interesting nuances with respect to addresses. First consider the condominium building shown in Figure 45. Since each unit is individually owned (and taxed) the county database has a record and address for each unit as seen in Figure 46.

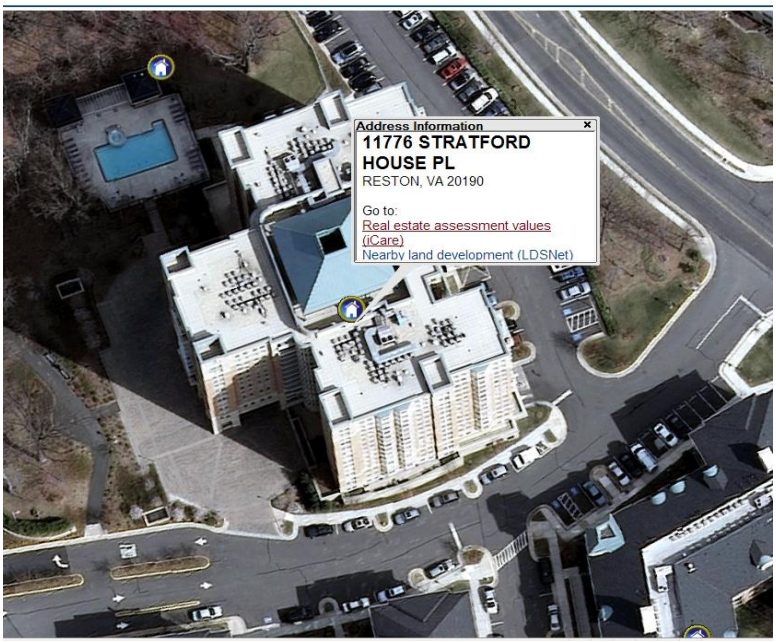


Figure 45. Multi-unit condominium building in Fairfax County. Source: http://www.fairfaxcounty.gov/gis/VirtualFairfax/VirtualFairfax.aspx#appTop

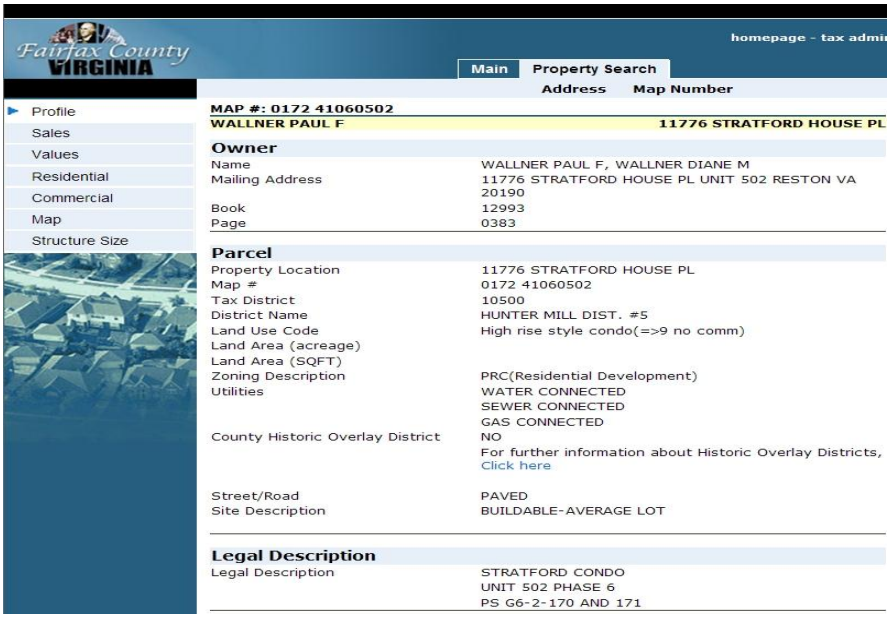


Figure 46. Address record for a unit within the multi-unit complex shown in Figure 2.
Source: http://icare.fairfaxcounty.gov/Main/Home.aspx

Now consider the case of an apartment building (Figure 47). Since it has one owner, there is only one property record (and address) in the database. The record does show that the apartment has multiple units (368), but does not supply addresses for each of these.

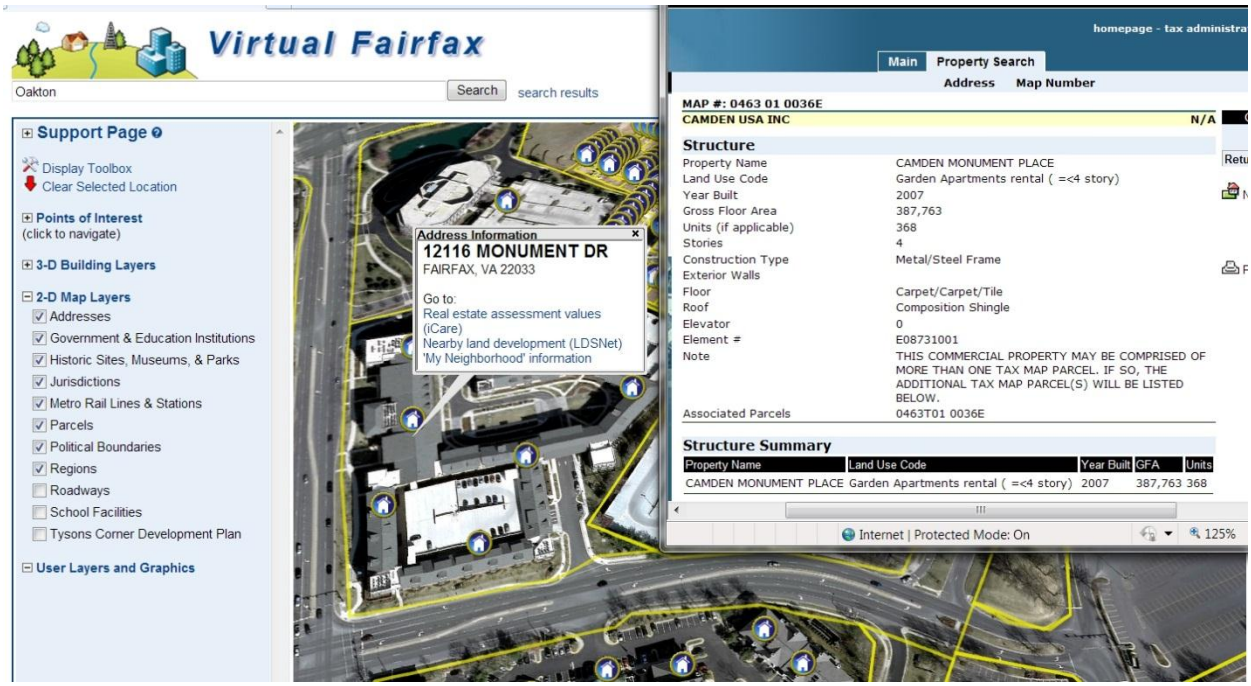


Figure 47. Address and property information for an apartment building in Fairfax County

If the Census Bureau requirement is to have an address for every dwelling unit, then the Fairfax County database may not be a sufficient source of information, particularly with respect to rental units. A blind adoption of the list from Fairfax County would result in significant errors of

omission. Otherwise, it would seem to be an authoritative source of data and could certainly be used, if not as a primary source for all addresses, as a source for quality evaluations of both the address list in the MAF and the location of the MSPs.

A similar situation with respect to data availability is found in Fredrick County Maryland (est. 2009 population 227,980). Over 70 types of digital data are available for the county; free for download or $200 for a CD. Interestingly, the Enterprise GIS unit of the Frederick County Government has the responsibility to provide property addressing assignment and road naming approval throughout the county. The Addressing and Road Naming Office of Frederick County assigns an address based on the county adopted addressing grid system, utilizing the county Geographic Information System (GIS). Following the assignment of addresses the address is reported to E911, other government agencies, utility companies, and the U.S. Postal Service. The address is recorded as a part of the data records of Frederick County and becomes the official address to identify your property (Source: http://ww3.frederickcountymd.gov/gis/addressing/home.html).

As was the case with Fairfax County, the information appears to be detailed and complete (Figure 48).

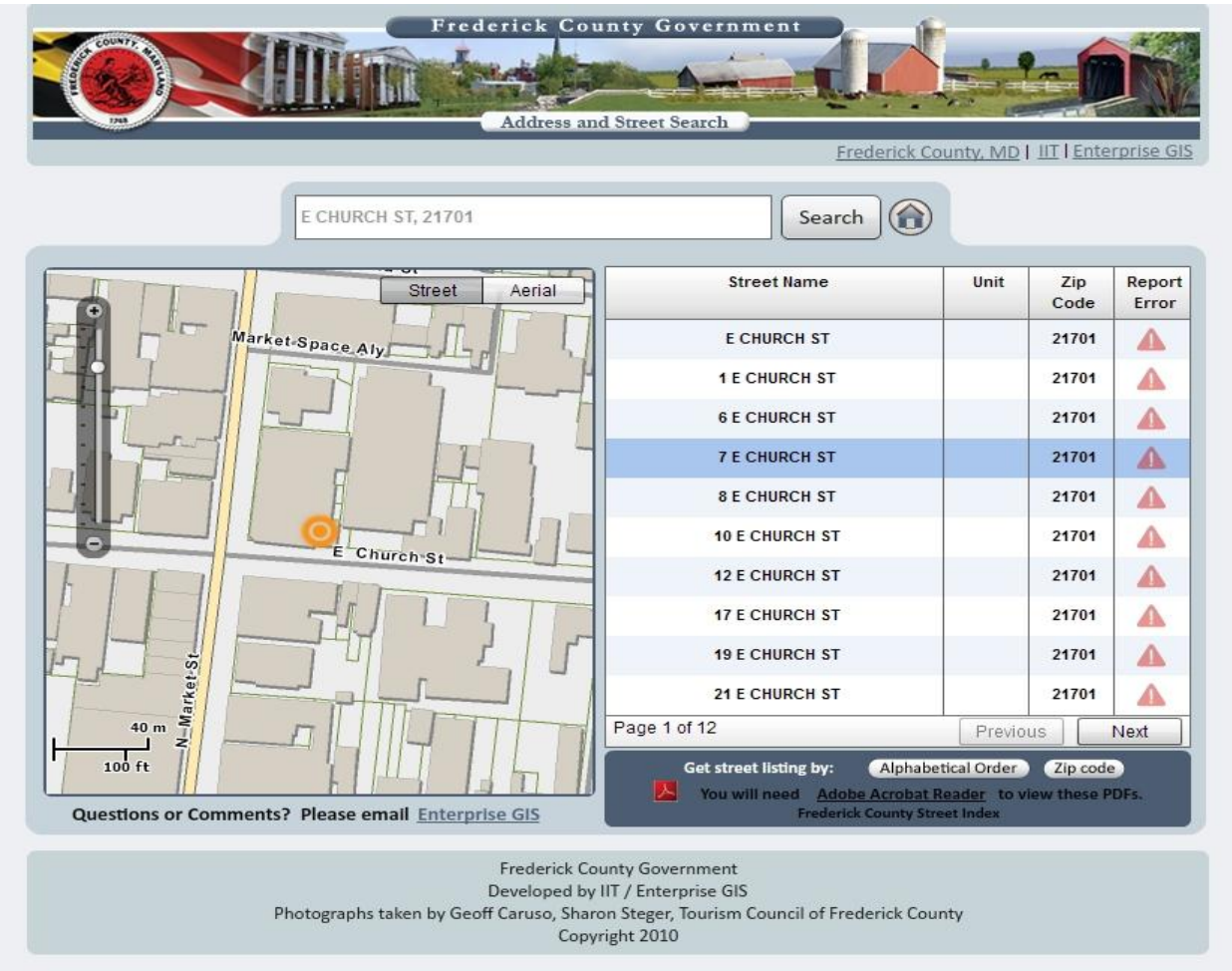


Figure 48. Example of addresses and address related data from Fredrick County Maryland
http://webmaps.frederickcountymd.gov/StreetSearch/#

As part of the public safety efforts (E911) it appears that all parcels have been assigned city style street addresses, regardless of whether any structure exists at that location. See Figure 49.

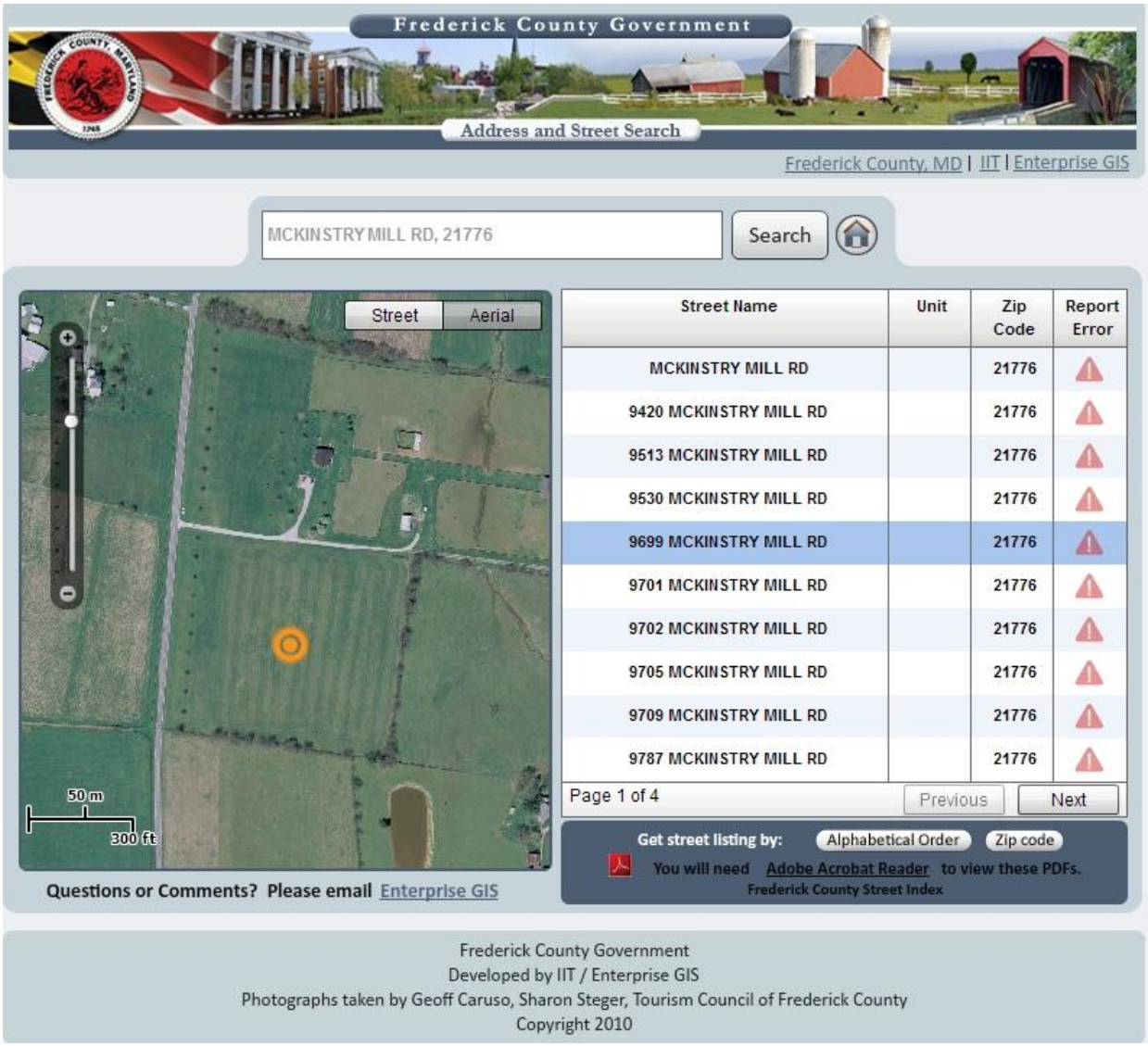


Figure 49. Address assigned to a vacant field in Fredrick County Maryland
http://webmaps.frederickcountymd.gov/StreetSearch/#

This inclusion of vacant land results in potentially significant errors of commission, were the Census Bureau to do a wholesale incorporation of the Fredrick County address list.

This brief survey suggests that local governments may have large amounts of detailed address and address point information available. However, the data that counties maintain for their property records are not a perfect fit to satisfy the needs of the Census Bureau. In order to incorporate these data into the MAF, one would have to understand the subtleties of the information being provided.

Finally, we urge the Geography Division to focus effort on the local sources of E911 address conversions. In each community an addressing authority was convened that guided the address conversion process. As noted previously, many residents whose addresses have been converted to a city-style address have opted not to receive mail at this new address, but to continue receiving mail at their non-city-style address. In these cases, the new E911 address will never appear in the LACS database nor will these addresses ever appear in the DSF. While these new E911 addresses could be discovered through field operations and roadside imagery, the very best sources of the new and converted addresses will be the local authorities that managed the process. Although it would require a difficult and time consuming effort to discover these local authorities, doing so could substantially improve the coverage in the MAF in areas not represented by the DSF.

5.9 Summary – Sources of Addresses

While the universe of address sources available to the Geography Division is rich and diverse, each source must be considered based on characteristics of the data, availability, cost, and benefits that the data may provide. A major issue that we have not discussed here is the ability of the Census to ingest and process these data in a manner that is efficient and beneficial in terms of some measure of improvement. In Table 9 (below) we have provided a concise review of the sources of address information that we have described, as well as a number of variables that may help describe these data and their possible use profile.

In Table 9 we used several terms that require definition. In the Accuracy column, we used the term "USPS certified" to indicate that the source was in conformance with the USPS requirements for mass mailing (e.g. CASS). In the same column the notation "USPS Plus" is used to indicate that the source likely has address information that the USPS does not have. For example, UPS and FedEx have captured information about the location of E911 converted addresses, as some residences provide these addresses to the package delivery companies to receive shipments of products that cannot be delivered to the mailing addresses. The use of the term "High" in the same column is based on our evaluation that local sources of addresses are usually the most authoritative sources of addresses, that is, the term is used here in a relative sense. In the column labeled Cost, our metric again is ordinal encompassing low, moderate and high ("high" was not used for any source referenced in the table). We did not have any useful method of estimating how the Census Bureau could or would source these data. For example, government agencies that charge commercial entities for their data often provide it free of charge to the Census Bureau and other government agencies. In addition, we unable to evaluate the negotiating positions of Census or the leverage it might bring to bear on any negotiation related to licensing or otherwise procuring data from public or commercial sources. As a consequence, we provided a relative measure of cost.

Our measure of "Difficulty of use" was, also, relative and based on our understanding of the data handling capabilities of the staff of the Geography Division. We do not feel that any of these data from data sources mentioned in the chart should prove to be difficult to use. Certainly, some of the data, such as parcel data, will require considerable thought on how to employ the data, but the actual implementations would not be, in our opinion, of significant difficulty.

We, also, used a relative scale to define the benefits of working with any of the data sources listed. Our low category reflects our opinion that working with these sources would be of limited benefit based on the reasons described in the text. High values reflect our belief that the data would help resolve addressing problems of interest to the Geography Division. Those sources marked with "Moderate" in general reflect sources with many participants and would require significant data gathering resources to collect and analyze the data, which, we suspect, would vary tremendously in quality.

Table 9. Alternative Sources of Address Information

Sources of Addresses	Postal Addresses	Address Points	Alternate Address Descriptions	Photography	Aerial Imagery	Accuracy	Comprehensiveness	Up-to-Date	Available to Census	Cost	Difficulty of Use	Benefits
Direct Marketer/Mass Mailers	Yes	No	No	No	No	USPS Certified	Best sources of postal address data	Monthly updates	Yes	Low	Low	Low
Package Delivery Companies	Yes	Yes	Yes	No	No	USPS Plus	Possibly most comprehensive sources for postal, address points and address points for non-postal locations	Monthly updates, but ability to update in near real-time if required.	No			
Parcel Mapping Companies	Yes	Yes	Yes	Mixed	Yes	USPS Plus	Available coverage is authoritative for addresses and address points, but coverage is incomplete on a national scale	Targeted updates, authoritative	Yes	Moderate	Low	High
Automobile Navigation Companies	Yes	No	No	Mixed	Yes	USPS Plus		Quarterly updates	Yes	Moderate	Low	Low
White Page Publishers	Yes	No	NO	No	No	High - limited to customer base	High - limited to customer base	Yearly or longer	Yes	Low	Moderate	Low
Utilities and Infrastructure Providers	Yes	Yes	Yes	Mixed	Yes	High - limited to customer base	High - limited to customer base	Updated immediately for new residences for change of customer	No			
Local Sources (GIS or E9-1-1)												
Statewide	Yes	Mixed	Mixed	Mixed	Mixed	High -	Mixed	Targeted updates for E9-1-1, longer term for GIS	Yes	Low	Low	Moderate
Regional	Yes	Mixed	Mixed	Mixed	Yes	High	High	Changes once notified by appropriate municipality.	Yes	Low	Low	Moderate
City/County	Yes	Mixed	Mixed	Mixed	Yes	High	High	Changes recorded daily, but data propagation may take longer	Yes	Low	Low	Moderate

6. Summary and Recommendations – Addresses and Addressing

The range of topics covered in this research on Addresses and Addressing has been broad, but each of the topics contains information on aspects of addresses and addressing of critical importance to the success of the Census Bureau's Geography Division. Pursuing the threads that we have discovered will take additional research and a determined strategy. However, the benefits of doing so may enhance the MAF to the benefit of the 2020 Census, as well as the ACS and other surveys using the address data in the MAF.

Rural address forms are of concern to the Census Bureau and capturing the location aspect of these residences through field operations remains difficult. Increased development of parcel mapping databases across the country may contribute to more accurate identification of rural and urban addresses and the ability to link them with a specific location. Next, advances in road-side imagery may provide a useful method of capturing enduring and unequivocal "Location Descriptions" of all address types. In addition, we think that the innovative use of crowdsourcing in the form of Volunteered Geographic Information (VGI) may eventually prove to be a new source of address information for the MAF. Local and commercial sources of addresses are another area of addressing we recommend that the Geography Division further examine as sources. Finally, we note that new and emerging address standards are factors that may have ramifications for the Geography Division in general and the MAF in particular.

Researching the diverse topics examined in this deliverable has led us to speculate about the appropriate address compilation model to support expanding, augmenting, and maintaining the MAF with quality information on addresses and addressing.

We recommend that the Geography Division consider developing a formal fusion-based compilation process that is influenced by address standards, such as Street Address Standard being adopted by the FGDC in January 2011. Field data collection of addresses should be are supplemented by the incorporation of road-side imagery and other forms of imaging. Additional address data could be gathered by purpose built web crawlers seeking either address information or surrogate information denoting spatial change that would help the Geography Division focus its address research on discovery of areas where addresses may have changed or new addresses may have come into existence.

Data partnerships, including commercial licenses and crowdsourced data, should become the primary source of addressing information, as they reflect and benefit from sources that are local, regional, and national. While address standards are overlooked by many, we believe that standards and their influence on data exchange are of primary importance in a fusion model such as shown below in Figure 50, since their widespread adoption would lead to greater uniformity in the address data harvested by address data collection methods.

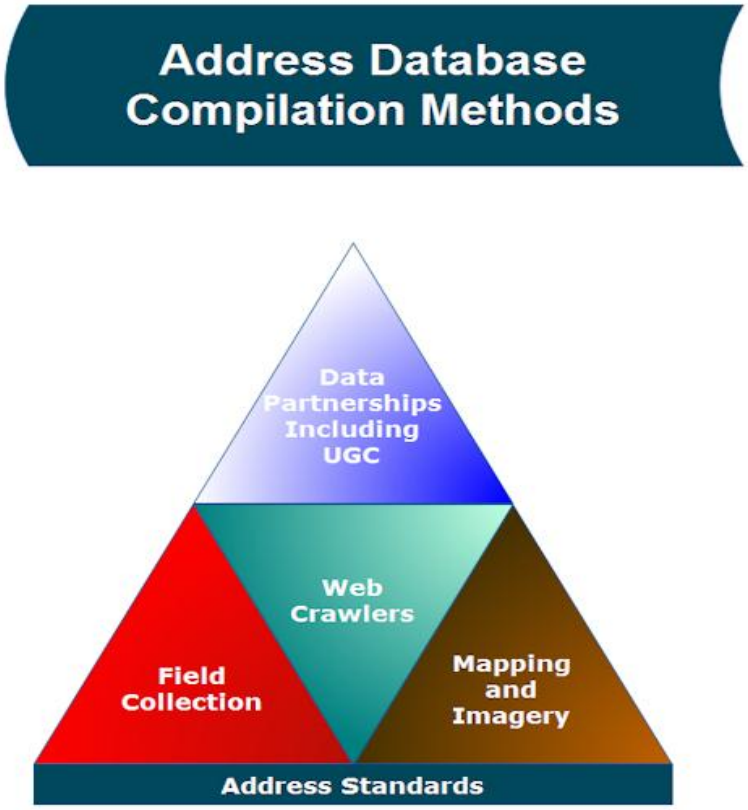


Figure 50. Address compilation Model showing the fusion of various types and sources of address data.

We view the fusion –based compilation model as a formalization of many of the data gathering and integration tasks that are currently conducted by the Geography Division and the Census Bureau. However, tightly integrating these efforts could produce a virtuous compilation update cycle that would improve the quality and comprehensiveness of the address data in the MAF.

7. References

Bernert, Joe. 2004. Oregon Geospatial Data Addressing Standard. See http://www.oregon.gov/DAS/EISPD/GEO/standards/docs/ORAddressContentStdv1.pdf?ga=t

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).

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

Clay County, Minnesota. 1998. Property Address Standard. See http://www.co.clay.mn.us/Depts/GIS/GISAddSt.pdf

Coetzee , Serena. 2010. The Addressing project by the International Organization for Standardization (ISO). 2010. PowerPoint presentation. See https://www.euradin.eu/Lists/Announcements/Attachments/32/3_1_EURADIN_ISOAddressing_SerenaCoetzee_15June2010.pdf

Cork, Daniel. L., Michael L. Cohen, and Benjamin F. King (Editors), 2004. Reengineering the 2010 Census: Risks and Challenges. Washington, DC: The National Research Council

Dobson, Michael W. 2010. MapQuest on Botox, Tele Atlas on Detox. Exploring Local http://blog.telemapics.com/?p=334#comments

EURADIN. 2010. EURADIN Project 2008 -2010, PowerPoint presentation. https://www.euradin.eu/Lists/Announcements/Attachments/32/1_1_EURADIN%20project.pdf

EURADIN. 2008. EURADIN European Address Infrastructure, PowerPoint presentation. https://www.euradin.eu/Documents/General%20project%20presentation%20oct08.pdf

Federal Geographic Data Committee (FGDC). 2010. United States Thoroughfare, Landmark, and Postal Address Data Standard (Draft). Reston, VA: FGDC. Available online at http://www.fgdc.gov.

Federal Geographic Data Committee. (FGDC) 2008. Geographic Information Framework Data Content Standard: FGDC-STD-014.7-2008. See http://www.fgdc.gov/standards/projects/FGDC-standards-projects/framework-data-standard/GI_FrameworkDataStandard_Part7_Transportation_Base.pdf

Federal Geographic date Committee (FGDC). 2010. Postal Addressing Profile of the United States Thoroughfare, Landmark and Postal Address Data Standard. Provisional Draft (under review by the USPS). (The Postal Addressing Profile is included in Committee draft at http://www.fgdc.gov/standards/projects/FGDC-standards-projects/street-address/?searchterm=address%20standard

Goodchild, Michael F. 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69, (2007), 211-21.

Goodchild, Michael F. 2010. Twenty years of progress: GIScience in 2010, Journal of Spatial Information Science, 1 pp. 3-20.

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

Haklay, Dr. Muki. 2008. Understanding the quality of user generated mapping, PowerPoint Presentation 2008 http://www.slideshare.net/mukih/osm-quality-assessment-2008-presentation?from=ss_embed

Hanks, Shawn (2010), Addresses 101: A Background Primer, PowerPoint Presentation, The GSS Initiative Offsite, January 19-21, 2010,

Harte Hanks. Undated. Advanced Data Quality Dictionary. See https://www.adqdirect.com/HDS.SSP.WebFiles/Documents/Data%20Dictionary/ADQ_Direct_Data_Dictionary_Standard.pdf

INSPIRE RDM. 2002. INSPIRE position paper: Reference Data. http://www.ec-gis.org/inspire/reports/position_papers/inspire_rdm_pp_v4_3_en.pdf

INSPIRE. 2008. Drafting Team "Data Specifications" – Definition of Annex Themes and Scope. http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.3_Definition_of_Annex_Themes_and_scope_v3.0.pd*f*

INSPIRE. 2009. INSPIRE Data Specification on Addresses. INSPIRE_DataSpecification_AD_v3.0. http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_AD_v3.0.pdf

Intelligent Addressing. 2010. Summary Product Description for the NLPG, 2010. See http://www.iahub.net/docs/1263829255172.pd*f*

IS0/TC211. 2010. Geographic Information/Geomatics, Draft Review summary of project 19160, Addressing. http://www.isotc211.org/address/docs/Review_summary_19160_20100622.pdf

Kosar, Kevin, 2008. The U.S. Postal Service's Use of Contractors to Delvier Mail, CRS Report for Congress. http://www.nrlca.org/userfiles/File/USPS_Use_of_Contractors.pdf

LaMarca, Anthony and Eyal deLara. 2008. Location Systems, An Introduction to the Technology Behind Location Awareness. Breinigsville, PA: Morgan and Claypool

.

Lexington County, South Carolina GIS. 2010. Lexington County Card Address Range. See http://www.lex-co.com/GIS/documents/CARD.pdf

Lombard. Martha, et. al. 2006. Overview of Draft Street Address Standard, PowerPoint presentation. Available through the FGDC website at http://www.fgdc.gov/standards/projects/FGDC-standards-projects/street-address/06-09-21.ASWG.Presentation.URISA2006a.pdf/view

National Research Council. 2007. National Land Parcel Data: A Vision for the Future. Washington, DC: National Academies Press. See http://books.nap.edu/catalog.php?record_id=11978

National Emergency Number Association. 2010. NENA Master Glossary of 9-1-1 Terminology. USA: NENA. See http://www.nena.org/sites/default/files/20100929_MG_00-001_V14.pdf

NLPG. Undated. BS7666 Explained. See http://www.iahub.net/docs/1183553456634.pdf

Shirky, Clay. 2008. Here Comes Everybody, The Power of Organizing Without Organizations. New York NY: Penguin Press

Staab, Jennifer M. and Vincent G. Iannacchione. 2003. Evaluating the use of residential mailing addresses in a national household survey. Joint Statistical Meeting. See http://www.amstat.org/sections/SRMS/Proceedings/y2003/Files/JSM2003-000471.pdf

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

Tapscott, Don and Anthony D. Williams. 2008. Wikinomics, How Mass Collaboration Changes Everything. London: Penguin Books

U.S. Census Bureau, 2003, Census Bureau to test changes in questionnaire, new response technology, Suitland Maryland: Public Information Office.

U.S. Census Bureau. 2005. Global Positioning Systems (GPS) Field Data Collection Guidelines (v.4.0). Suitland, MD: GPS Application Strategy Working Group. (Internal document)

U.S. Census Bureau, 2008 ACS Design and Methodology, Suitland, MD: American Community Services Office.

U.S. Census Bureau. 2010. Planning for the Geographic Support System Initiative. Report from the January 19-21, 2010 offsite meeting, Geography Division, Suitland, MD: U.S. Census Bureau.

U.S. Census Bureau. 2010. Geographic Support System Initiative (GSS Program). Version 20090528, updated 2010.01.06, Suitland, MD: Geography Division.

UK GOVtalk. 2010. UML Diagram for BS 7666. See http://www.cabinetoffice.gov.uk/govtalk/schemasstandards/e-gif/datastandards/address/bs7666_address.aspx

United States Postal Service. 1995. 1994 Comprehensive Statement on Postal Operations. United States Postal Service. Washington. DC: United States Postal Service.

United States Postal Service. 2002. 2001 Comprehensive Statement on Postal Operations. United States Postal Service. Washington. DC: United States Postal Service.

United States Postal Service. 2006. 2005 Comprehensive Statement on Postal Operations. Washington. DC: United States Postal Service.

United States Postal Service. 2009. Cass Technical Guide 2009-2010 Cycle. Washington, DC: United States Postal Service. See http://ribbs.usps.gov/files/CASS/TECHNICAL_GUIDES/CASSTECH.PDF

United States Postal Service. 2009. 2008 Comprehensive Statement on Postal Operations. Washington. DC: United States Postal Service.

United States Postal Service. 2010. 2009 Comprehensive Statement on Postal Operations. Washington. DC: United States Postal Service.

Universal Postal Union. 2010. UPU Addressing Conference, 2010. See http://www.globaladdress.org/wp-content/uploads/downloads/2010/10/UPU-adressing-conference.pdf

URISA. 2006. Executive Summary: Final Draft, Street Address Standard. Available from the FGDC at http://www.fgdc.gov/standards/projects/FGDC-standards-projects/street-address/06-09-17.ASWG.ExecSumm.Rev4a.pdf/view

Vieira, Veronica M., Gregory J Howard, Lisa G Gallagher, and Tony Fletcher, Geocoding rural addresses in a community contaminated by PFOA: a comparison of methods. Environmental Health 2010, 9:18doi:10.1186/1476-069X-9-18 http://www.ehjournal.net/content/9/1/18

Watelet, Marcel. 2009. ICT Policy Support Programme in the CIP – WP 2009, Teleregions Network Workshop, PowerPoint Presentation. 2009. http://www.teleregionsnetwork.org/filebank/EC-Marcel%20Watelet%20-%20TeleRegions2009.pdf

Watelet, Marcel. 2010. Econtentplus and the INSPIRE Directive: a framework for address harmonization in Europe. 2010. PowerPoint presentation. See https://www.euradin.eu/Lists/Announcements/Attachments/32/2_1_EURADIN%20INSPIRE-GEO%20BXL%2015.06.2010.pdf

Wells, Martha McCart, et. al. 2009. Overview of Draft U.S. Address Data Standard. *PowerPoint presentation. See* http://www.slideshare.net/NSGIC/address-standard

Zacharias, Hanes. 2010. Fostering collaboration with Spatial Data. 2 See http://www.magicgis.org/magic/symposiums/2010/presentations/zacharias_ses.pdf

Zielstra, Dennis and Alexander Zipf. 2010. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany, 13th AGILE International Conference on Geographic Information Science, http://plone.itc.nl/agile_old/Conference/2010-guimaraes/ShortPapers_PDF%5C142_DOC.pdf

Appendix 1. DSF Field Descriptions

These are the fields describing the layout of the data provided to the Geography Division of the Census Bureau by the United States Postal Service.

Field #	Field Description	Length	Start	End		Notes
1	Copyright Detail Code	1	1	1		
2	Delivery ZIP Code	5	2	6		
3	Delivery Carrier Route ID	4	7	10		
4	Delivery Sequence Number	4	11	14		
5	Action Code	1	15	15		A = Add D = Delete (Always 'A' for base file)
6	Record Type Code	1	16	16		F=Firm G=General Delivery H=Highrise P= PO Box R = Rural Route/Hwy Contract S=Street
7	Delivery Address Number	10	17	26		
8	Street Pre Drctn Abbrev	2	27	28		
9	Street Name	28	29	56		Street Name includes any prefix street type
10	Street Suffix Abbrev	4	57	60		
11	Street Post Drctn Abbrev	2	61	62		
12	Secondary Address Abbrev	4	63	66		
13	Apt Room Number	8	67	74		
14	Building or Firm Name	40	75	114		
15	Delivery Point Type Code	1	115	115		**RANGES:**
						A THRU H = RESIDENTAL (INCLUDED IN DELIVERY STATISTICS)
						R THRU W = RESIDENTAL (EXCLUDED FROM DELIVERY STATISTICS)
						Y THRU Z = RESIDENTAL (EXCLUDED FROM DELIVERY STATISTICS)
						I THRU P = BUSINESS (INCLUDED IN DELIVERY STATISTICS)
						1 THRU 8 = BUSINESS (EXCLUDED FROM DELIVERY STATISTICS)
						Q = GENERAL DELIVERY (INCLUDED IN DELIVERY STATISTICS)
						X = UNKNOWN (EXCLUDED FROM DELIVERY STATISTICS)

					DETAILED VALUES:
					A = RESIDENTIAL CURBLINE (INCLUDED IN DELIVERY STATS)
					B = RESIDENTIAL NDCBU (INCLUDED IN DELIVERY STATS)
					C = RESIDENTIAL CENTRAL (INCLUDED IN DELIVERY STATS)
					D = RESIDENTIAL OTHER (INCLUDED IN DELIVERY STATS)
					E = RESIDENTIAL FACILITY BOX (INCLUDED IN DELIVERY STATS)
					F = RESIDENTIAL CONTRACT BOX (INCLUDED IN DELIVERY STATS)
					G = RESIDENTIAL DETACHED BOX (INCLUDED IN DELIVERY STATS)
					H = RESIDENTIAL NONPERSONNEL UNIT BOX (INCLUDED IN DELIVERY STATS)
					I = BUSINESS CURBLINE (INCLUDED IN DELIVERY STATS)
					J = BUSINESS NDCBU (INCLUDED IN DELIVERY STATS)
					K = BUSINESS CENTRAL (INCLUDED IN DELIVERY STATS)
					L = BUSINESS OTHER (INCLUDED IN DELIVERY STATS)
					M = BUSINESS FACILITY BOX (INCLUDED IN DELIVERY STATS)
					N = BUSINESS CONTRACT BOX (INCLUDED IN DELIVERY STATS)
					O = BUSINESS DETACHED BOX (INCLUDED IN DELIVERY STATS)
					P = BUSINESS NONPERSONNEL UNIT BOX (INCLUDED IN DELIVERY STATS)
					Q = GENERAL DELIVERY, INCLUDED IN DELIVERY STATS
					R = RESIDENTIAL CURBLINE (EXCLUDED FROM DELIVERY STATS)
					S = RESIDENTIAL NDCBU (EXCLUDED FROM DELIVERY STATS)
					T = RESIDENTIAL CENTRAL (EXCLUDED FROM DELIVERY STATS)
					U = RESIDENTIAL OTHER (EXCLUDED FROM DELIVERY STATS)
					V = RESIDENTIAL FACILITY BOX (EXCLUDED FROM DELIVERY STATS)
					W = RESIDENTIAL CONTRACT BOX (EXCLUDED FROM DELIVERY STATS)
					Y = RESIDENTIAL DETACHED BOX (EXCLUDED FROM DELIVERY STATS)
					Z = RESIDENTIAL NONPERSONNEL UNIT BOX -
					(EXCLUDED FROM DELIVERY STATS)

						X = EXCLUDED FROM DELIVERY STATS, NO OTHER INFO PROVIDED
						1 = BUSINESS CURBLINE (EXCLUDED FROM DELIVERY STATS)
						2 = BUSINESS NDCBU (EXCLUDED FROM DELIVERY STATS)
						3 = BUSINESS CENTRAL (EXCLUDED FROM DELIVERY STATS)
						4 = BUSINESS OTHER (EXCLUDED FROM DELIVERY STATS)
						5 = BUSINESS FACILITY BOX (EXCLUDED FROM DELIVERY STATS)
						6 = BUSINESS CONTRACT BOX (EXCLUDED FROM DELIVERY STATS)
						7 = BUSINESS DETACHED BOX (EXCLUDED FROM DELIVERY STATS)
						8 = BUSINESS NONPERSONNEL UNIT BOX (EXCLUDED FROM DELIVERY STATS)
16	Addr Vacant Ind	1	116	116		Y=Vacant For 90 Days or More N=Not Vacant
17	Delivery Point Drop Ind	1	117	117		Y=Delivery Point Is A Drop N=Delivery Point Is Not A Drop
18	Bus Fam Served Cnt	3	118	120		# of Businesses or Families Served at Drop Point
19	Seasonal Delivery Ind	1	121	121		Y=Seasonal Delivery N=Not Applicable
20	Zip Add On Low No Zip Sector No Zip Segment No	 2 2	 122 124	 123 125		
21	Zip Add On High No Zip Sector No Zip Segment No	 2 2	 126 128	 127 129		
22	Base Alt Code	1	130	130		B=Base A=Alternate
23	Lacs Status Ind	1	131	131		L=Lacs Converted Blank=Not Applicable
24	Addr Sort Instruction Code	1	132	132		T=PO Box Delivery Blank=Not Applicable
25	Finance No	6	133	138		Postal Finance Number (PFN)
26	State Abbrev	2	139	14		Alphabetic
27	Congressional District ID	2	141	142		Numeric
28	Municipality City State Key	6	143	148		Link to PO Name in City-State File
29	Urbanization City State Key	6	149	154		(Puerto Rico Only) Link to Name of Urbanizacion in City-State File
30	Prefd Last Line City State Key	6	155	160		Link to Preferred PO Name, State, and ZIP in City-State File
31	Scheme Zip Code	5	161	165		ZIP Code that Cases This Address

32	Scheme Carrier Route ID	4	166	169		
	Permanent DSF ID					Permanent DSF ID is comprised of char's. 170 -208. These values are supposed to remain constant, even if the corresponding field above changes for this address
33	DELVPT-SYSTEM-ID	26	170	195		
34	REC-TYPE-CODE	1	196	196		**Same Legal values as Record Type Code Char. 16**
35	SCHM-ZIP-CODE	5	197	201		
36	BASE-ALT-CODE	1	202	202		**Same Legal values as Base Alt Code Char. 130**
37	DELVPT-TYPE-CODE	1	203	203		**Same Legal values as Delivery Point Type Code Char. 115**
38	CNTY-STATE-ABBREV	2	204	205		Alphabetic State Abbrev
39	CNTY-NO	3	206	208		FIPS County Code

Appendix 2. Highway Carrier Route Addressing Formats

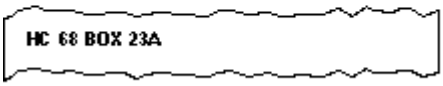
For more information on HCR addresses, see USPS Publication 28 on Addressing. The online version for the section on HCR Address formats in Publication 28 can be found at http://pe.usps.com/text/pub28/28c2_005.html .

25 Highway Contract Route Addresses

251 Format

Print highway contract route addresses on a mailpiece as: HC N BOX NN. Do not use the words *HIGHWAY CONTRACT, ROUTE, NUMBER, NO., STAR ROUTE,* or the *pound sign (#).*

Incorrect

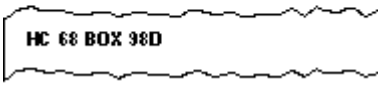


Correct

252 Leading Zero

A leading zero before the highway contract route number is not needed.

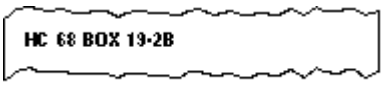
Acceptable



Preferred

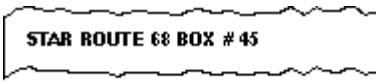
253 Hyphens

Print hyphens as part of the box number only when they are part of the address in the ZIP+4 file.

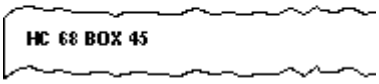


254 Star Route Designations

Change the designation *STAR ROUTE,* which usually refers to highway contract route, to *HC.*



Incorrect



Correct

255 Additional Designations

There should be no additional designations, such as town or street names, on the Delivery Address Line of highway contract route addresses. Street names used together with route and box numbers can create potential matching difficulty. Mailers are encouraged to use only one style of addressing. If secondary name information is used, however, place it above the Delivery Address Line.

Incorrect

Correct

256 ZIP+4

When applying a ZIP+4 Code to a highway contract route address, an exact match is preferred. If a box number is included in the address, the mailpiece must bear the appropriate ZIP+4 Code representing the range for that box number. When box number information is not available, the highway contract base record must be used.

Appendix 3. E911 Definitions

(The definitions in italics below were taken from the NENA Master Glossary of E 9-1-1 Terminology, NENA 00-001, Version 13a, June 3, 2010. See http://www.nena.org/sites/default/files/20100929_MG_00-001_V14.pdf)

Automatic Location Identification (ALI) in the E9-1-1 system provides for the automatic display of the caller's telephone number and the address of the residence where the call originated, if the call originated as a landline call. In addition, the ALI supplies information on the emergency services for the location from which the call originates. The ALI is informed, by another database, usually called the Master Street Address Guide, although it may have other names, such as the E9-1-1 data table.

The Master Street Address Guide:

The MSAG is data base of street names and house number ranges within their associated communities defining Emergency Service Zones (ESZs) and their associated Emergency Service Numbers (ESNs) to enable proper routing of 9-1-1 calls.

Master Street Address Guide Address:

Address recognized by Public Safety for the dispatch of emergency first responders. It is an absolute and unique address in that variants for directions, street spelling, street suffixes, and community names are not allowed. It is preferred that MSAG Addresses be in Civic Address format. The community name associated with this address format is assigned by the Addressing Authority in cooperation with the 9-1-1 Administrator and may or may not be the same as the community name assigned by the USPS.

MSAG addresses are used to route 9-1-1 calls and for ALI display.

NOTE: MSAG Address data format is not standardized throughout the country. This is generally attributed to legacy system limitations that have been continued as operational practices on the part of 9-1-1 administrative entities. This fact gives rise to the need for two subtending MSAG definitions.

More recently, a graphic version of the MSAG databases has become popular and is known by various names. Contact One, a company providing "…solutions for 9-1-1 and GIS Response" provides a highly regarded graphic MSAG management product called MapSAG that synchronizes address difference in the various databases used in 9-1-1 systems. MapSAG is based on ArcGIS technology. More about the product and company can be found at http://www.contactone.com/prod_mapsag.htm .

Civic Address

Any city-style address that includes a house number and a street name is considered a Civic Address. Civic Addresses include a community name that may or may not be recognized by the USPS or be MSAG valid. Civic Addresses may be used as Postal address if recognized by the USPS. Civic Addresses may be used as MSAG addresses if they are an exact match to the MSAG address. A rural route delivery address or FPO or APO address is not considered a Civic Address. See Civic Location.

Civic Location

A set of elements that describe detailed street address information. See Civic Address.

Civic to MSAG Translation

Translations consist of processes, tables or rules that can be used to:

Translate a Civic Address from/to an MSAG format address.

Translate a Postal Address from/to an MSAG format address.

Translate Landmark from/to an MSAG-format address (e.g. Empire State Building --- 350 5TH AVE).

A civic address may need to be translated so that the data is consistent to the format within the existing PSAP equipment and processes (i.e., CAD, mapping, CPE).

Appendix 4. Specific Plans for NTIA Broadband Addressing/Mapping

The data in this table were gathered from the NTIA Website sections on the State Broadband Data and Development Program (http://www2.ntia.doc.gov/SBDD). It was current as of November 2010.

State or Territory	Address File Development/Mapping
American Samoa	American Samoa: NTIA has awarded the American Samoa Office of the Governor approximately $558,000 for broadband data collection and mapping activities over a two-year period and $500,000 for broadband planning activities over a two-year period in American Samoa, bringing the total grant award to approximately $1.1 million. The Office of the Governor is the designated entity for the territory of American Samoa.
Arizona	The Arizona State Cartographer's Office at the Arizona State Land Department will partner with the Arizona E-911 Program to assist counties in developing accurate and current geocodable street networks and address point locations in 15 Arizona counties. This data will greatly assist the Arizona Broadband Mapping Project by providing a quality data resource for data development and verification, especially where census blocks are larger than two square miles.
Arkansas	In collaboration with the Arkansas Geographic Information Office (AGIO) and using the state E-911 file and U.S. Postal Service address database, Connect Arkansas will further develop a statewide address file, focusing first on unserved and underserved communities. During the three-year life of the project the AGIO will complete the data set and will have the files available for public use and download for 20 counties.
Colorado	OIT will collect and add to current county data and deploy a crowdsourcing tool to improve address file quality. OIT will validate county information against data from the Colorado Division of Motor Vehicles and Secretary of State's Office. This project will also work with counties to train emergency service personnel to collect and verify address location information.

Indiana	The State of Indiana Office of Technology will support address development of raw data submitted by counties, data hosting and transfer costs for address point and centerline aggregation, and a small portion of the costs required to refresh orthophotography with other state partners.
Louisiana	The State of Louisiana will build out address files through collaboration with all parishes, but specifically focusing on rural communities and Acadian Ambulance, which holds a point address file for 15 counties. Once the files are complete, Louisiana will possess an authoritative address file for the entire state, a street layer of census blocks greater than two square miles, and an electronic solution for managing addresses available to parishes.
Maine	The State of Maine will conduct a two-stage process that conflates the existing E-911, ConnectME and Department of Transportation centerline files, and then utilizes dynamic segmentation to generate an address file for the majority of communities and a traditional assignment approach for the few communities that did participate in the previous address normalization project.
Massachusetts	MBI is proposing two innovative and cost-effective solutions to create address point files to improve geocoding accuracy: 1) Focusing on 45 towns and census blocks greater than two square miles, MBI will utilize existing digital parcel data and real estate files to automate the creation of accurate address files; 2) MBI will implement a proof-of-concept model for crowd sourcing addresses across 35-45 towns, and will expand the work to all towns in the state after reviewing the effectiveness of the crowd sourced mechanism.
Missouri	The awardee will develop a situs point-based address file that will support more granular broadband data collection and also the needs of local governments and the public safety community. Participating agencies and localities will provide local match for the project, increasing local ownership of the activity.
Montana	Montana proposes to partner with 16 counties and 7 tribal governments to geocode current cadastral information in an efficient and cost-effective manner.
New Hampshire	The New Hampshire Broadband Mapping Program proposes an innovative strategy that leverages relationships between UNH, Regional Planning Commissions, local/state agencies and community volunteers to develop the first public master address file for the state.

North Carolina	The North Carolina Center for Geographic Information and Analysis (CGIA) will build on the statewide addressing effort that began prior to the 2010 census. CGIA and e-NC will form a working group of local governments to document best practices identified in the first phase and will also develop an automated procedure for filtering the data that has remained unchanged between address file data collection cycles
Ohio	The State of Ohio maintains an existing partnership, the Location Based Response System (LBRS), between the state and local governments, to build accurate field verified address databases. This project will support direct funding to rural counties that lack the technical and/or financial resources required for LBRS participation
Rhode Island	The Office of Information Technology will collect and improve current county data and deploy a crowdsourcing tool to improve address file quality. OIT will conflate that information with data from the Rhode Island Department of Motor Vehicles and the Secretary of State's Office. This project includes additional support to train emergency service personnel to collect and verify address location information.
South Dakota	This project will create a comprehensive geocodable point dataset for the state. The Department of Revenue and Regulation will assist BIT with collecting existing addresses, and will coordinate the datasets through county offices. The counties will receive funding to assist them in creating county-level address files. Gap analysis will be performed on the master list from the counties and missing addresses will be crosschecked against data held by the county office and through 9-1-1 Public Safety Answering Points. A web portal will be then created to allow stakeholders to see the addresses and to validate information.
Utah	Utah PSC proposes to develop a crowd sourcing application for address development through a public-private partnership with Blue Stakes of Utah, and will also leverage state-to-county relationships and Utah's 911 Committee to develop a statewide address file.
Virginia	Virginia: NTIA has awarded the Virginia Center for Innovative Technology $1,799,979 for broadband data collection and mapping activities over a two-year period and $500,000 for broadband planning activities over a two-year period in Virginia, bringing the total grant award to approximately $2.3 million. The Virginia Center for Innovative Technology is the designated entity for the state of Virginia.

Virgin Islands	VIPFA will begin address file assignment and collection in areas greater than two square miles by year three of the project period. This will be especially valuable because of the current numbering system, which uses a Danish system in some areas and which is non-existent in others. The requested funding covers planning and residence-by-residence assignment of addresses in target areas.
Wisconsin	The Wisconsin Land Information Program (WLIP) was formed to meet the parcel data needs of state agencies, regional planning commissions, emergency managers, tribal agencies and other stakeholders. SBDD funding will support WLIP's design, development, integration, testing, outreach, support, and data production necessary for securing statewide parcel and address data.