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 Materials from the Offsite remain unpublished and are not available for dissemination.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

SPECIAL NOTE: Throughout this report, wherever the authors refer to "a product database," they use the term in a generic manner and do not imply that the Geography Division's Product Database would be the specific database in which processes related to exchange of data with external partners would occur. References to the Geography Division's Product Database and the processes related to it are intended as examples of the way in which data exchange processes might be modeled.

Department of Commerce
Bureau of the Census
Geography Division

Researching Address and Spatial Data Digital Exchange and Data Integration

A Report to the Geography Division, U.S. Bureau of the Census
Deliverables #8 and #9 – Syneren Technologies Contract – Task T005

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

Purpose of the Study

Dictionary definitions of "exchange" refer to the concept of reciprocal giving and receiving. In the context of GIS, data exchange is a data communication process perhaps better characterized as data import and data export. The data exchange process is not typically reciprocal, rather it involves users importing address and spatial data they have obtained (perhaps purchased) from a set of data exporters (or providers), typically government agencies or commercial data sources.

The need to be able to receive data from a number of sources and to exchange data between dissimilar geodatabases presents a number of challenges. From the perspective of Census and TIGER, the challenge involves integrating a heterogeneous data environment. The various aspects of this heterogeneous environment that affect the import and export operations must be understood for the data exchange or transfer process to be meaningful and successful.

Every geodatabase has its own, unique (to varying degrees), conceptual and logical data models that manifest themselves in a unique physical data model (e.g. TIGER in Oracle Spatial Topology Model; ESRI Geodatabases in ArcSDE + RDBMS). To export a data set, the internal physical data model is converted to a specific file structure and transferred via physical media or networks. To import a data set, the process is reversed. A variety of file structures are in use to export and import data. While these formats may be capable of exchanging data between disparate systems, the goal of data exchange is the transfer of information to communicate an understanding of the phenomenon being represented. What on the surface may seem to be a straightforward task has many added levels of complexity, particularly when data content and quality issues are considered. Each of these data content issues needs to be understood for the data exchange process to be completely successful.

The mechanisms for data exchange exist, but how (or on what basis) should data exchange between the parties occur? In the report that follows, we have pursued this question guided by the requirements of the Geography Division.

Scope of the Study

Four areas related to the concept of spatial data exchange and integration were identified for study and analysis.

We begin with an overview of address and spatial data sources, including a description of the challenging issues commonly encountered by the Geography Division during the data exchange process with its data partners. Data exchange methodologies and evaluation are the second focus of our research. Next, we identify and report on relevant components for the successful exchange of address and spatial data between the Census Bureau and its partners. The information discovered in the course of this research (Deliverable #8) is used to plan a work breakdown schedule of a data exchange pilot described in the final section of this report (Deliverable #9).

The interrelationships between the selected research topics are many and by necessity, there is some content overlap in this presentation with the information in the deliverables for Tasks 1, 2, 3 and 5. The scope of our assignment spans both current and future directions. In some cases, our focus reflects benefits that could accrue from adopting existing practices, while in other instances; it is focused on new or emerging techniques and practices that may become standard in the future. We believe that the research topics presented are relevant to understanding, improving, and supporting the spatial data exchange and integration activities of the Geography Division. In turn, these are topics that could lead to improvement in the MTdb and the Census programs that rely on the use of this database, particularly those uses related to the goals of the GSS program.

Research Topics

Task 1. Spatial Data Sources

We have been asked to evaluate problems with the sources of address and spatial data that have been exchanged by the Geography Division, as well as discovering sources of address and spatial data that have not been extensively used by the Geography Division. Discovering and describing address and spatial data sources, as well as their benefits and inadequacies, is a task that has been woven into our previous reports. In our report on addresses and addressing (Deliverable 2, Task 1), we described specific and general sources for address data, including governmental organizations, commercial entities, and not for profits. We note here that many of the sources we examined in that report provide spatial data in addition to the address data, which was the focus of Task 1. In our final deliverable for Task 3 on the future directions of technology, we surveyed sources of spatial data based on their underlying technology. In both Task 1 and Task 3 we described crowdsourcing and its potential application as a data source for addresses and other spatial data. In addition, we identified commercial and not-for-profit organizations gathering geospatial data through the use crowdsourcing.

1.a Identification of existing problems related to address and spatial data that have been exchanged between Geography Division and its internal and external partners.

The Geography Division has had considerable experience in exchanging and integrating address and spatial data with various partners. The yearly Boundary and Annexation Survey (BAS) and the once a decade Local Update of Census Addresses (LUCA) program are prime examples of the current efforts of the Geography Division to “exchange” data with external partners. The

2011 BAS, for example, allows participants to choose to submit data on boundary changes on Census supplied paper maps, using the MAF/TIGER Partnership Software (MTPS), or using their own GIS to modify shape files supplied by the Census Bureau. The 2010 LUCA program, which was focused on address update, offered participants the ability to review and update city-style addresses on the Census Bureau's address list, challenge address counts in census blocks, make map updates and appeal feedback results after address canvassing.[1] Several modes of updating were available, including the use of paper maps, the MTPS software, or using a GIS to update Census supplied shapefiles. Because LUCA is focused on updating addresses and involved Title 13 issues, some of the use-options included only a subset of the topics described above.

From an overall perspective, the Geography Division's data exchange programs appear to be well-structured, thoughtful and robust approaches to the requirements defining the individual programs. Participants were provided assistance seminars, workshops, computerized training, self-training modules, and a help desk for those using the MTPS software. Throughout the process changes were made in procedures and new practices were implemented aimed at accommodating the needs and preferences of the data exchange partners.

Our review of LUCA and BAS indicates that these programs fell short of being successful examples of data exchange and integration. In numerous cases the data provided by contributors was difficult for the Census to integrate, and equally often did not satisfy the needs of data partners for ease of preparation or even provide data that was intended to be exchanged. However, it is important to note that these programs are designed for jurisdictional bodies to provide data to the Census Bureau to allow the Bureau to fulfill its requirements in support of the Decennial Census, the American Community Survey and a host of related, official, surveys and census activities. In essence, these programs are more focused on the Census collecting and integrating the required data within the MTdb, than on the exchange of specific data that would be of benefit to both parties. While it is clear that the government entities participating with the Census in these types of programs receive significant indirect benefits from their efforts, the participants appear less satisfied with the ease of contributing their data and the relative absence of operational benefits from sharing data with the Census Bureau. In turn, the Geography Division often expended considerable effort to convert the data supplied to it into a form that could meet its requirements.

In the section of the report that follows, we provide a brief overview of the difficulties with the data exchange process between the Geography Division and its partners, as currently practiced. We raise these issues not to criticize any party, but to indicate current practices inhibiting successful exchange of data that need to be modified to improve the future results of data exchange efforts.

[1] U.S. Census Bureau, 2008, 2010 LUCA Promotional Workshop, Suitland , Maryland: U.S. Bureau of the Census see http://www.census.gov/geo/www/luca2010/luca_promo_work.pdf

Problems with existing approaches to data exchange

One of the most common problems in data exchange appears to be the inability or unwillingness of data partners to meet Census Bureau's requirements to preserve topological relationships and spatial accuracy. The Geography Division is very serious about topology and its requirements for topology in the MTdb serve to distinguish it from most Geographical Information Systems (GIS) or Land Information Systems (LIS). For example, the illustration below from the BAS Respondent Guide: Digital BAS, 2010 shows the variation between parcel-based and topologically-based approaches. Many governments use parcel-based systems for recording their legal property and right-of-way-mapping (left image). The approach taken by Census (right image) is based on topologically integrated spatial data which make handling offsets (e.g. between property boundaries along roads) inefficient. Census prefers that data supplied to it have the boundary along the front of the lot snapped to the road. Users of parcel-based systems may see this action as extra work and consider it to be inaccurate extra work that cannot be imported back into their system. However, these agencies may be required to make these changes, since the Census suggests that snapping an entity boundary to the centerline, when applicable, will help to establish more accurate population counts.[2]

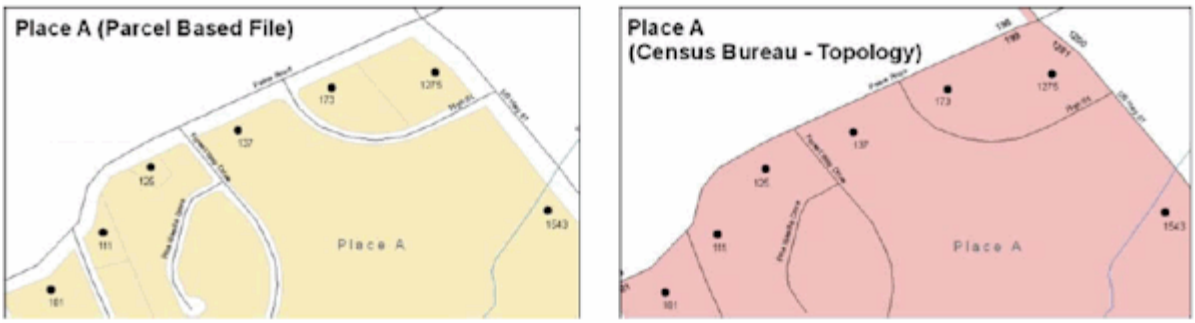
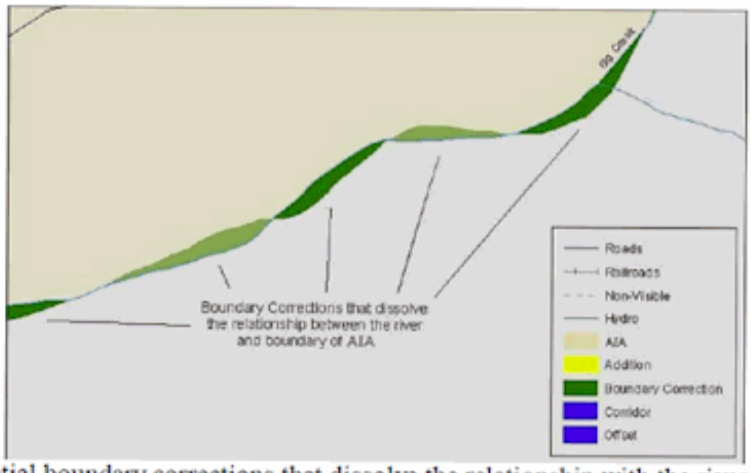


Figure 1. Parcel versus Topology-based files.

A related problem is the inability or unwillingness of some data exchange partners to edit a Census Bureau-supplied shapefile. Census requires that entities update Census Bureau shapefiles with boundary and feature changes, rather than submitting a shapefile from a local Geographic Information System. The issues here are accuracy of location and topology. During the last decade through the MAF/TIGER Accuracy Improvement Program (MTAIP), the Census Bureau precisely, and at considerable expense, repositioned the elements of the MTdb to accurately reflect their positions in the real world, as well as to correctly delineate the addresses of living quarters allowing the Census Bureau to enumerate population and assign the results to the correct Census geography. The Geography Division believes that their representations of geographical space are accurate and, in order to preserve topological relationships between these elements, prefers that all editing be done on the shapefiles they supply because these shapefiles

[2] U.S. Census Bureau, 2010, Boundary and Annexation Survey Respondent Guide: Digital BAS, Suitland, Maryland; Geography Division, U.S. Bureau of the Census (PDF) see http://www.census.gov/geo/www/bas/bas11/11_RespondentGuide_DigitalBAS_Local.pdf

benefit from the realignment and topological processing that the Census has undertaken to improve the address and spatial data quality of the MTdb. (See Figure 2.)



Example 4.4: Small spatial boundary corrections that dissolve the relationship with the river have been created. These boundary corrections will not be incorporated into the MAF/TIGER database.

Figure 2. Spatial boundary representations by partners often provide a very inaccurate representation of Census Geographic Areas.[3]

It is our understanding that numerous partners believed that the GIS data that they had already prepared should be usable by the Census and, in some cases, joined their data to the data supplied by the Census. Doing so created a data file that was so complex that the Geography Division had to spend significant time trying to untie the data to determine the changes that had been made.

It appears that the problem with partners being unwilling to edit the shapefiles extracted from the MTdb reflects a difference of opinion on whose data is authoritative. Local authorities, quite reasonably, might be inclined to assume that their data reflects local conditions. On the other hand, the requirements of local GIS systems may not be based on specifications as rigid as the Census specifications for data quality and topology. For example, if linear feature changes do not align with current MAF/TIGER linear features, the Census Bureau may not incorporate the submitted updates.

Another common problem is the inability or unwillingness of external partners to follow relatively straightforward exchange procedures set by Census. For example, under LUCA's option 1 for the 2010 program, in order to update a city-style address (add, delete or otherwise correct) the participant was required to provide the Census geocodes including state code, county code, census tract number and census block number, as well as to identify any additions, deletions or corrections to jurisdictional boundaries, road, or other physical features on the Census Bureau maps, or submit digital feature updates.[4] It is our understanding that the required

[3] U. S. Census Bureau, Ibid.
[4] U.S. Census Bureau, 2008, 2010 LUCA Promotional Workshop, Suitland , Maryland: U.S. Bureau of the Census see http://www.census.gov/geo/www/luca2010/luca_promo_work.pdf

geocodes often were not provided by participants and when provided often included attributes that did not refer to the actual geography associated with the addresses. In turn, the Geography Division's requirements for geographic offsets and corridors[5] to enhance geocoding are likely confusing to many local data partners.

In a somewhat curious example of the inability of some participant organizations to follow what were thought to be relatively straightforward directions, the MAF extracts provided to communities participating in the LUCA program were delivered in pipe-delimited format. Working in the pipe-delimited format seemed to have caused a number of participants significant difficulties, even though the Census provided nine pages of specific instructions on how to work with pipe-delimited ASCII text files in Excel 2002.[6,7]

Our understanding is that it was a rare occurrence when an external partner was able to provide a dataset that met the Census Bureau requirements and did not require some degree of alteration. It was common for address data to require editing and for map data to be re-digitized by Census staff before it was entered into the MTdb. However, not all limitations were due to action or inaction on the part of those providing data to the Census Bureau. Many of the contributors found that meeting the requirements set by the Census required considerable sweat equity and in some cases required extraordinary budget allocations to fund the required work. For example, Joseph Salvo claimed that

> "In New York City, our LUCA operation took place over three years, involved about 60 people, leveraged the resources of the largest planning department in the country. I had to get a $250,000 grant for three separate field operations to work on the address list, to get it to a point where we felt like we had something that was real."[8]

The New York City effort appears to have added approximately 130,000 addresses to the MAF for purposes of validation during Address Canvassing operations which were conducted in 2009.

Shoreh Elhami, GIS Coordinator for Delaware County, Ohio, manages an inventory of approximately 77,000 addresses and found that preparing materials for LUCA required the efforts of two and half full time equivalents over a four month period.[9] In order to process the county's data and update the MAF she and her staff spent a great deal of time analyzing data and

---

[5] U.S. Census Bureau, 2010, Boundary and Annexation Survey Respondent Guide: Digital BAS, Suitland, Maryland; Geography Division, U.S. Bureau of the Census PDF see http://www.census.gov/geo/www/bas/bas11/11_RespondentGuide_DigitalBAS_Local.pdf

[6] U.S. Census Bureau, 2008, FAQ 77. Converting Microsoft Excel 2002 to Pipe-Delimited ASCII Text Files. Suitland, Maryland: U.S. Bureau of the Census, (online FAQ). see http://www.census.gov/geo/www/luca2010/luca_faq/FAQ_77_for_2010_LUCA.pdf

[7] U.S. GAO, 2007, 2010 Census: Census Bureau Has Improved the Local Update of Census Addresses Program, but Challenges Remain, Washington, DC; US. Government Accounting Office. p. 8. See http://www.gao.gov/new.items/d07736.pdf

[8] Salvo, Joseph, 2010 Counting Every Address in the Census: Joseph Salvo at Pew Research Center, New York City; Pew Research Center see http://census.pewsocialtrends.org/2010/counting-every-address-in-the-census-joseph-salvo-at-pew-research-center

[9] Elhami, Shoreh, 2010, personal communication.

matching the MAF to the county's Master Address Point File.[10] Shoreh Elhami was able to submit approximately 10,000 additions, 5,000 corrections and 7,000 deletions to help improve the MAF.

While successes such as those in New York and Ohio show the worth of data exchange programs, they also indicate that the processing of the data required by the Census Bureau can take considerable time, effort, and resources on the part of its data partners.

Another problematic factor in the exchange of data with the Census Bureau is the dissatisfaction of external partners with the non-acceptance of data they submit. While the Census Bureau must be the arbiter of what is or is not a valid address identifying a living quarter, or other data entered into the MTdb, the quasi-adversarial nature of the interactions with participants may not build the desired participatory framework. For example, some of the requirements imposed on LUCA participants while reasonable from the Census Bureau's point of view, must be vexing to the participants. For example, consider this section of the LUCA Feedback User Guide Supplement:[11]

> "Note: In order to determine which version of an address to use on the Full Address List for census questionnaire mailing, the Census Bureau defers to:
>
> 1. The U.S. Postal Service's (USPS) Delivery Sequence File (DSF)
> 2. Field findings
>
> For example, the initial mailing address sent to a LUCA participant contained the street name "PNE ST." The participant submitted a correction (C) to the street name as "PINE ST." Although address canvassing found the street name as "PINE ST," the Census Bureau deferred to the USPS DSF, since the mailing address supplied by the USPS is "PNE ST," which is used on the Full Address List."

Participants who change address names knowing that the name they supplied is correct and used on a local street sign may be somewhat displeased by those situations where they took the time to change the street name, know that it was verified by Census field workers, and changed back to its original but inappropriate spelling used by the Census. The issue here, again, is one of authority and authoritativeness. Our point is that while actions of this nature are understandable given the mission of the Census Bureau, these kinds of conflicts do not build strong partnerships and should be bred out of future data exchange strategies.

Another unavoidable but contentious issues for participants is the confidentiality that must be enforced with any LUCA process related to addresses, due to Title 13 restrictions. Participants must appoint a LUCA liaison and assume responsibility for Title 13 compliance. All reviewers

[10] Elhami, Shoreh, 2008, Census 2010 – Local Update of Census Addresses (LUCA) Program in Delaware County, Ohio, Presentation at URISA 2008 Conference, New Orleans, LA.
[11] U.S. Census Bureau, 2008, LUCA 2010 Feedback User Guide Supplement, Suitland, Maryland: U.S. Bureau of the Census. See http://www.census.gov/geo/www/luca2010/user_guide_suppl.pdf, p3.

must follow guidelines and sign Confidentiality Agreements.[12] The punishment for violating Title 13 U.S.C. is a fine of up to $250,000 and/or 5 years imprisonment. It is likely that these requirements cause some communities to avoid participating in data exchange. It is our understanding that more than 39,000 communities (e.g. states, counties, federally recognized tribes, Incorporated places, functioning minor civil division, etc.) were invited to participate in LUCA. Of that number 11,500 organizations registered for the program, but only 8,185 of these actually provided data.[13] Only approximately 7,000 participating communities chose to provide data under either Option 1 or Option 2, which required signing the confidentiality documents and observing Title 13 restrictions.

Joseph Salvo has indicated that New York City has 60 agency personnel who have signed the Address Confidentiality form required to work with Census addresses under Title 13 Restrictions and wonders how these numbers might balloon under continuous updating.[14] Craig has indicated that the confidential nature of the address information shared by Census with participants in the LUCA program (they can see but not use the address information provided by the Census Bureau for any other purposes) may cause various communities and tribal governments not to participate in the program.[15] Changing Title 13 may be unlikely since it would require legislative action. It appears, however, that the influence of confidentiality restrictions, and more likely the threat of substantial fines or prison sentences, serve to limit the number of communities willing to participate in LUCA.

Summary

Many of the data exchange programs conducted by the Geography Division are not exchanges, although some may be examples of asynchronous exchanges. For example, the data contributed by partners in BAS may be returned to them in the eventual releases of the TIGER Line files, although it may require several years before all of the data elements supplied are processed. Address data, however, are not really exchanged with the LUCA participants, as they can use the Census supplied address data for no other purpose than LUCA and must destroy the address data they received from the Census at the end of the project.

Exchanging address and spatial data with the Geography Division is restricted by confidentiality issues and guided by their stringent requirements for spatial accuracy and topology. It is our opinion that most of the potential participants are either unable or unwilling to provide data deliverables that will satisfy the requirements posed by the Census. We suspect that the present model is unworkable and suggest that creating a new system that avoids or minimizes the problems outlined above, may be the most reasonable strategy for developing future data exchanges.

[12] U.S. Census Bureau, 2008, 2010 LUCA Promotional Workshop, Suitland , Maryland: U.S. Bureau of the Census see http://www.census.gov/geo/www/luca2010/luca_promo_work.pdf
[13] Personal communication David J. Bush, U.S. Census Bureau
[14] Salvo, Joseph op. cit.
[15] Craig, William J., 2006, A Master Address File For State and Local Government, URISA, 2006 Conference Proceedings, see www.nsgic.org/committees1/showDoc.cfm?docID=99

1.b Identify the range of public, commercial and non-profit organizations that are systematically and actively acquiring data that would be relevant to the Geography Division.

For this section of our report, we were asked to survey promising sources of address and spatial data that have not been used extensively by the Geography Division. Note that our inclusion of entities in categories is not meant to imply that they are either new or unfamiliar to the Geography Division, rather that these sources appear to be promising and may have data of relevance to the Geography Division. Rather than pursue our original plan to create a comprehensive inventory of data providers, we agreed with the Geography Division's request to summarize the "pulse" of the state of the available address and geospatial data maintained by local governments and providers of commercial data.

Providers of Spatial Data with National Coverage

a. Commercial

Our initial list of candidates includes NAVTEQ, Tele Atlas, Google, Microsoft, and CoreLogic (formerly First American). [16] The companies listed were selected because they utilize many unique data gathering techniques, for example User Generated Content and probe data (also known as floating car data), as well as other techniques that rely on purpose-built instrumentation, unique data collection vehicles, and advanced imaging systems.

NAVTEQ

Founded in 1986, the former Navigation Technologies was acquired by Nokia in 2008 and is now managed as a stand-alone acquisition not tightly integrated within the parent company. NAVTEQ provides a navigation quality database that is often used for GIS and mapping applications. NAVTEQ does not license address information, as many of the addresses in its database are not proprietary and are licensed from CoreLogic, among other sources.

The company provides street level vector data that can be attributed with up to two hundred-sixty attributes per segment, classified by 14 categories (e.g. road classification, administrative areas, points of interest, land use, country, nodes, conditions, traffic, etc.) (See Figure 3). Our experience indicates that their data coverage of the United States is comprehensive, well attributed, positionally accurate, and includes metadata, although the metadata is not distributed with the database.

[16] Originally, we had intended to research Pitney Bowes Business Insight (MapInfo). However, the company is a reseller of other companies' spatial data and its address data is based on the DSF^2 with no unique address data added. As a consequence, we did not focus on PBBI.

Attribute	Field Name	Format
Link ID	Link_ID	Number(10)
Street Name	Street_Name	VarChar2(80)
Feature ID	Feature_ID	Number(10)
Street Name Language Code	St_LangCode	VarChar2(3)
Number of Street Names	Num_StNames	Number(2)
Street Name Prefix	StNm_Pref	VarChar2(2)
Street Type Before	StType_Before	VarChar2(30)
Street Name Base	StNm_Base	VarChar2(35)
Street Name Suffix	StNm_Suff	VarChar2(2)
Street Type After	StTyp_After	VarChar2(30)
Street Type Attached	StType_Attached	VarChar2(1)
Address Type	Address_Type	VarChar2(1)
Left Side Reference Address	Left_RefAddr	VarChar2(10)
Left Side Non-Reference Address	Left_NonRefAddr	VarChar2(10)
Left Address Scheme	Left_AddrScheme	VarChar2(1)
Right Side Reference Address	Right_RefAddress	VarChar2(10)
Right Side Non-Reference Address	Right_NonRefAddr	VarChar2(10)
Right Address Scheme	Right_AddrScheme	VarChar2(1)
Right Address Format	Right_AddrFormat	VarChar2(1)
Reference End Intersection ID	Ref_Intrsect_ID	Number(10)
Non-Reference End Intersection ID	Nref_Intrsect_ID	Number(10)
Number of Shapepoints	Num_ShapePoints	Number(10)
Functional Class	Func_Class	VarChar2(1)

Figure 3. NAVTEQ's street data attributes (partial) (source NAVTEQ 2006)[17]

[17] NAVTEQ, 2006, NAVTEQ's NAVSTREETS Street Data Reference Manual, Chicago, Il: NAVTEQ v.2.0 see http://faculty.unlv.edu/jensen/gisdata/navteq/NAVSTREETS_Street_Data_Reference_Manual_Q2-2006.pdf

The NAVTEQ data is kept reasonably up-to-date, as one might expect of a database that is used with autonomous, in-car navigation systems.[18] The company's data is updated in a continuous process by its field collection teams, but released quarterly to its licensees. We consider NAVTEQ data to be the highest quality and most comprehensive provided by the commercial database providers.

NAVTEQ utilizes a multifaceted approach to data collection and updating that involves data mining, but is focused on field data collection utilizing its approximately 1200 geographic analysts distributed in 46 countries around the world. Recently the company has developed an advanced mapping collection technology called NAVTEQ True that includes rotating LiDAR sensors capable of capturing 1.5 million 3D data points per second, combined with panoramic and high resolution cameras tracked by GPS and IMU (Inertial Measurement Unit) to collect detailed mapping data in key commercial areas. In addition, the company uses crowdsourcing through active contributions made using a simple interface available at its website.[19,20] NAVTEQ regards data that it collects through field observations to be its most authoritative data, followed by data that relies on official mapping sources in national, state, and local governments. NAVTEQ conflates external data using Esri software, as well as a variety of purpose built tools.[21]

NAVTEQ provides data in a wide variety of exchange formats including Geographic Data File (GDF), ESRI Shapefile, SIF+, POI/XML, and ODF (Oracle Delivery Format)

NAVTEQ's technical capabilities, knowledge base, and skillsets are excellent and their familiarity with local data sources is comprehensive. At a GIS-T conference in 2010, Skip Turner of NAVTEQ indicated that public/private partnerships are workable, but the company would need to charge for what they consider "the secret sauce".[22] He added that NAVTEQ has field geographers to do the validation and that the company would be willing to work with government to put "the geometry in the public domain". Unfortunately, NAVTEQ has made statements on this intent previously, but has never shown significant interest in their data being redistributed in the public domain as a part of TIGER or any other system. NAVTEQ is known as an aggressive negotiator in the industry and we believe that it is unlikely that they would agree to what others might consider reasonable terms. We do not regard NAVTEQ as a suitable candidate for a data exchange pilot project.

---

[18] See the NAVTEQ website for more details http://www.nn4d.com/site/global/learn/basics_of_map_data/10_coremap.jsp

[19] See http://mapreporter.navteq.com/dur-web-external/secured/submitDur.do?userType=CONSUMER&language=en

[20] For more information on crowdsourcing, see our report on Addresses (Deliverable 2) and for information on NAVTEQ and crowdsourcing see our report on New Technology (Deliverable 6).

[21] The commercial data providers were unwilling to be specific about the tools or the approaches that they use for conflation or other data evaluation activities.

[22] TFTN-Strategic Planning Group, 2010, Conference Notes –What is TFTN and what does it mean? Charleston, West Virginia: Konag Technology Solutions (unofficial meeting notes), see http://www.slideshare.net/KoniagTFTN/gist-notes-public

Tele Atlas

Tele Atlas was founded in 1984 to meet the potential data needs of the navigation marketplace. The company was acquired by Personal Navigation Device (PND) manufacturer TomTom, one of its largest customers, in 2008. Tele Atlas has been tightly integrated into the TomTom corporate infrastructure and is now an operating division within the company. TomTom has been hurt by a significant downturn in demand and pricing for its products and we believe this factor has negatively influenced the research efforts to update the Tele Atlas database.

The Tele Atlas database provides comprehensive coverage of the streets and roads in the United States and powers in-car navigation systems, Internet mapping and GIS applications. The company's vector data is focused on solving navigation problems and is highly attributed with data relevant to this task (address related administrative data, road function and characteristics, traffic characteristics, etc.). Tele Atlas classifies these attributes as Guide attributes, Find attributes and Display attributes.[23] We consider the Tele Atlas database to be less comprehensive and less accurate than the data available from NAVTEQ, but Tele Atlas seems to be closing this gap rapidly, in part due to their adoption of crowdsourced data as an indicator of areas in their database that need updating. (See Figure 4)

The Tele Atlas production systems has feature level data, but its current database does not contain the rich metadata that was the hallmark of Geographic Data Technologies, a company acquired by Tele Atlas in 2000 to enhance its product in the United States. We were told by a Tele Atlas representative that the company is now substantially extending their metadata and expect to complete this project at the end of the second quarter of 2011. Tele Atlas use of metadata is functional rather than descriptive. The company uses its metadata to permit or restrict certain actions in the editing environment. They also employ the concept of "trumping" that allows the updating of an attribute only if the metadata associated with a new source exceeds the confidence level of the existing source. Tele Atlas data is updated continuously, although released to its licensees on a quarterly basis.

[23] Tele Atlas, 2010, Tele Atlas Multinet, Eindhoven, The Netherlands; Tele Atlas

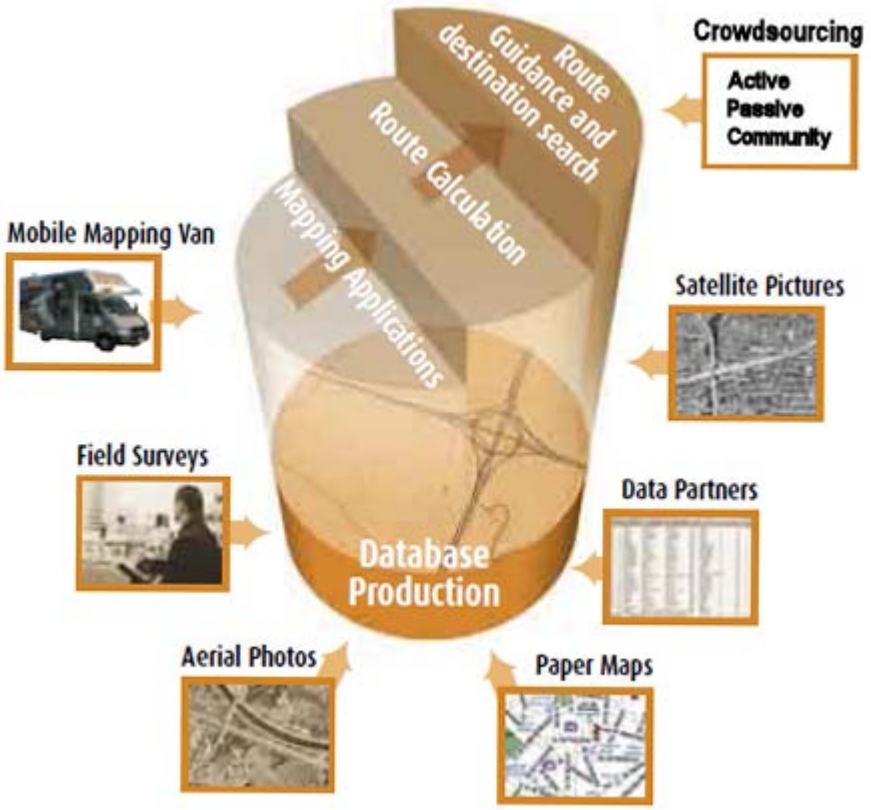


Figure 4. Data sources used by Tele Atlas (after Tele Atlas 2007).

Tele Atlas claims that its map data is compiled from over 50,000 trusted sources. The company employs a complex set of tools to compile their data including the use of data mining, precision aerial and satellite images, paper maps, field surveys, their fleet of mobile mapping vans, and crowdsourcing.[24] Tele Atlas is the most prominent user of crowdsourced data in the world of commercial map suppliers. The company relies on active and passive crowdsourcing provided by users of TomTom PNDs and other products to update and maintain its MultiNet database.

Tele Atlas provides data in a wide variety of formats including shapefile, GDF-AS (ASCII Sequential), GDF-AR (ASCII Relational) and ODF. A representative from Tele Atlas indicated that the company was unlikely to license data that might end up in TIGER (and thus become distributable). Finally, Tele Atlas does not distribute its metadata to its licensees.[25] For these reasons, we do not regard Tele Atlas as a suitable candidate for a data exchange pilot project.

[24] The efforts of Tele Atlas in crowdsourcing are described in our report on Future Technologies (Deliverable 6).
[25] For more information on Tele Atlas, see their website http://www.teleatlas.com/index.htm

CoreLogic

CoreLogic was spun out of the First American Corporation in 2009 and now operates as a separate company.

CoreLogic creates parcel (property) databases that include parcel boundaries, parcel identification numbers, parcel/lot dimension, parcel centroids, property address or situs, street addresses and street names, as well as other data typically used in property location and ownership 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. (See Figure 5.) Coverage of the United States is not complete, as some counties resist providing data to commercial providers, preferring to resell their data to interested parties. Somewhat curiously, another group within CoreLogic provides the same parcel data based through RQGIS, a "WMS–like" hosted solution that is reported to integrate with Microsoft's Virtual Earth. Unfortunately, detailed information on this service is not yet publicly available from CoreLogic.

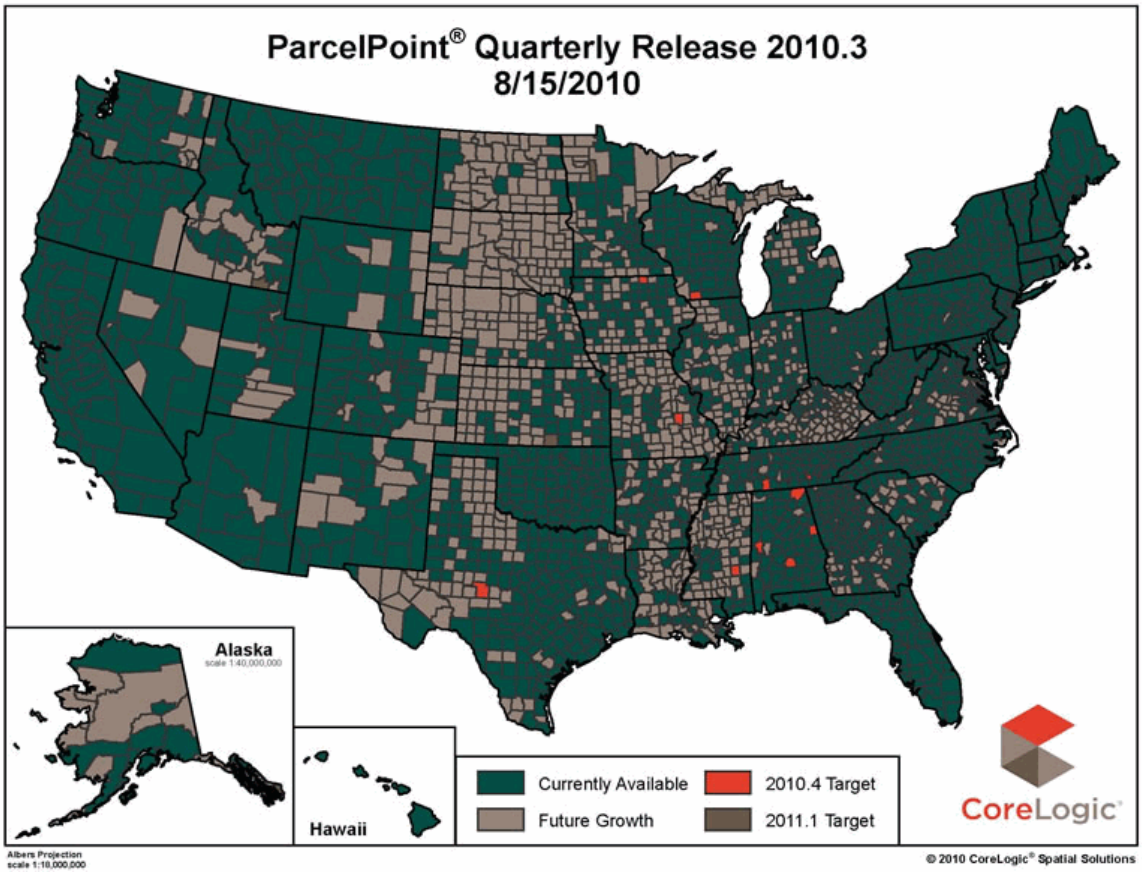


Figure 5. CoreLogic ParcelPoint coverage map, 2010. (Image courtesy of CoreLogic Spatial Solutions Group.)

CoreLogic updates its parcel database in a continuous manner. The company claims to employ a network of stringers (sometimes employees, usually contractors) 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 their core ParcelPoint product.[26,27]

CoreLogic does not utilize crowdsourcing in the collection of its data since the value of their parcel product is to represent the official, legal, property descriptions approved by the local taxation authorities throughout the United States.

The company provides its database in shapefile format.

Representatives of CoreLogic indicated that the company would not be interested in licensing these data in a manner that would allow the parcel data to be redistributed without a license fee. As a consequence, we do not recommend CoreLogic as a candidate for the data exchange pilot.

Google

In 2009, Google stopped using Tele Atlas data to create its map and navigation-based products and changed over to a spatial database compiled by the company using a variety of techniques shown in the figure 6 below.[28]

[26] For information on ParcelPoint see http://www.corelogic.com/Products/ParcelPoint-Technology.aspx

[27] We describe additional CoreLogic products in our Report on Addresses and Addressing (Deliverable 2), where we provide a current coverage map and other relevant information.

[28] Those interested in the procedures adopted by Google might want to review these articles in the TeleMapics blog, Exploring Local, written by one of the authors of this report: http://blog.telemapics.com/?p=218 , http://blog.telemapics.com/?p=222 , http://blog.telemapics.com/?p=226 and http://blog.telemapics.com/?p=238

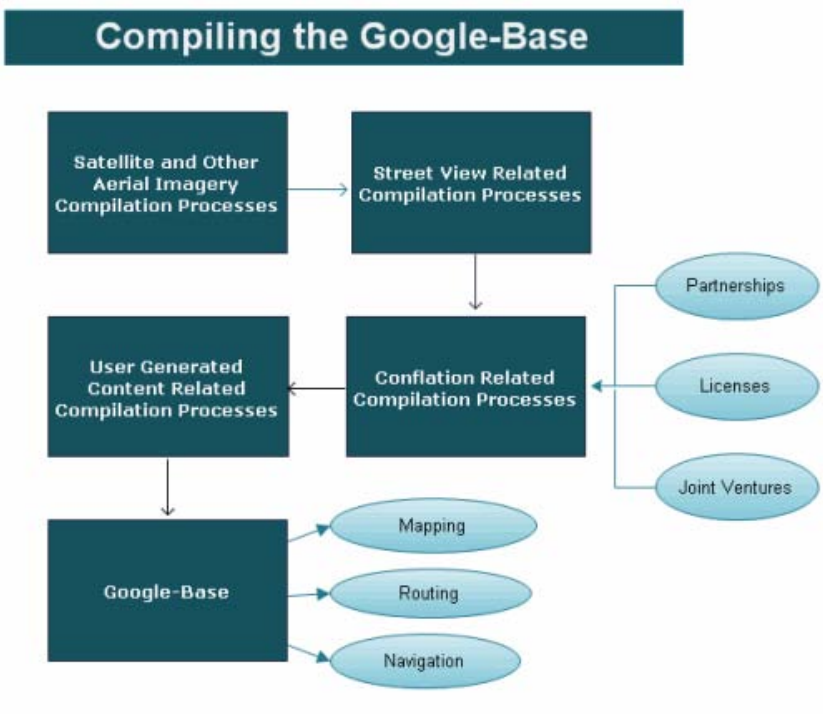


Figure 6. Google's approach to map compilation (image courtesy of TeleMapics, LLC.)

Note that Google is a secretive company that has not yet revealed in any detail the actual working of its compilation process. Figure 6 was created based on knowledge Google's map compilation related capabilities and research on the development and use of these tools. For instance, while it is clear that Google relies on conflation, the exact tools that Google uses for these processes remain unknown. For several years Google worked with deCarta and it is believed that their current process is an even more sophisticated version of the deCarta tools.

Google obviously has a wide range of capabilities, but it is their use of User Generated Content (UGC) from a variety of sources that differentiates their methods from other commercial map database providers. (See Figure 7.)

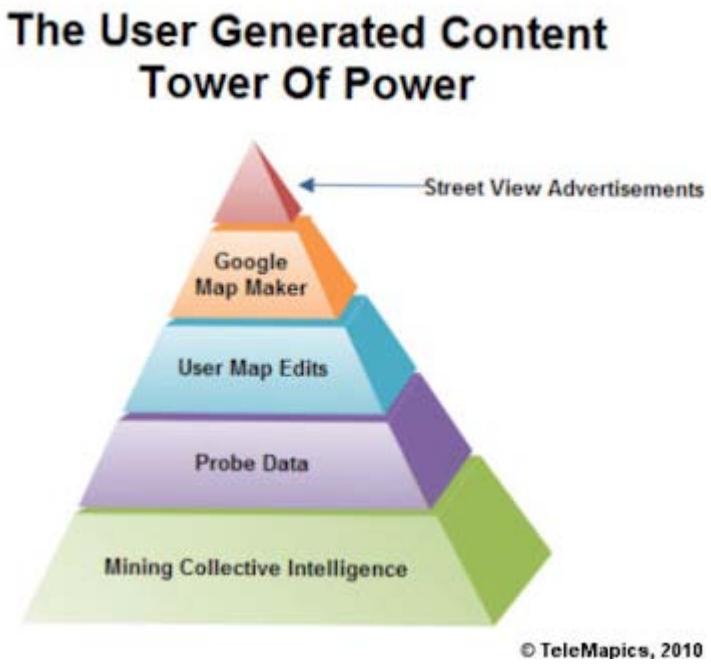


Figure 7. Google's use of UGC spans many sources (image courtesy of TeleMapics LLC).

The company uses the tools developed for its search engine technology for purposes of data mining. Probe data anonymously associated with users is gathered from mobile phones running the Android platform and map edits are suggested directly by users through direct interaction with Google Maps.[29]

Google has revealed little information on its data structure or the characteristics of its map and navigation database. Our belief is that the database contents and management are highly similar to those provided by NAVTEQ and Tele Atlas, but we are unable to offer direct proof of this assertion. What is obvious, however, that Google data could not be used to route or present maps in the manner it does without the normal complement of navigation and display related data attributes.

Google's map database is in a state of continual update, and the updating process utilizes the same model as described above for initial compilation, but we believe that Google relies on crowdsourcing (in terms of map user map suggestions) as the prime ingredient in the updating process. In addition, Google has hired a team of approximately 300 software engineers to help update their maps over the next year.[30]

Google's model for distributing its map and navigation database does not provide for data exchange. Rather the database is available as a public service, as well as a fee-based service known as Google Enterprise, which offers Google Maps API Premier.[31]

[29] Details of this interaction are discussed in crowdsourcing section of our report on Technology (Deliverable 6)
[30] See this article for more details on the Google map update effort http://blog.telemapics.com/?p=282
[31] See this page for more information on Google Enterprise http://www.google.com/enterprise/earthmaps/maps_features.html

Google has, as of this date, shown little interest, in sharing data or participating in data exchange with any party, although, like Census, it is interested in ingesting mapping data from authoritative sources through its Map Content Partners program.[32] The company's interest in mapping is focused on the use of mapping to support its local advertising business and as a method of providing navigable routes to deliver potential customers to the stores of advertisers. As a consequence, Google is not interested in allowing other sources to share or ingest their data. It is our opinion that Google would not be a candidate for a data exchange pilot program.

In our paper on Addresses and Addressing, we described additional categories of sources for commercial address information and direct the interested reader to that report for details on address sources. However, we do not feel that any of the commercial sources of address and map data that we have described here or in our previous reports would be candidates for a data exchange pilot project.

b. Non-profit

OpenStreetMap

OpenStreetMap[33] (OSM) is an example of crowdsourcing applied to spatial data. Initiated in the United Kingdom in 2004 by Steve Coast, the creation of OSM was, in part, a response to the high cost of and restrictive licenses for the use of Ordnance Survey data in that country. The goal behind OSM was to provide free geospatial data that could be edited, updated, and augmented by anyone, as long as the changes that they made to the data were available to everyone under a Creative Commons ShareAlike license.[34] Interest in OSM blossomed and its current coverages of UK, Germany, and France are quite good. Coverage is building in the rest of the world, including the United States.

For its United States coverage, OSM has imported a version of Tiger 2005 that was not uploaded until 2007.[35] While OSM has indicated there have been many edits in some cities, they also indicate that many other cities have not received any editing. Curiously, it is unlikely that a more up-to-date version of TIGER will ever be imported by OSM due to the amount of editing that has occurred since the original upload. OSM has indicated 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".[36]

The OSM data base is comprised of a few simple elements known as nodes, ways (paths), and relations (areas/polygons). It is important to note that OSM has open tagging with no fixed scheme and users are free to use tags that they feel appropriately identify the elements in an

[32] See this article for details on the Map Content Partners program http://maps.google.com/help/maps/mapcontent/
[33] OSM was extensively described in our report on Addresses and Addressing (Deliverable 2), as well as our in our report on Technology (Deliverable 6). We refer the reader interested in OSM and crowdsourcing to these reports.
[34] See http://creativecommons.org/licenses/by-nc-sa/2.0/ . The OSM database is now licensed under the Open Database License (ODbL). For more information see http://www.opendatacommons.org/licenses/odbl/
[35] Wiki OpenStreetMap. See this link http://wiki.openstreetmap.org/wiki/TIGER
[36] Ibid. Wiki OpenStreetMap.

"appropriate" manner. The OSM data model includes objects with position/extent, but lacks class attributes, authoritative ontology of tags and provides no hierarchy. Geometric objects are stored with their history, but the system does not include formal metadata at the feature level, although free text Wiki explanations of the data are sometimes provided.[37] OSM does use metadata, but the intent of the metadata is for tracking the user, user id, OSM visibility, version, change set, time stamp, etc., and not the source of the data (which is assumed, in most cases, to have been collected independently by the contributor from non-copyrighted sources).

OSM data capture is based on a modified probe-collection mechanism. Most data is collected by participants using GPS receivers to record paths as they are driving, bicycling, or walking through local areas. Data from GPS receiver s are uploaded to OSM in GPX (a lightweight XML-based GPS Exchange Format) and there are a number of open source editors, such as GPSBabel, that can be used to convert GPS tracks to GPX from the internal file format of most GPS receivers.[38] The GPS traces are then uploaded online to OSM from the main OSM interface page (http://www.openstreetmap.org/) and can be marked for public sharing or kept private. Once uploaded, the data (in either raw or mapped format) can be downloaded and edited.

The crowdsourcing model used by OSM is "self-healing" over time and based on the notion that as the database is exposed to more eyes, users will correct inaccuracies with which they are personally familiar. As a consequence, the database is under continuous expansion, editing, and updating. Conflation is not a formal part of the crowdsourced model. When a newer source of information is discovered that appears more comprehensive that the data in the database this new information is imported as is and then edited by users to correct various inaccuracies. There is considerable debate in the OSM community about the use of any automated tools, including automated editing tools, since these may change valid information provided by a contributor who has spent a great deal of time researching the area of interest.

OSM data (in the form of nodes, ways, relations, and tags) can be exported in an XML format. Data downloaded in XML format (from the export tab for small areas for from Planet.osm for large areas) can be translated into Keyhole Markup Language (KML), Geography Markup Language (GML) (to convert to shapefiles) or a variety of other formats.[39] As an example of this process, MapQuest is using OSM data in a beta in the United Kingdom.[40] The company updates the maps every ten minutes and updates the map tiles every ten to fifteen minutes, updates the search results to reflect the map changes every 2- 24 hours, and updates changes in directions related to the map changes daily.[41]

Since OSM is an open format, its data could be used for a data exchange pilot, as long as the uses of the data meet the requirements of the OSM license. OSM data, until recently, was licensed under a CC by SA agreement which translates to a creative commons license with a ShareAlike attribution.[42] The ShareAlike attribution requires that if you redistribute the work, you must do

[37] See this presentation http://www.slideshare.net/nichtich/openstreetmap-presentation for an overview of the OSM database.
[38] See http://sourceforge.net/projects/gpsbabel/files/
[39] See http://wiki.openstreetmap.org/wiki/Export
[40] See http://open.mapquest.co.uk/
[41] Personal communication with Antony Pegg, MapQuest developer.
[42] See this page for the creative commons SA 2.0 generic license http://creativecommons.org/licenses/by-sa/2.0/

so under a license similar to the CC by SA. 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.[43] As noted in our report on Addresses and Addressing, OSM is not a great source for addresses and its companion OpenAddress.org is not yet well enough developed to be a useful source of address information for the MAF. We believe that working directly with an OSM extract might not be particularly useful and described the reasons for this assertion in our previous reports. The greatest issue is that OSM is crowdsourced and not authoritative, nor perhaps even "trusted". While we think that an extract of OSM might be a good vehicle for indicating change detection, it is unlikely to be a useful tool for more direct updating of the MTdb. However, OSM concepts and tools might provide an interesting approach to data exchange.

Open Address

The OpenAddress (OA) project was launched in Switzerland in 2007 at the University of Applied Sciences Northwestern Switzerland and expanded in 2009 into a project to create a global directory of addresses. OpenAddress aims to be the OSM of the address community and is described as an Open Source web portal for the management of open, worldwide, geolocated postal addresses. The effort is geared at storing only geolocated postal addresses. The addresses are not associated with persons, companies, points of interest, or any other identifiable entity, other than the location of the address.

OpenAddress is tightly integrated with OSM, but unlike its partner does provide for the use of address attributes[44], including the following:

- Username: the person or entity who created the address. We use it ONLY for statistic purpose. This information can be viewed by all. You can use a pseudo, a group name or whatever you consider as pertinent.
- The street name. Try to use the name used by the person living in this area.
- The house number, if one exists
- The postal code of the city
- The city
- The country of the city
- The quality. If you use the web interface, this will be "Digitized".

- The house name if one exists
- The region. This can be a department, a state, a canton.

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 City of Portland, OR. Address data is available under the Creative Commons 2.0 Share and ShareAlike license and must be associated with OpenAddress web site.

[43] See http://www.opendatacommons.org/licenses/odbl/
[44] See http://code.google.com/p/openaddresses/wiki/UserGuideEn

It is too early in the game to predict with any certainty, but our estimate is that OpenAddress is going to find it difficult to collect enough data to make them a valuable partner. While its initial announcement was greeted with enthusiasm, the site has recorded few major data partners since that time. In addition, the OA system of recording addresses lacks standardization and verifying the actual existence of the contributed addresses appears to be a weak link, at least at this time.[45] We do not recommend considering OpenAddress as a candidate for the data exchange project, although the OpenAddress model is an interesting method to capture addresses.

Summary

Although the commercial providers may have comprehensive address and spatial databases, it is unlikely that these companies would be willing to license their data for re-use in TIGER. As we noted in our report on Address and Addressing some commercial firms might be willing to license address data due to the potential protection Title 13 would provide them in terms of the redistribution of their data. However, we believe that the most valuable and up-to-date address and spatial data are provided by state and local sources and that these sources would be the best candidates for a data exchange pilot project. For that reason, the remainder of this section is focused on state and local sources of address and spatial data that we feel are potential candidates for data exchange.

c. Providers of Federal, State and/or Local Spatial Data

The objective of this section of our report is to provide an update on the public geospatial data community. It covers the current situation within the Federal Government, a discussion of the relationship between states and Federal Government through the NSGIC/FGDC 50 states initiative and a review of some exemplary systems at the county, regional and state levels. It highlights trends in technology such as cloud computing that have greatly facilitated the access to and sharing of data relevant to the GSS Initiative. The overarching conclusion is that local governments have embraced GIS as the way to conduct business and are increasingly sharing these data. As a result, they routinely maintain highly accurate (often higher than those specified for MTdb) representations of roads, buildings, address and parcel data. While some local governments continue to restrict access to their data resources, the trend is to share them with state GIS coordinators. This trend is strongly supported through regulatory and financial incentives designed to support emergency response. Many local governments are taking advantage of web based technology to provide easy access that allows their citizens to display, query and even download current data. Agencies with restrictive data sharing policies also provide mapping services that enable users to locate and view their data holding. Even these limited types of services could assist the GSS Initiative by providing the opportunity to inspect or evaluate the location of potential living quarters from an office environment, rather than requiring field work. For example, the Richland County, South Carolina Geographic Information System group, which elects not to share its data with the state government, provides state of the art web viewing tools for its citizen users. (Figure 8.)

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

With the aid of the NSGIC/FGDC 50 states initiative, many states have taken an active role in consolidating and standardizing street centerlines, address points, and parcel data. For example, the Ohio OGRIP program has an extremely active local / state partnership to create the Location Based Response System (LBRS) that helps fund a consistent state-wide set of street centerlines with a positional accuracy of about one meter. Furthermore, this is a very current data base with one county maintaining its data on a daily basis with updates published to an ftp server on a weekly basis. As the GSS Initiative begins, the Geography Division should make every effort strengthen its partnerships with state agencies and to acquire, on a regular schedule, the data collected and published by them. Since an increasing number of these local and state partners are willing to "throw the goodies over the fence" the Geography Division should develop efficient tools to harvest the data, extract changes, and stage the results for updating the MTdb. From a historical perspective, this represents a reversal of roles from earlier decades when "throwing the goodies over the fence" was the de facto distribution policy of the UGSG and Census Bureau. (See: *Spatial Data Needs: The Future of the National Mapping Program* (NRC, 1990))

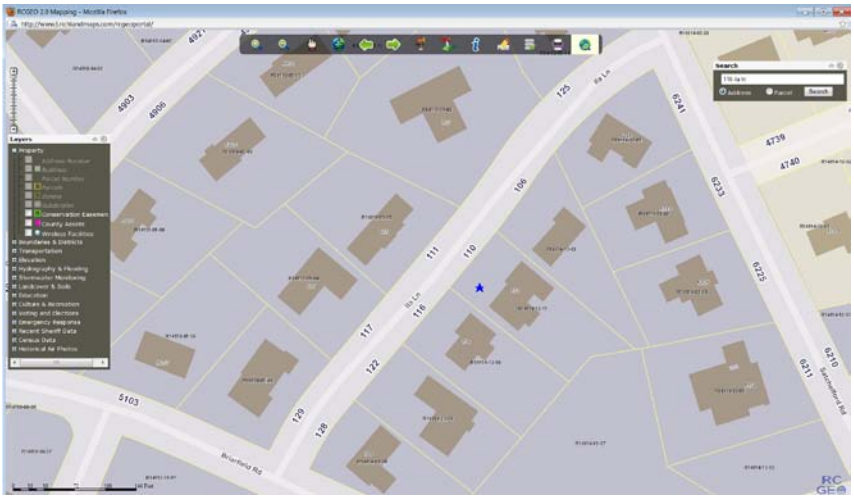


Figure 8. Richland County, SC Web mapping site (http://www3.richlandmaps.com/rcgeoportal/)

1.b.2 The Current Federal Environment

From an institutional perspective Federal Geospatial activities are subject to oversight under OMB circular A-16. The basic governance structure is described as follows

> "The Circular affirms and describes the National Spatial Data Infrastructure (NSDI) as the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data. The Circular describes the management and reporting requirements of federal agencies in the acquisition, maintenance, distribution, use, and preservation of spatial data by the Federal Government. The Circular establishes the FGDC as the interagency coordinating body for

> NSDI-related activities, chaired by the Secretary of the Interior with the Deputy Director for Management, Office of Management and Budget (OMB) as Vice-Chair."[46]

Milo Robinson (2008) provides a summary of the evolution of A-16 and the Federal Geographic Data Committee (FGDC) (Table 1). He offers a useful summary:

> "Many of the coordination policies that we have today have well-established roots. While the technology has markedly improved, many of the policy goals are the same as in the past, including standards, having an authoritative geospatial information source, and having broad participation in the coordination process. Coordination of geographic data is important and the need for such coordination is well established. However, ways to improve the organizational effectiveness of often stovepiped organizational structures within the Federal Government are needed, as are ways to stimulate cross agency coordination and collaboration. History has shown that improving coordination and fostering organizational change that will benefit the development of the NSDI is difficult. If it was easy it would have been done 100 years ago."

[46] See (http://www.whitehouse.gov/omb/circulars_a016_rev#1)

Table 1. Evolution of OMD Circular A-16 (Robinson, 2008)

Executive Order 9094	1942	Abolished Board of Surveys and Maps and authorized Director of OMB to perform the functions of the Board.
OMB Circular A-16	1953	Described responsibilities of Federal agencies with regard to the coordination of surveying and mapping activities.
A-16 Exhibits A,B,C, & D	1953-1964	Attachments to A-16 that outlined programming and operations for specific activities (Topographic Mapping, National Atlas, Geodetic Control, International Boundaries). Some Exhibits were revised.
A-16 Revised	1967	Better described responsibilities of Federal Agencies to provide leadership and coordination.
OMB Memo 83-12	1983	Established coordination of Federal digital cartographic data programs.
A-16 Revised	1990	Established Federal Geographic Data Committee and expanded Circular to include more programs
Executive Order 12906	1994	Established the National Spatial Data Infrastructure (NSDI).
A-16 Revised	2002	Strengthened coordination responsibilities of Federal agencies and incorporated NSDI into the Circular.
OMB Memo M-06-07	2006	Required agencies to designate Senior Agency Official for Geospatial Information (SAOGI) at Assistant Secretary-level.

Table 1. Evolution of OMB Circular A-16

The FGDC is a multi-tiered bureaucracy that focuses on thematic subcommittees (Figure 9). From the perspective of the GSS Initiative, it is important to understand the federal stewardship of roads, governmental units, and addresses. The Department of Transportation (DOT) is the steward of base transportation, railways, roads, air, and transit. The US Army Corp of Engineers oversees inland waterways. DOT has worked with other federal partners to create content standards for base, rail, roads, and transit. DOT and the Bureau have been discussing the role of TIGER as a component of the federal data base (Lewis, Steve 2010), but it appears that little progress has been made on this opportunity.

The Bureau is an active leader of the FGDC Cultural and Demographic Statistics Working Group. In that role it has helped formulate standards for Governmental Unit and Other Geographic Area Boundaries. The standard codifies important definitions and responsibilities. It also includes an interesting set of topological relationships that impact boundary polygons.

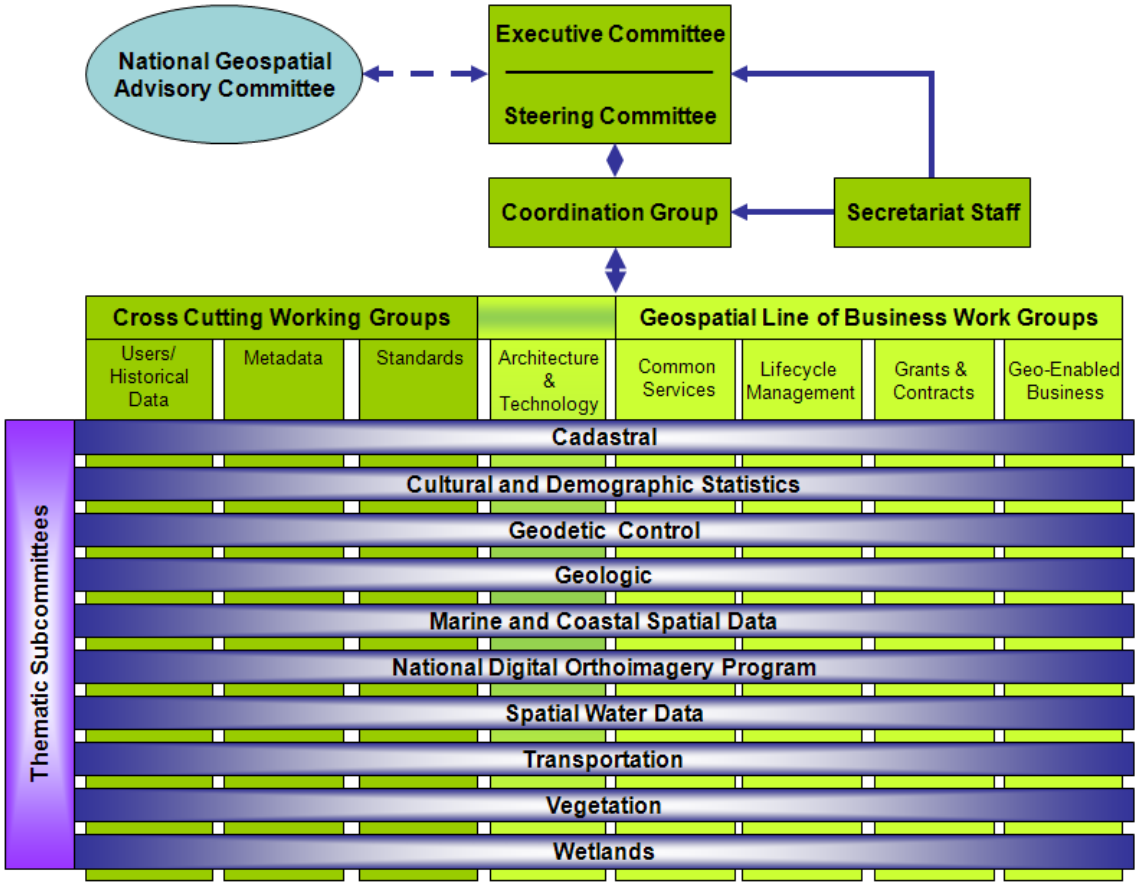


Figure: Overview of the structure of the various components of the FGDC

Figure 9. Organizational Structure of the FGDC (fgdc.gov)

The Bureau has also provided leadership in the development of the 555 page standard for *United States Thoroughfare, Landmark, and Postal Address Data* which was published in final draft version in November. This standard was discussed in detail in our report *Reporting the State and Anticipated Future Directions of Addresses and Addressing.* When the standard is accepted, the Bureau will be the official steward for address information.

OMB Circular A-16 Supplemental Guidance

In November, Vivek Kundra, The Federal Chief Information Officer issued new A-16 Supplemental Guidance. In his cover letter Kundra stated:

> "Data-management, and particularly geospatial data-management, is one of the essential components for addressing the management of the business of government and for supporting the effective and economical use of tax dollars. It is, however, susceptible to constant renewal, information quality, and information management challenges. A portfolio-centric model cures the single agency, stovepipe model by applying consistent policy, improved organization, better governance, and understanding of the public to deliver outstanding results." http://www.whitehouse.gov/sites/default/files/omb/memoranda/2011/m11-03.pdf

The major changes in A-16 emphasize a Portfolio Management of National Geospatial Data Assets (NGDA) that includes:

Inventory – identify existing geospatial datasets that may be considered NGDAs

Select - determine if the datasets are NGDA Datasets and if they should become part of the NGDA Portfolio

Manage – develop and manage NGDA Datasets per the Supplemental Guidance

Evaluate/monitor – manage NGDA Datasets according to NGDA Theme plan, milestones and performance measures

Identify and execute – develop NGDA Theme and NGDA Dataset development and funding and priorities

The guidance also expands the roles and responsibilities for participants in the stewardship of each NGDA theme to include:

Executive NGDA Theme Champion

NGDA Theme Lead

NGDA Dataset Manager

Data Steward

The FGDC has announced (Siderelis, Karen and DeLoatch, Ivan 2010) a schedule for implementing the charges of the new guidance. These include the following objectives before the end of April 2011:

The FGDC will make final adjustments to NGDA themes

Associate the current inventory of National Geospatial Data Assets Themes

Grow the NGDA inventory

The FGDC will distribute guidance to agencies on how to comply with the Supplemental Guidance

Agencies will send their timelines for complying with the Supplemental Guidance to the FGDC

The FGDC will select a portfolio management pilot for the FY13 budget process

The FGDC will finalize the maturity model/other scorecard content that will be used to evaluate NGDAs.

The National Geospatial Advisory Committee (NGAC)

The National Geospatial Advisory Committee (NGAC) is a Federal Advisory Committee sponsored by the Department of the Interior under the Federal Advisory Committee Act. The NGAC reports to the Chair of the Federal Geographic Data Committee. The mission of the NGAC is "To provide strategies regarding the creation, management and dissemination of cohesive geospatial data, information and knowledge to enable commercial, academic, and nonprofit organizations and all levels of government…" NGAC is an active group of 28 members who meet quarterly to review a wide range of federal geospatial programs and policies.

Geospatial Platform

A current topic of interest in the Federal geospatial community is the Geospatial Platform that was included in the President's Budget for Fiscal Year 2011. That budget includes the following statement:

> "In 2010 and 2011, federal data managers for geospatial data will move to a portfolio management approach, creating a Geospatial Platform to support Geospatial One-Stop, place-based initiatives, and other potential future programs. This transformation will be facilitated by improving the governance framework to address the requirements of State, local and tribal agencies, Administration policy, and agency mission objectives. Investments will be prioritized based on business needs. The Geospatial Platform will explore opportunities for increased collaboration with Data.gov, with an emphasis on reuse of architectural standards and technology, ultimately increasing access to geospatial data." President's Budget, Fiscal Year 2011 (FGDC, 2010)

Over the past several months the FGDC, with lead from the Department of Commerce, has worked to define the geospatial platform. Many documents and presentations about the Platform can be found at http://www.geoplatform.gov/

The key components of the proposed platform are:

Common Data, Services, and Applications

Shared Infrastructure

Geospatial Segment Architecture

Collaborative Governance

Portfolio Management

A conceptual diagram of the Geospatial Platform emphasizes the need for streamlined access to resources (Figure10.)

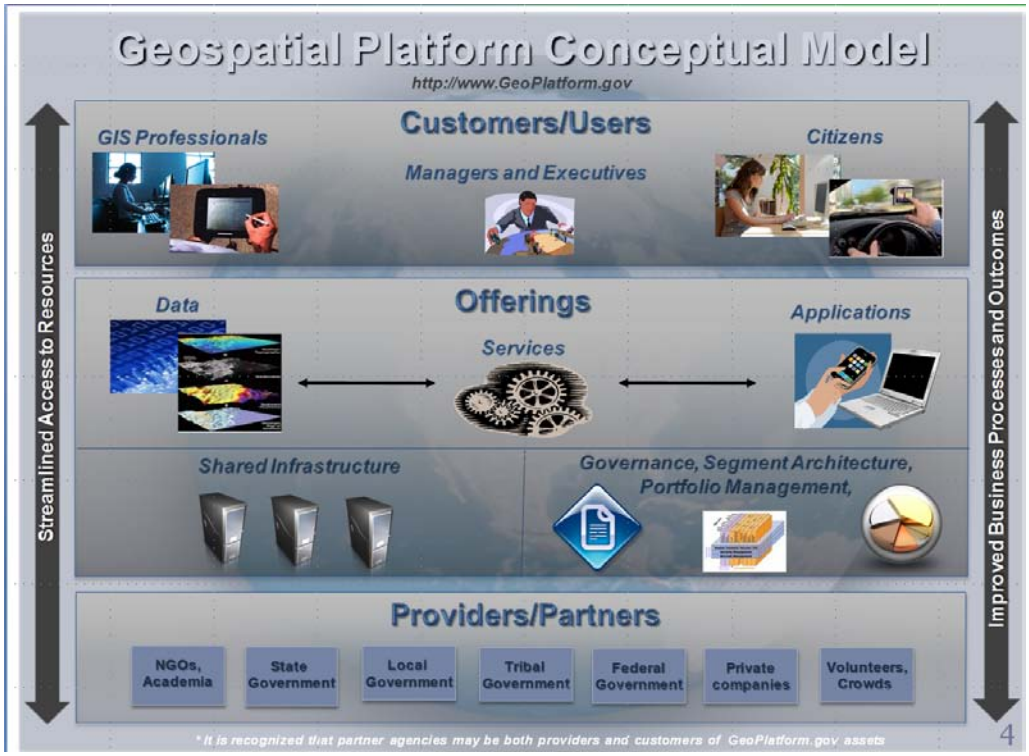


Figure 10. Geospatial Platform Conceptual Model (Siderelis, 2010)

The current debate focuses on determining which agency will be the managing partner and how the platform will be funded. The approach to establishment of the platform is outlined in the *Modernization Roadmap for the Geospatial Platform.* From the perspective of the GSS Initiative the following specific action items were listed in roadmap:

Time Box 1 (Jul – Sep 2010)

DoD Geospatial Visualization Enterprise Solution (GV-ES)

Homeland Security Infrastructure Protection (HSIP) data

Time Box 2 (Oct – Dec 2010)

Common geocoding service(s) requirements

National database pilot projects

National Urban Change Indicator (NUCI) dataset

Time Box 3 (Jan – Sep 2011)

Common geocoding service

Permanent Managing Partner

Final structure for intergovernmental body

National study by trusted institution

Of particular interest would be specific action to create the National Urban Change Indicator (NUCI) dataset which is a Landsat based change detection program offered by MDA Federal Systems with current support from The Department of Homeland Security. More details about this program can be found in our report *Identifying the Current State and Anticipated Future Direction of Potentially Useful Developing Technologies.* The other proposed program would be a Common Federal Geocoding Service. Since geocoding is a vital component of all aspects of Census operations it is useful to note that HUD operates a geocoding service.[47] The service provides the following functions:

Agency-wide services for address validation, standardization, and matching (3.5+ Million records monthly on-line volume experienced)

Options for batch, on-line, or point-of-entry address validation and geocoding

High data quality, timely information, cost-effective, and tailored to customer needs

Assigns geographic codes to an address:

Census Codes: Congressional District, County, Place, Tract, Block

HUD Codes: Qualified Census Tracts, HUD Underserved Areas, Revitalization Areas

Postal Codes: ZIP+4, Delivery Point Bar Code, Address Type

Geographic Coordinates: Lat/Long

Return Codes: Data Quality/ Confidence Levels

It is also noteworthy that a table of potential opportunities for common data, services and applications highlights the role of The National Map for public streets and roads, as well as address points and a geocoding service. (Table 2.)

[47] For more details see (http://www.fgdc.gov/geospatial-lob/documents/hud-geocoding-factsheet.pdf)

Table 2 Potential Opportunities for common data, services and applications (FGDC, 2010)

Table A.1. Potential opportunities for common data, services and applications

Geospatial Data	Citation Reference	Immediate	Mid Term
Imagery	TNM, NGAC, LLR	TNM-Imagery	IFTN
Elevation/LiDAR	TNM, CSWG, NGAC, LLR	TNM-NED	NEEP
Public Streets and Roads	TNM,	TIGER	
Surface Water	TNM,	TNM-NHD	
Civil Boundaries	TNM,		
Parcels	TNM, NGAC, LLR		
Land Cover	TNM,		NLI
Geographic Names	TNM,	TNM-GNIS	
Public Land Survey System	TNM,		
Vertical and Horizontal Control	TNM,		
Federal and Native American Land Boundaries	TNM,		
Address Points	TNM, CSWG		
National Wetlands Dataset			USFWS
Geospatial Platform Services			
Cloud Computing	CSWG		
GPS receivers	CSWG		
Geocoding Service	CSWG, TNM		X
Instruction/Guidance/How-to manual		How-to manual	
Common Data Access URL			X
Remotely Sensed Land Cover Change Det			X
Strategy for Service Discovery			X
Geospatial Applications			
Land Cover Change	CSWG,		
Map Viewer/Visualization	CSWG, TNM	Palanterra™	
Meta Data Editor	CSWG		
Map Viewer for Mobile Platform	TNM		X
Coastal Communities			X
Tightening Mortgage Practices			X
Cities in Transition			X

(TNM) The National Map Customer Requirements: Findings from Interviews and Surveys, OFR 2009-1222, U.S. Department of the Interior, U.S. Geological Survey, December, 2009
(CSWG) Common Services Work Group Survey of Federal Agency Geospatial Requirements, Summary of Results, February, 2010
(NGAC) National Geospatial Advisory Committee, Summary of key decisions/recommendations, November, 2009
(LLR) National Geospatial Advisory Committee Messaging Committee Lessons Learned and Recommendations

General Services Administration (GSA)

Historically, the FGDC has been under the general guidance of the Department of Interior. However, over the past few years the General Services Administration (GSA) has taken an active role in E-Government, cloud computing, and even the geospatial platform. The Bureau has been a major contributor to the E-Government GEODATA catalog. At the December meeting of NGAC, Dr. David L. McClure, Ph.D., Associate Administrator Citizen Services and Innovative Technologies of GSA, elaborated on the concept of Creating a Cloud Based Geospatial Platform. He also specifically commented on the potential role of GSA as the Managing Partner of the Geospatial Platform. His comments focused on how cloud computing could and may be required to serve as the base for the geospatial platform. In fact, it appears that there is a "cloud first" policy emerging across the Federal Government. (See Figure 11.) The following statement summaries the goal:

> **"**Moving to consolidate data centers and deploy cloud computing technology to reduce IT, real estate, and energy costs**.** Already, we have implemented a zero-growth policy on federal data centers, we are working with agencies to review their plans, and we are on

track to meet the president's objective to consolidate and significantly reduce the number of data centers within five years”.[48]

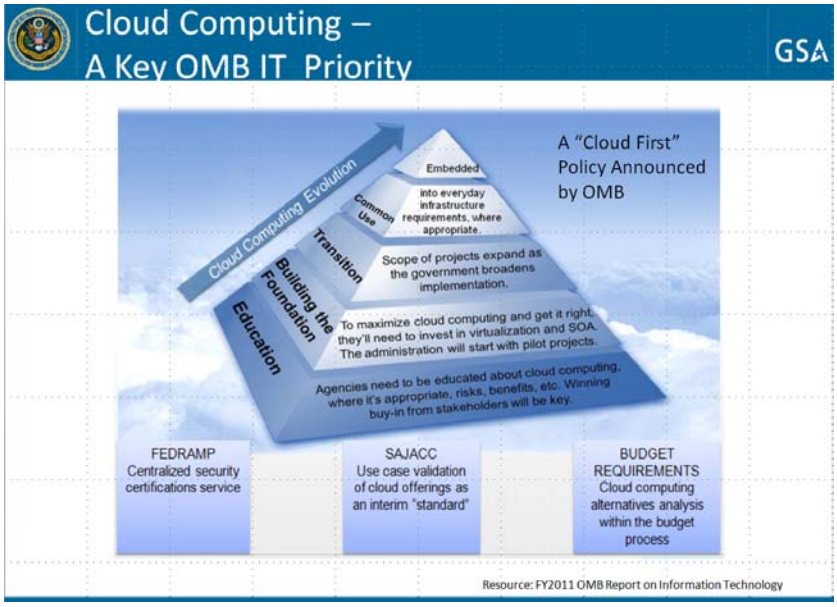


Figure 11. Cloud First pyramid (McClure, David. 2010)

The National Institute of Standards and Technology (NIST) defines cloud offerings as having five essential characteristics:

On-demand Self-Service

Ubiquitous Network Access

Location Independent Resource Pooling

Rapid Elasticity

Measured Service

David McClure (2010) cited the following Benefits of this shared services model include:

- Faster acquisition and deployment of computing resources
- Economies of scale through easier sharing of IT services across organizations
- Lower capital equipment expenditures

He also presented a conceptual model for a Geospatial Platform as a Service (Figure 12)

[48] See (http://www.whitehouse.gov/blog/2010/10/25/closing-it-gap-update)

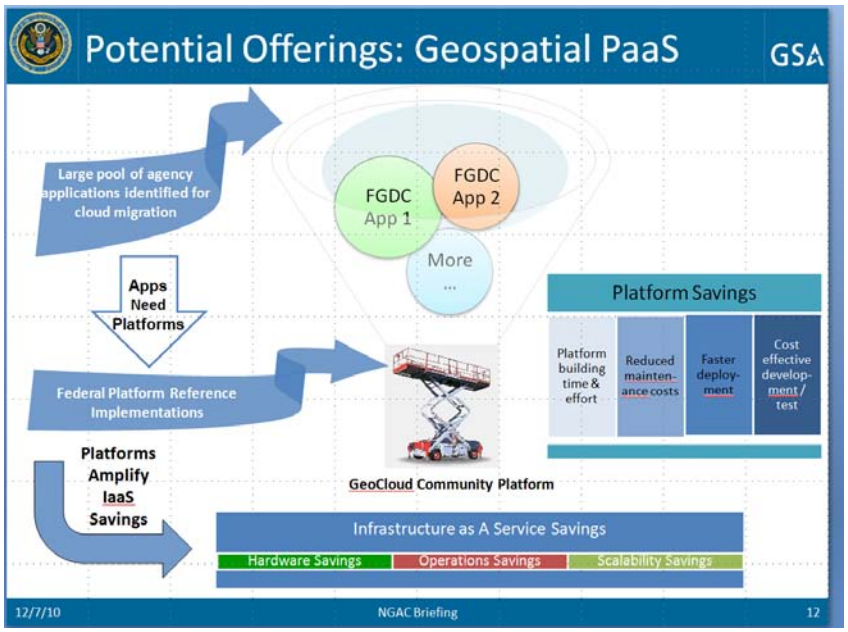


Figure 12. Geospatial Platform as a Service Model (McClure, David. 2010)

David McClure identified twelve initial target applications for the community platforms. It is interesting to note that hosting and distributing Census TIGER shapefiles is one of these experiments. (See Figure 13.) A member of the FGDC staff (Nebert, 2010) has been supervising these cloud experiments. He presented the following GeoCloud Status and Plans, as of December 2010:

- Providing login credentials (secure key pairs) to AWS and its management interface to individuals using AWS
- Initiating image builds of OS plus base software dependencies for CentOS (RedHat) and scripting the geospatial module activation
- Scripting Windows 2008 Server using hardening guidelines used by GSA and other agencies with ESRI ArcGIS Server as image baseline
- Installation and monitoring of the deployment environment with agency-facing Web services (December-February)
- Initial public-facing Web Services (January).

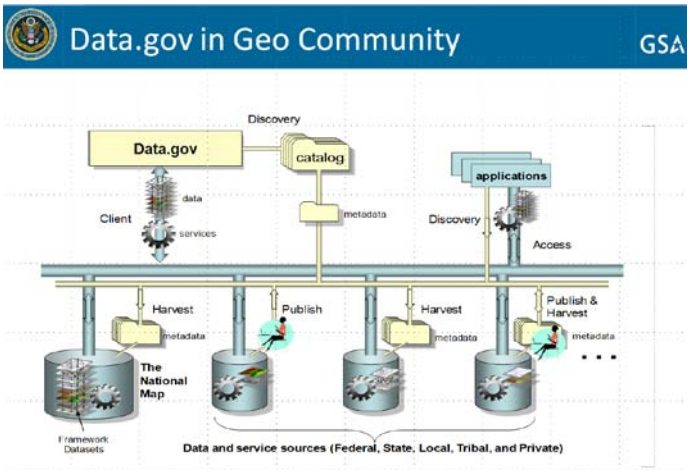


Figure 13 Data.Gov configuration (McClure, David. 2010)

Summary

In the current federal environment the Geospatial Platform is a significant new initiative stemming directly from the Executive Branch. However, no decision has been made regarding the managing partner or the funding for the Geospatial Platform. The Census Bureau has been a major player in FGDC coordination efforts by taking the responsibility to complete the address and governmental units standards. At the same time, it is unclear what role TIGER will play in the DOT stewardship of transportation. The Bureau has also taken a pioneering position with respect to participation in the Data.gov and the current GSA/FGDC experiments to utilize cloud computing as a platform. The supplemental guidance for OMB circular A-16 could result in a review of the data stewardship responsibilities for data content. The Bureau should carefully consider its role with respect to National Geospatial Data Asset of transportation. While DOT may be the Executive NGDA Theme Champion there are additional roles as the Theme Lead, Dataset Manager and Data Steward. The Census Bureau should also be cognizant of the effort to establish a federal geocoding service either within HUD or as part of the USGS National Map.

1.b.3 NSGIC / FGDC 50 States initiative

In 2005 the National States Geographic Information Council (NSGIC) worked with the FGDC to create the Fifty States Initiative. It was designed to bring all public and private stakeholders together in statewide GIS coordination bodies that would help to form effective partnerships and lasting relationships. The initiative is now managed by the FGDC which issues Cooperative Agreement Program (CAP) grants each year that enable the states to improve their coordination mechanisms. This initiative is one of twelve planning activities that were begun as part of the FGDC *Future Directions* strategic planning process (NSGIC, 2009). The Action Plan identifies the following list of activities that can be used to measure the successful implementation of the NSDI by statewide coordination councils/authorities:

1. Local, state, tribal and federal agencies have data sharing agreements in place unless they routinely provide data in the public domain.

2. A published list of local, state, tribal and federal data stewards and integrators for each of the framework layers is available.

3. Local, state and tribal framework data are being posted to the statewide clearinghouse or otherwise being made available through OGC interfaces.

4. Local, state and tribal data producers use the GIS Inventory to create metadata for data holdings and post it to the Geospatial One Stop (GOS) Portal.

5. A functioning clearinghouse or appropriate inventory tool is available to all interested sectors in a state.

6. Local, state and tribal agencies participate in The National Map.

7. Local, state and tribal agencies adopt and incorporate OGC, FGDC, ANSI and ISO standards as appropriate.

Cooperative Agreement Program (CAP) grant program

Critical components of the 50 States Initiative are Cooperative Agreement Program (CAP) grants that help fund efforts to improve data sharing – which is a major goal of the GSS Initiative. According to NSGIC (Craig, 2010) 42 states received CAP grants between 2006 and 2009. According to FGDC (Urban-Mathieu, Brigetta 2010) 49 projects of up to $50,000 were funded between 2006 and 2010. A total of $2,275,000 was provided with at least 50% as an in-kind match. These grants have focused on strategic and business plan development and implementation. A significant outcome of the Fifty States Initiative has been the creation of state GIS data portals (Table 3) that provide a point of contact for discovery and access to GIS data. It is also significant that these organizations have pushed to develop data that meets state and often national standards and document these data with FGDC compliant metadata. While these efforts represent huge steps forward it must be noted that there is a wide range of approaches to data sharing policies and programs.

Table 3. List of State Data Portals

List of State GIS Data Portals	
State	Data Portal
Alabama	http://portal.gsa.state.al.us/Portal/index.jsp
Alaska	http://www.asgdc.state.ak.us/
Arizona	http://agic.az.gov/ (Not in NSGIC 2009 Survey)
Arkansas	http://www.geostor.arkansas.gov

California	http://gis.ca.gov/catalog
Colorado	http://coloradogis.nsm.du.edu
Connecticut	http://www.ct.gov/gis
Delaware	http://datamil.delaware.gov/geonetwork/srv/en/main.home (Not in NSGIC 2009 Survey)
District of Columbia	http://data.octo.dc.gov
Florida	http://www.dep.state.fl.us/gis/ (Not in NSGIC 2009 Survey)
Georgia	http://data.georgiaspatial.org
Hawaii	http://hawaii.gov/dbedt/gis/ (Not in NSGIC 2009 survey)
Idaho	http://www.insideidaho.org
Illinois	http://www.isgs.uiuc.edu/nsdihome
Indiana	http://indianamap.org
Iowa	http://www.iowagis.org
Kansas	http://www.kansasgis.org
Kentucky	http://kygeonet.ky.gov/
Louisiana	http://doa.louisiana.gov/lgisc/
Maine	http://megis.maine.gov
Maryland	http://www.marylandgis.net
Massachusetts	http://www.michigan.gov/csstp
Michigan	http://www.michigan.gov/csstp
Minnesota	http://www.lmic.state.mn.us/chouse/index.html
Mississippi	http://www.gis.ms.gov/Portal/ (Not in NSGIC 2009 Survey)
Missouri	http://msdisweb.missouri.edu
Montana	http://gisportal.mt.gov/Portal/
Nebraska	http://www.dnr.ne.gov/databank/geospatial.html
Nevada	http://keck.library.unr.edu/

New Hampshire	http://www.granit.sr.unh.edu/
New Jersey	http://njgin.state.nj.us
New Mexico	http://rgis.unm.edu/
New York	http://www.nysgis.state.ny.us/
North Carolina	http://www.nconemap.net/
North Dakota	http://www.nd.gov/gis/
Ohio	http://ogrip.oit.ohio.gov/ServicesData/OhioMetadataServer/tabid/88/Default.aspx
Oklahoma	http://geo.ou.edu/DataFrame.htm
Oregon	http://gis.oregon.gov/DAS/IRMD/GEO/alphalist.shtml
Pennsylvania	http://www.pasda.psu.edu
Rhode Island	http://www.edc.uri.edu/rigis
South Carolina	http://gis.sc.gov/data.html
South Dakota	http://arcgis.sd.gov/IMS/sdgis/Data.aspx
Tennessee	http://www.tngis.org/ (Not in NSGIC 2009 Survey)
Texas	http://www.tnris.state.tx.us
Utah	http://agrc.utah.gov/sgid
Vermont	http://www.vcgi.org/
Virginia	http://gisdata.virginia.gov/Portal/
Washington	http://metadata.gis.washington.edu/
West Virginia	http://wvgis.wvu.edu
Wisconsin	http://sco.wisc.edu/wisclinc/index.php
Wyoming	http://wgcs.state.wy.us/ (Not in NSGIC 2009 Survey)
Source: National States Geographic Information Council 2009 State Summary - except as noted http://www.gisinventory.net/summaries/	

NSGIC GIS Inventory

The GIS Inventory (formerly called Ramona) was implemented in 2006 with grant funds from NOAA. Ramona was developed to inventory the GIS data holdings of tribal, state and local governments, and their partners. It provides one consistent platform for the nation that is designed to work in concert with the Geospatial One Stop (GOS) Portal, but it can also be customized for use by each state. It provides a simple tool to enable users to create their own profile, including information on their organization, systems, and data distribution and management policies. It automatically creates "starter" metadata data that facilitates data discovery. It is important to note that one of the stated objectives of the GIS Inventory is to support the Bureau. As noted in the following:

> "Using Ramona will help state, local and tribal government agencies reduce the number of data inventory requests made by federal agencies. For example the Census Bureau routinely contacts agencies to obtain information for the TIGER Enhancement." (See: http://gisinventory.net/about.html)

The GIS Inventory has great potential for the GSS Initiative. Ideally, it would be possible for the Geography Division to discover which public agencies have updated their street centerlines. For example, a search of street centerlines completed in 2009 returned 39 entries (Figure 14). A link is provided to each of these. For example, the starter metadata indicates that the Iowa Department of Transportation completed a set of "Roads/Street Centerlines - Road Centerlines 2009, at a scale of 1:600 (1in=50ft). For further information examine: http://www.iowadot.gov/gis".

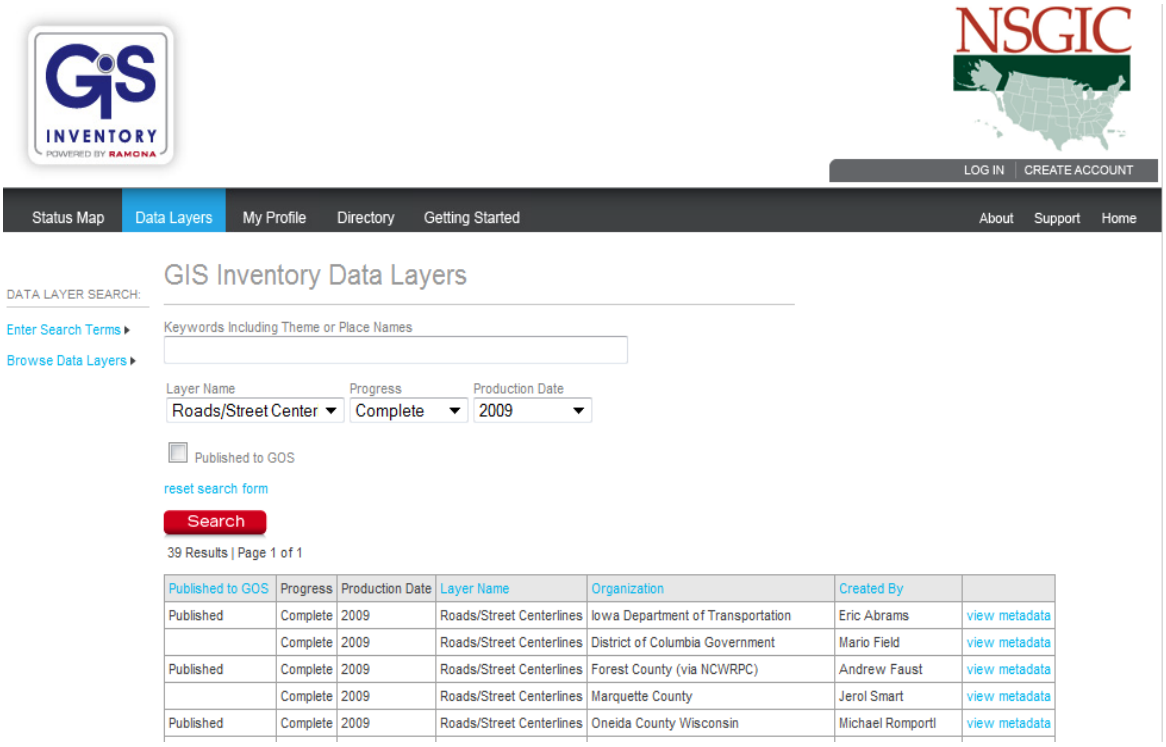


Figure 14 GIS Inventory example for roads completed in 2009 (http://gisinventory.net)

1.b.4 State Coordination of roads and addresses

While it is not the intent of this report to provide an in-depth analysis of all state data portal efforts, it is noted that state wide standardization of street centerlines and address points is an important function of public safely and emergency response systems. As we noted in our report *Reporting the State and Anticipated Future Directions of Addresses and Addressing* twenty two states are coordinating such data at the state level (Table 4). A good example of such a program is Ohio's The Location Based Response System (LBRS). This program is described as:

> The Location Based Response System (LBRS) is an initiative of the Ohio Geographically Referenced Information Program (OGRIP). The LBRS establishes partnerships between State and County government for the creation of spatially accurate street centerlines with address ranges and field verified site-specific address locations. Funding to support the development of LBRS compliant systems is available to counties through a Memorandum of Agreement (MOA) that establishes roles and responsibilities for program participation. Participating counties provide project management and QA/QC on road names, addresses, etc., to develop data that is compatible with the state's legacy roadway inventory. [49]

[49] See (http://ogrip.oit.ohio.gov/ProjectsInitiatives/LBRS.aspx)

Table 4. Statewide address databases. (The data in this table were provided, in part, by William Burgess, Washington Liaison, National States Geographic Council.)

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

An example of the Ohio street centerlines attribute table is provided as table 5, below.[50]

[50] See http://gis3.oit.ohio.gov/LBRS/_downloads/docs/LBRS%20Specifications%203%20FINAL.pdf

Table 5. Ohio LBRS street centerline specifications

2.2.2 Dataset Name

Name	Type
STREET_CENTERLINES	Line Layer

2.2.3 Attribute Definition

Field Name	Length	Type	Description	Standard
SEG_ID	20	Numeric	Unique identifier of the street segment used to join to the Street Alias Table and Site Address tables	
STR_PRE	4	Alpha	The primary street prefix	USPS Pub 28
STR_NAME	30	Alpha	The primary street name. Example: Main, 1ST, Lake Forest, Second	USPS Pub 28
STR_TYPE	4	Alpha	The primary street type. Example: ALY, AVE, BLVD, DR, LN, PKWY, RD, PL, ST, CIR	USPS Pub 28
STR_SUFFIX	2	Alpha	The primary street directional suffix. Example: N, S, E, W, NW, NE, SW, SE	USPS Pub 28
L_ADD_FROM	5	Alpha	Potential address range left side at the segment's start node	USPS Pub 28
L_ADD_TO	5	Alpha	Potential address range left side at the segment's end node	USPS Pub 28
R_ADD_FROM	5	Alpha	Potential address range right side at the segment's start node	USPS Pub 28
R_ADD_TO	5	Alpha	Potential address range right side at the segment's end node	USPS Pub 28
L_TWP	18	Alpha	The Township name on the left side of the street segment	ODOT, FIPS
R_TWP	18	Alpha	The township name on the right side of the street segment	ODOT, FIPS
L_CITY	30	Alpha	The municipality name on the left side of the street segment	OSOS
R_CITY	30	Alpha	The municipality name on the right side of the street segment	OSOS
L_COUNTY	3	Alpha	The 3 character county code on the left side of the street segment. ODOT designation	ODOT
R_COUNTY	3	Alpha	The 3 character county code on the right side of the street segment. ODOT designation	ODOT
L_STATE	2	Alpha	The State name abbreviation on the left side of the street segment. Example: OH, MI, IN	USPS Pub 28
R_STATE	2	Alpha	The State name abbreviation on the right side of the street segment. Example: OH, MI, IN	USPS Pub 28
L_ZIP	5	Alpha	The Zip Code on the left side of the street segment	USPS Pub 28
R_ZIP	5	Alpha	The Zip Code on the right side of the street segment	USPS Pub 28
L_FIPS	5	Alpha	Left side FIPS code (County, Municipal or Township)	ODOT
R_FIPS	5	Alpha	Right side FIPS code (County, Municipal or Township)	ODOT
RD_NUM			Route or local road number	ODOT
JURISDIC	1	Alpha	Designation for jurisdiction (Same as previous DOT Road Type)	
CARDINAL	1	Alpha	Indicated as C for cardinal N for non-cardinal	ODOT
NLFID	14	Alpha	Calculated Network Linear Feature Identifier.	ODOT
BEG_LOG	.001	Numeric	3D Traversed distance from beginning point of route in miles	ODOT
END_LOG	.001	Numeric	3D Traversed distance of route from BEG_LOG in miles	ODOT
LENGTH3D	.001	Numeric	3D Traversed distance of segment BEG_LOG to END_LOG in miles	ODOT
STYPE	2	Alpha	Surface Type, P=Paved, U=Unpaved	ODOT
DIR_TRAV	1	Alpha	Direction of Roadway, B=Bidirectional, F=One Way Cardinal, T=One Way Non-Cardinal	ODOT
LANES	2	Alpha	The total number of through lanes per road segment	ODOT
REV_DIR	4	Alpha	A value of TRUE indicates direction of ascending logpints does not follow direction of geometry	
DIV	1	Alpha	Divided Road Y = yes , the section is a dual carriageway or; N = no, the section is a single carriageway only	
SPEED	2	Alpha	Posted Speed Limit in MPH	ODOT

The specification also states that the positional accuracy of approximately one meter for address points and centerlines is much more stringent than the 3 meter and 7.6 meter standards for the MTdb. The requirements call for 100% of the feature attributes to have valid domain and range values and for all the edges within a county to be topologically matched. It is interesting to note how the Ohio LBRS specifies how an address point should be collected:

> "The Address Points layer contains point (X, Y) locations for all addressable structures. Each point represents a single address, which is located along the access to the structure from the road, typically the structure's driveway."

This deviates significantly from the desire to gather a MSP near the front door or the common practice of placing a point on the rooftop.

Summary

The efforts by NSGIC to foster statewide coordination and standardization are a real success story. These efforts have a major impact on how the GSS Initiative can be implemented. In many cases municipal level data are being coordinated by counties and shared with state offices. In at least 22 states there are one stop shops for address databases that are being maintained according to a high level of consistent standards.

1.b.5 Exemplary State Data Portal – Indiana

It is beyond the scope of our project to comment on each of the state geospatial data portals listed in table 3. The NSGIC 2009 survey provides an excellent summary of the progress the states are making. When following the links to each state, one quickly encounters a hodgepodge of access modes and constraints. Rather than attempt a comparative analysis we have highlighted the efforts by Indiana to collect, standardize, and disseminate geospatial data. We believe that The Indiana Map initiative is an example of what is possible with the right leadership, an enlightened set of policies, and the appropriate technical capabilities. Most importantly, the state has developed an extensive set of partners. As the developers note:

> "IndianaMap is a collaborative effort of the Indiana Geographic Information Officer (GIO), the Indiana Geographic Information Council (IGIC), the Indiana Department of Transportation (INDOT), the Indiana Geological Survey (IGS), University Information Technology Services (UITS) of Indiana University, and other federal, state, and local partners."
>
> "The IndianaMap would not be possible without the active cooperation of local government. To date, 85 [of 92] counties are sharing their data on addresses, streets, parcels, and boundaries. Their continuing participation makes it possible for emergency personnel to coordinate joint responses, economic development

> corporations to showcase and attract new business, and planners to manage growth that takes advantage of Indiana's natural resources." [51]

The data include approximately 261 different themes (Table 6). The Indiana Spatial Data Portal (ISDP) provides access to more than 20 terabytes of Indiana geospatial data. Most datasets are available to the public for download and have no use restrictions. Indiana University's (IU) high performance networks and computing infrastructure support the ISDP which archives and provides web access to imagery provided by data partners within and outside IU. From the view point of the GSS Initiative, access to roads, address points and parcels would be most important. For 85 counties access to these layers is facilitated by a number of technical alternatives. For starters, the web mapping interface provides a convenient access mechanism that enable one to navigate and query the data.[52] For example one can quickly display imagery, streets, and address points (Figure 15). From this interface it is possible to select specific address points, generate a map, or go directly to the FTP site to download the data in Esri shapefile format (Figure 16). The downloaded data can be added to an ArcGIS project and analyzed in relationship with other web basemap services (Figure 17). Alternatively, one can add any of the features directly as a web service or open the IndianaMap ArcGIS map document. As a map document one has desktop access to almost all the state within a standard GIS operating environment. The example in Figure 18 incorporates the IndianaMap themes with a newly created Esri TIGER 2009 map service. This process provides a quick way to compare TIGER with high resolution local resources almost anywhere in Indiana. These data could be exported to GeoPDF or moved to a mobile device for field verification.

[51] See (http://www.indianamap.org/about.html)
[52] See http://inmap.indiana.edu/viewer.htm

Table 6. List of Indiana Map data themes

Indiana Map Data Themes - Non Geology		
PLSS & Quad Boundaries	**Other Infrastructure**	
Counties	Airports (NTAD)	**Management Areas & Misc.**
Landsurvey - Sections	Airports - Public (INDOT)	Air Quality (NTAD)
Landsurvey - Townships	Airports - Boundaries (ESRI)	Bacteria Monitoring
State	Alternative Fuel Sites	Managed Lands (IDNR)
USGS Quadrangle (1:24,000)	Broadband - Wireless Coverage	Hoosier Natl. Forest Mgt. Areas
USGS Quadrangle (1:100,000)	Broadband - Wireline Coverage	Hoosier Natl. Forest Ownership
USGS Quadrangle (1:250,000)	BMV License Branches	Patoka River NWR
Contours, Imagery, & Other	Care Facilities - Long Term (	Wind Speed at 50 m
Aerial Photos - 2008 (NAIP)	Cell Towers (FCC)	Wind Speed at 100 m
Aerial Photos - 2007 (NAIP)	Cemeteries	Wind Power at 50 m
Aerial Photos - 2007 (Central Ind	Dams (IDNR)	Wind Power at 100 m
Aerial Photos - 2006 (NAIP)	Dams 1996 (EPA)	**HYDROLOGY**
Aerial Photos - 2005	Electric Service Territories	Hydrography
Aerial Photos (Infrared) - 2005	Emergency Facilities (HAZUS)	Canal Routes - Historic
Aerial Photos - 2005 (NAIP)	Emergency Med. Services (IDHS	Canal Structures - Historic
Aerial Photos - 2004 (NAIP)	Fire Stations (IDHS)	Floodplains - DFIRM
Aerial Photos - 2003 (NAIP)	Hospitals (IDHS)	Floodplains BFE - DFIRM
Aerial Photos - 1998	hospitals and Rural Health Cl	Floodplain X-sect. - DFIRM
Benchmarks (NOAA)	Hospitals (HAZUS)	Hydrography Points (NHD)
Benchmarks - GPS (NOAA)	Industrial Parks (INDOT)	Lake Shore Features
Elevation	Intermodal Terminals (NTAD)	Rivers (NHD)
Elevation Contours	Libraries	Lakes (NHD)
Shaded relief	Museums	Rivers - Outstanding (NRC)
Surveyor Tie Cards	Pipelines (IGS)	Rivers - Inventory (NPS)
Topo map (USGS 1:24,000)	Ports (NTAD)	Stream Features
DEMOGRAPHICS	Power Facilities	Streamflow Gauges (USGS)
2000 Census & Population	Powerlines - Not publicly ava	Streams (NHD)
Census Blockgroups	Rail Crossings (INDOT)	Weather Stations (NCDC)
Census Blocks	Rail System - Active	Wetlands
Census Counties	Rail System - Active and Aban	Wetland Lines
Census Tracts	Railroads (1:100,000)	Wetland Points
Incorporated Areas (INDOT)	Railroads - Amtrak Stations	**Watersheds & Quality**
Minor Civil Divisions (Civil Twps	Recreational Facilities	Lakes - Impaired (IDEM)
Place Names	Religious Facilities	Streams - Impaired (IDEM)
Populated Areas	Schools - Higher Education (I	Sediment Inventory
Urban Areas	School Districts	Water Quality Observations
Demographic Data	Schools (HAZUS)	Water Quality Statistics
Age (Median)	Trails (IDNR)	Watershed HUC06
Census Data (Historical)	**Environment**	Watershed HUC08
Family Size (Average)	Brownfields	Watershed HUC11
Children in Poverty	Cleanup Sites	Watershed HUC14
Population Change (1990-2000)	Composting Facilities	**Hydrogeology**
Population Density (Blockgroups)	Confined Feeding Operations	Aquifers - Unconsolidated (USGS)
Population - American Indian	Construction Demolition Waste	Aquifers - Bedrock (USGS)
Population - Asian	Corrective Action Sites	Cave Entrance Density
Population - Black	Industrial Waste Sites	Hydrologic Terrains
Population - Hispanic	Institutional Control Sites	Karst Area Dye Lines
Population - White	Leaking UG Storage Tanks	Karst Area Dye Points
Political & Other Boundaries	Manufactured Gas Plants	Karst Springs
111th Congressional Districts	NPDES Facilities	Sinkhole Areas and Sinking-Stream
116th Gen. Assembly - House	NPDES Pipe Locations	Observation Wells (USGS)
116th Gen. Assembly - Senate	Old Landfills	Water Wells (IDNR)
Misc. Govt. Boundaries (IDHS)	Open Dump Waste Sites	Water Wells (iLITH)
Land Parcels (IDHS)	Restricted Waste Sites	**ENVIRONMENT/BIOLOGY**
Time Zones	Septage Waste Sites	Agriculture & Land Cover
Voting Districts	Solid Waste Landfills	Agriculture Census
ZIP Code Tabulation Areas 2005	Superfund Sites	Big Trees 2005
INFRASTRUCTURE	Tire Waste Sites	Crops 2008
Roads	Underground Storage Tanks	Crops 2006
Bridges - County and City (INDOT)	Voluntary Remediation Program	Crops 2003
Bridges - System 1 (INDOT)	Waste Transfer Stations	Crops 2002
Highways (INDOT)	Waste Treatment Storage Dispo	Crops 2001
Interstates (TIGER)		Cultivated Areas 2004
Mile Markers - System 1 (INDOT)		Ecoregions
Ramps (INDOT)		Land Cover 2001 (USGS)
Address Points (IDHS)		Impervious Surfaces 2001 (USGS)
Street Centerlines (IDHS)		Tree Canopy 2001 (USGS)
Roads 2005 (INDOT - labeled)		Land Cover 1992 (USGS)
Roads 2005 (TIGER - unlabeled)		Prime Farmland % (STATSGO)
Roadways (INDOT)		Soil Associations (STATSGO)
Scenic Byways		Soil Survey (SSURGO)
Traffic Counts (INDOT)		
Traffic Zone Analysis		

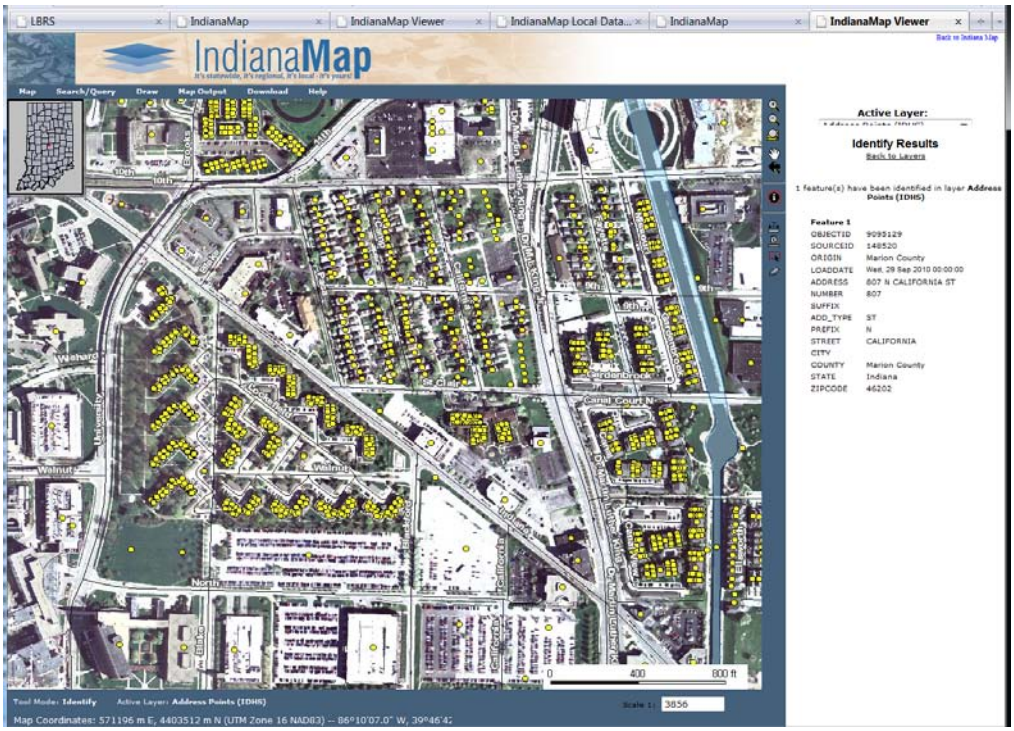


Figure 15. IndianaMap web interface with example of address query results (http://inmap.indiana.edu/viewer.htm)

Figure 16. IndianaMap download site (http://inmap.indiana.edu/download.html)

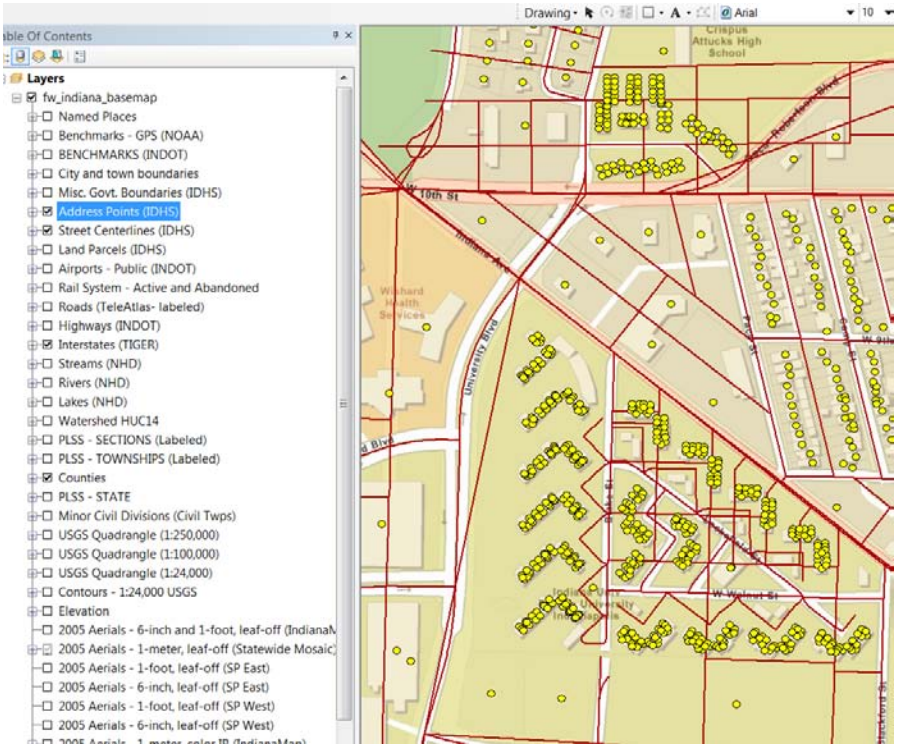


Figure 17. Example of IndianaMap web services with Esri basemap (Cowen, David)

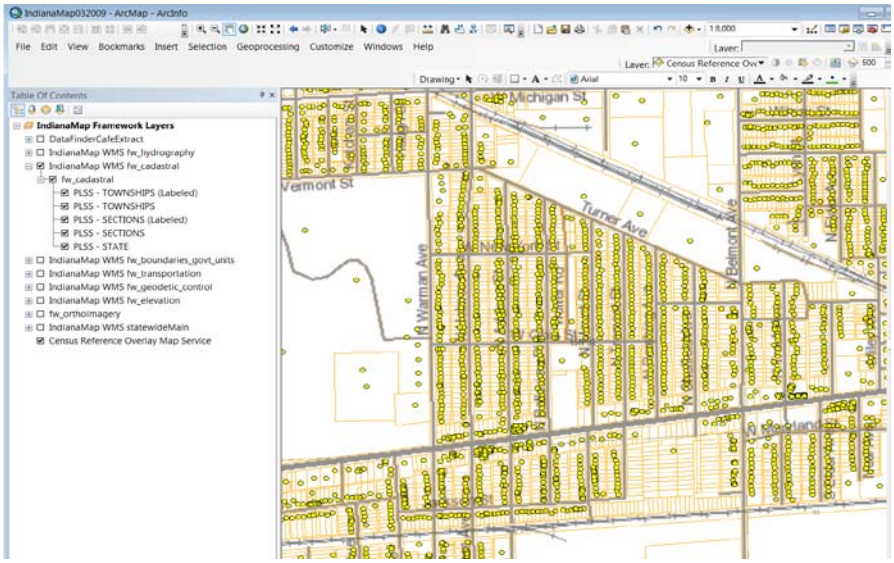


Figure 18. IndianaMap ArcGIS Map document with Esri TIGER map service (Cowen, David)

1.b.6 Exemplary County Geospatial Data Portal - Delaware County, Ohio

Just as Indiana has set a high bar in terms of the creation and sharing of geospatial data, it hard to find a local government that has incorporated enlightened leadership, organizational needs, and technical competency better than Delaware County, Ohio. The system, operated within the auditor's office as the Delaware Appraisal Land Information System (DALIS), is used by a variety of county and city jurisdictions and in numerous applications such as: appraisal, planning, economic development, E-911 related applications, elections, building and code compliance, utilities, engineering, water and waste water mapping to name a few.

From the view of the GSS Initiative, DALIS provides a number of ways to access and acquire address points, centerlines, parcels and several other themes that are updated quarterly and even weekly (Table 7). Publication versions of these data are posted on their FTP server (Figure 19). As part of the Ohio LBRS, all of the data follow positional accuracy standards of approximately one meter. It is also important that the county handles the annual BAS for all the municipalities in the county and maintains these boundary files with reference to the parcel data.

Table 7. Delaware County Ohio download and Metadata table[53]

LBRS Files (Weekly) - 3D fields are blank		
Address Points	Updated Weekly	Download (7.7 MB Zip File) Metadata
Street Centerlines	Updated Weekly	Download (3.4 MB Zip File) Metadata

LBRS Files (*Quarterly) - 3D fields are populated

[53] See http://www.dalisproject.org/%28S%285pg0d155hcsoymvkvqyamf55%29%29/pages/file_download.aspx

Address Points	Updated Quarterly	Download (8 MB Zip File) Metadata
Street Centerlines	Updated Quarterly	Download (3.5 MB Zip File) Metadata
Centerline Points	Updated Quarterly	Download (481 KB Zip File) Metadata

In addition to the ease of access to a download link to acquire current highly accurate shapefile representation of the themes, DALIS operates a web viewer that has multiple search tools, easy tools to clip and extract shapefiles for any part of the county, view Pictometry oblique imagery, and to export the view to Google Earth (Figures 20, 21 and 22).

Figure 19. DALIS project website [54]

[54] For more details see (http://www.dalisproject.org/%28S%285pg0d155hcsoymvkvqyamf55%29%29/pages/file_download.aspx

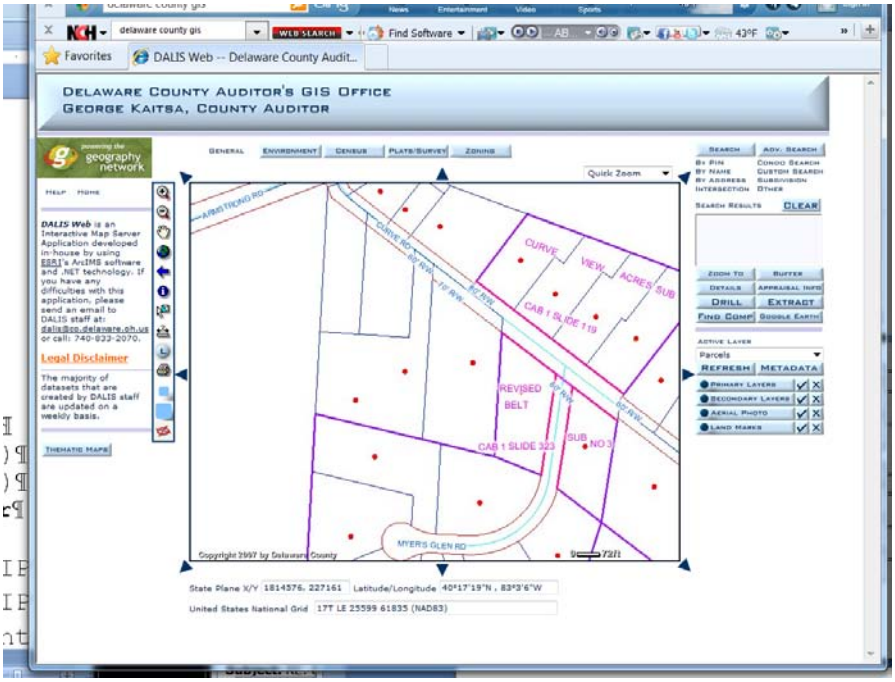


Figure 20. DALIS web mapping site with tools (http://131.187.129.83/map.aspx?INITHEME=General)

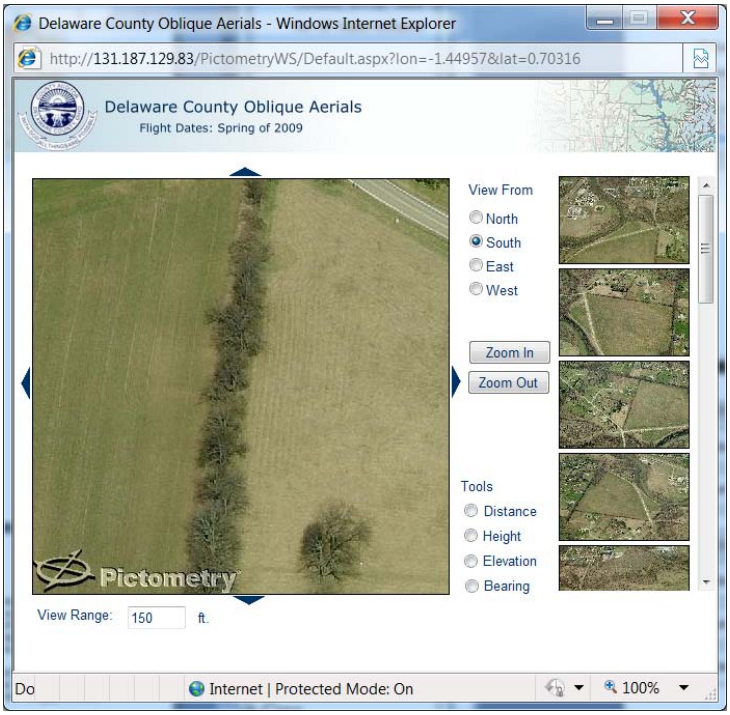


Figure 21. DALIS interface for Pictometry Oblique photos

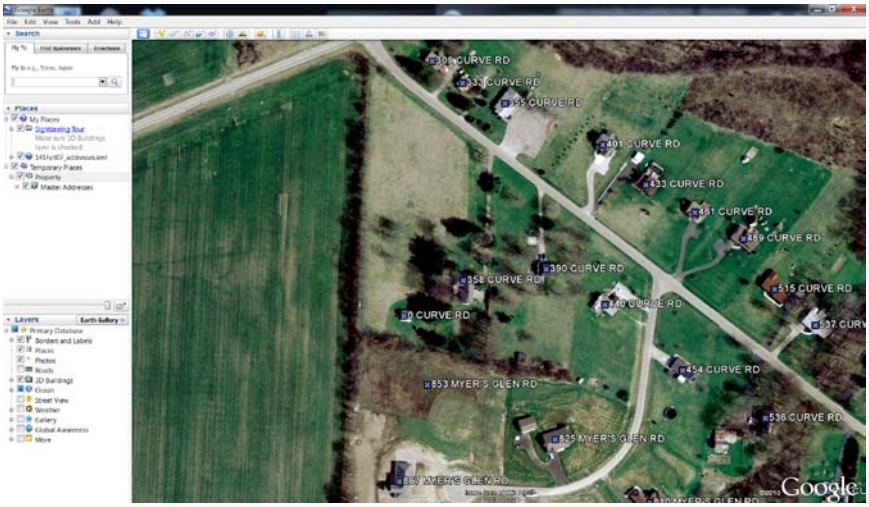


Figure 22. Google Earth Interface directly from DALIS

1.b.7 Exemplary Regional Government- Metro GIS

Our final exemplary geospatial portal candidate is MetroGIS, a regional system serving the seven county Minneapolis-St. Paul metropolitan area. Since its creation in 1996 MetroGIS has focused on coordinating the development of standardized data assets that can meet the need of a multicounty functional region that include "nearly 300 local (city, county, school and water management districts) and regional government institutions". MetroGIS's role is to define shared needs and, through consensus-based processes, define sustainable collaborative solutions to these needs and oversee the negotiations to secure willing custodian organizations to support them. All solutions are also intended to be work toward the broad objectives of the National Spatial Data Infrastructure (NSDI). Representatives from state and federal agencies as well as academic, non-profit and for-profit interests also participate in the MetroGIS decision making. One of the most intriguing aspects of MetroGIS is its organizational structure According to a recent statement by its director (Johnson, Randy 2010):

> "MetroGIS, since its launch in 1996, has been governed by a board of elected officials even though it does not have a statutory standing. MetroGIS is not formally created by law and has no ability to manage funds or retain staff of its own. In other words, MetroGIS is a voluntary collaborative forum through which the stakeholder organizations collectively define geospatial needs they have in common, agree upon collaborative solutions to those needs, and most importantly, voluntarily agree to serve as the custodians of the various operational components for each on behalf of the broader community. This unconventional structure remains an anomaly in the United States, yet the public value created is undisputable in terms of the organizational efficiencies that have been gained. As importantly, the value of engaging elected officials in a range of organizational and operational aspects to defining and acting on shared geospatial needs has proven to provide substantive benefits not only to achieving proposed collaborative solutions but also to improving

understanding among elected officials of the public value that can be created and engaging key political leaders in peer-to-peer advocacy for geospatial initiatives."

MetroGIS has established robust web services to find, view, and retrieve any of its 235 geospatial datasets. These include DataFinder Catalog which provides metadata describing GIS data sets, many of which can be directly downloaded or used via map services. One can even subscribe to the RSS feed for notification of updated or new datasets. An example of the catalog format is provided in Table 8 and web services in Table 9

Table 8. MetroGIS DataFinder options[55]

Facilities, Buildings and Structures back to top ^

Information related to details of man made structures

Dataset name (links to metadata record)	Last Updated (★New/▲ Updated in last 6 months)	MetroGIS Endorsed	Show in Cafe	Map Services ■ Image ■ Feature	Download Data
Bus Garages			map it	■ ■	
Dakota County Building Lines					
Dakota County Building Polygons					
Geographic Names Information System (USGS Place Names)			map it	■ ■	
Major Shopping Centers			map it	■ ■	
MetroGIS Address Points	05/05/2010	●			
NCompass Landmarks - Points of Interest	▲ 12/09/2010				
Ramsey County Bridges					
Ramsey County Buildings					
Ramsey County Dams					
Ramsey County Watertowers					
Scott County Planimetric Data					

Table 9. MetroGIS Map Services[56]

Map services available from DataFinder back to top ^

◉ All Map Services ○ WMS ○ WFS ○ ArcIMS Image ○ ArcIMS Feature

Map Service Name	Type	OGC/ESRI	WMS/WFS Capabilities
MN_MetroGIS_DataFinder_WMS_Administrative_Political_Boundaries	Image	WMS	GetCapabilities
MN_MetroGIS_DataFinder_WFS_Administrative_Political_Boundaries	Feature	WFS	GetCapabilities
MN_MetroGIS_DataFinder__Administrative_Political_Boundaries	Image	ArcIMS Image	N/A
MN_MetroGIS_DataFinder_FS_Administrative_Political_Boundaries	Feature	ArcIMS Feature	N/A
MN_MetroGIS_DataFinder_WMS_Agriculture_Farming	Image	WMS	GetCapabilities
MN_MetroGIS_DataFinder_WFS_Agriculture_Farming	Feature	WFS	GetCapabilities
MN_MetroGIS_DataFinder__Agriculture_Farming	Image	ArcIMS Image	N/A
MN_MetroGIS_DataFinder_FS_Agriculture_Farming	Feature	ArcIMS Feature	N/A
MN_MetroGIS_DataFinder_WMS_Base_Maps_Scanned_Maps_Charts	Image	WMS	GetCapabilities
MN_MetroGIS_DataFinder_WFS_Base_Maps_Scanned_Maps_Charts	Feature	WFS	GetCapabilities
MN_MetroGIS_DataFinder__Base_Maps_Scanned_Maps_Charts	Image	ArcIMS Image	N/A
MN_MetroGIS_DataFinder_FS_Base_Maps_Scanned_Maps_Charts	Feature	ArcIMS Feature	N/A
MN_MetroGIS_DataFinder_WMS_Biology_Ecology	Image	WMS	GetCapabilities

[55] For more details see (http://www.mngeo.state.mn.us/GeoServiceFinder/)
[56] For more details, see (http://www.mngeo.state.mn.us/GeoServiceFinder/)

Of particular importance to the GSS Initiative is the plan that MetroGIS has implemented to acquire address points from 192 authoritative sources in the region (See Figure 23). This application was highlighted in our report *Reporting the State and Anticipated Future Directions of Addresses and Addressing*. From the view point of data exchange, MetroGIS has developed a general online tool that enables local governments to add new address points. These points are synchronized by the staff every night. The MetroGIS system could serve as a model for transactionally maintaining an address point file from authoritative sources.

The address locations visible in Figure 23 represent a compilation of address point data from cities and counties in the seven county Twin Cities, Minnesota metropolitan area. According to MetroGIS "…the dataset is intended to contain a point location and the official address (as defined by the address authority) for all occupiable units and any other official addresses within the jurisdictional boundary of each address authority that provides data. A number of other attributes are available in the dataset, but may not be populated by some address authorities."[57]

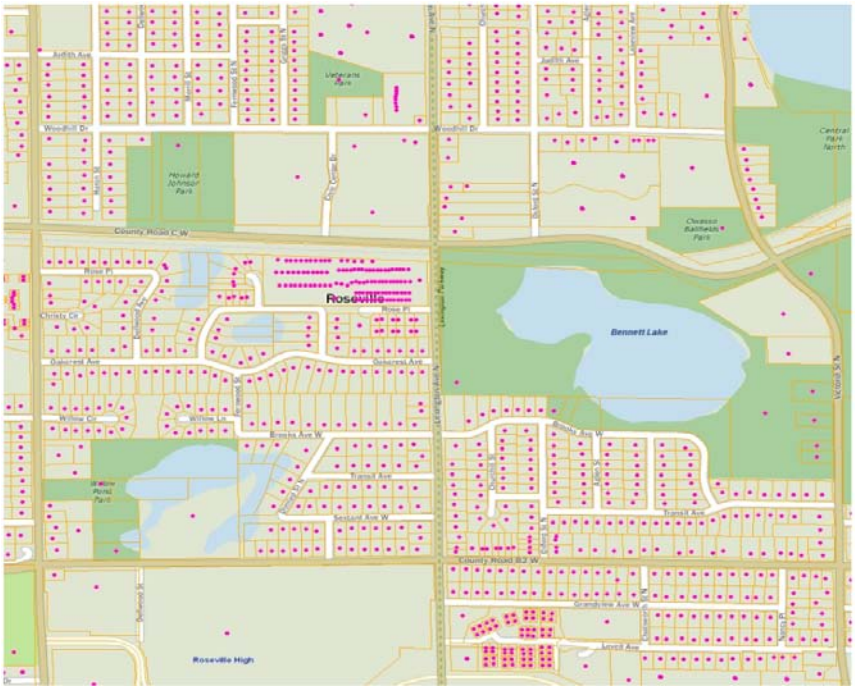


Figure 23. Example of MetroGIS address points

Summary

By highlighting three exemplary programs at different levels of government we have attempted to demonstrate how data can be discovered, retrieved, documented, and incorporated into a number of different user environments. We also believe that these three

[57] For more details, see http://www.mngeo.state.mn.us/GeoServiceFinder/

efforts are models for others. We believe that many other government agencies are adopting similar philosophies regarding data sharing and tools to find and use geospatial data.

As the Geography Division begins the GSS Initiative, it is in an excellent position to reap the benefits of major efforts by its non-federal partners. In the current environment, many local, regional, and state organizations want to be recognized as the authoritative source of geospatial information for their area. They want to demonstrate to the tax payers and elected officials that development of their data and systems has been a good investment. In other words, they need to demonstrate that Google Maps and Microsoft's Bing Maps do not represent the authoritative source of information for their community. A good manager understands that the best way to accomplish this is to establish and maintain systems that provide easy access to the most current and trusted source of information. The ultimate success of the GSS Initiative is directly related to the Geography Division's ability to take advantage of local resources. In order to accomplish this, we recommend the development of innovative procedures to utilize the "goodies that have been thrown over the fence" with the minimum of additional effort by the local partner.

Task 2. Data Exchange Methodologies and Evaluation

In this section of our report we will identify the relevant components that could lead to the successful exchange of address and spatial data between the Census Bureau and its partners. As part of that goal, we will describe available and appropriate data exchange methodologies and evaluate the pros and cons of exchange methodologies. Our original scope of work included detailing data exchange formats, but we were cautioned that the problem facing the Geography Division (which uses both FME and ArcGIS data interoperability tools extensively) generally lies not in unusable formats, but rather in issues related to data specifications, content, and quality. In other words, data standards or formats do not always result in standardized data and we have changed our focus, by de-emphasizing exchange formats and enhancing our focus on the identification and evaluation of methodologies for the exchange of address and spatial data.

2.a Available and Appropriate Methodologies of Data Exchange

As part of the data exchange operation we are assuming that, at a minimum, Census wants to obtain information on spatial features and addresses from various partners, say a local government agency. A major issue to be dealt with when considering the methodologies of data exchange among multiple parties concerns the architecture of the spatial data system.

One design consists of a common database, containing one set of addresses and spatial data, with shared access and joint maintenance by the data contributors. This design of a collection of cooperating but autonomous component database systems is sometimes called a federated data base system. From a technical standpoint, the approach of multiple parties editing a common data base (either centralized or distributed) has clear advantages. A common database ensures a uniform data schema and compatible data dictionaries. In a homogeneous data base environment, numerous tools exist to maintain the integrity of the system and perform data maintenance and update. A system built on this type of architecture was implemented in the Pennsylvania Federated GIS, under the auspices of the Pennsylvania Geospatial Technologies Office.[58]

The participants, a collection of counties in Pennsylvania, pooled their locally maintained data holdings into a seamless, shared data layer. In this distributed database architecture, each participant utilized the same database software (ArcGIS Enterprise Server), extracted the appropriate content from their individual operational databases, translated that data to conform to a common data schema (Esri's GIS for the Nation data model), and loaded that information into the federated database. When a county altered their data, the changes were posted to the federated system.[59] The system allowed the participants to effectively share the most current, accurate information with users locally, regionally, and statewide.

[58] http://www.portal.state.pa.us/portal/server.pt/community/geospatial_technologies_office/913

[59] Sweet, Thomas, O. 2009. Pennsylvania Federated GIS Enables Data Sharing across County Lines, *ArcNews*, Fall 2009. Available online at: http://www.esri.com/news/arcnews/fall09articles/pennsylvania-gis.html

However, this approach requires the implementation of a common technology platform across all of the collaborating parties. In many cases, even with a few partners, this is difficult to attain. Given that the Geography Division must interact with thousands of partners, this strategy is not feasible.

A more promising variant of this is to use a web-based approach in which the Geography Division maintains a database on-line (either open to the public or restricted to partners) and allow users to modify the contents, a la Open Street Map (OSM). An in-depth analysis of user generated content and the related technology is provided in our report *Identifying the Current State and Anticipated Future Direction of Potentially Useful Developing Technologies* (see section 3.5 on the potential use of crowdsourcing).

For this approach to work most effectively, the selected TIGER content needs to be presented in accordance with a well-defined, detailed data schema, or template (that starts with collection rules and includes feature definitions, feature instance rules, etc.). Local authorities can add/edit data (using web based editing tools) to the TIGER map with their contributions conforming to the rules imposed by the schema template. This forces the data that is provided to be consistent with all the rest of the data. This approach would seem to have merit to replace or supplement the use of the MTPS in the BAS updates.

Esri has told us that they want utilize this type of approach to be able to implement web editing of TIGER as part of their Community Maps Program, specifically the program dealing with local government.[60] This would require a harmonized approach and data model that Esri can provide to its local government customers through their Local Government Resource Center. It is our opinion that TIGER is viewed as the "starter" dataset for this program. Esri plans on providing more templates for its users to edit and manage their street centerlines as well as administrative boundaries and will also include tools for managing addresses for both as well. Esri would enable editing by various governmental authorities using a variety of tools (e.g. ArcGIS Desktops in GIS shops in local governments as well as a series of web editors to be used by selected web audiences). Esri would include Extract, Transform, and Load (ETL) procedures for these users to share their data with Census.

ArcGIS provides a mechanism to configure web editing on any shared geodatabase. This is done using ArcGIS Server, which can be set up to support a number of types of GIS services. The ones used for editing include map services and feature services. When you configure a map service, an ArcMap document is published as a map service and you can optionally enable a number of map service "capabilities". For example:

- Turn on a WMS service and/or a KML service for your map service (automatically has Simple Object Access Protocol (SOAP) and Representational State Transfer (REST).
- Enable a feature service and editing if any of your map layers reference an enterprise (ArcSDE) geodatabase. You can also specify particular geodatabase versions that web users would work with to track and manage how these edits are ultimately

[60] Brown, Clint. Director of Software Products, Esri. Personal communication. December 13, 2010.

inserted into your shared geodatabase. This enables the use of your own version-based workflows for web editing.

In ArcMap Version 10 and higher, users are able to define the actual editing rules and attribute values that are collected when features are edited. You specify "feature templates" that implement editing rules and tools for selected types of features in your dataset. An example of this is shown in Figure 24.

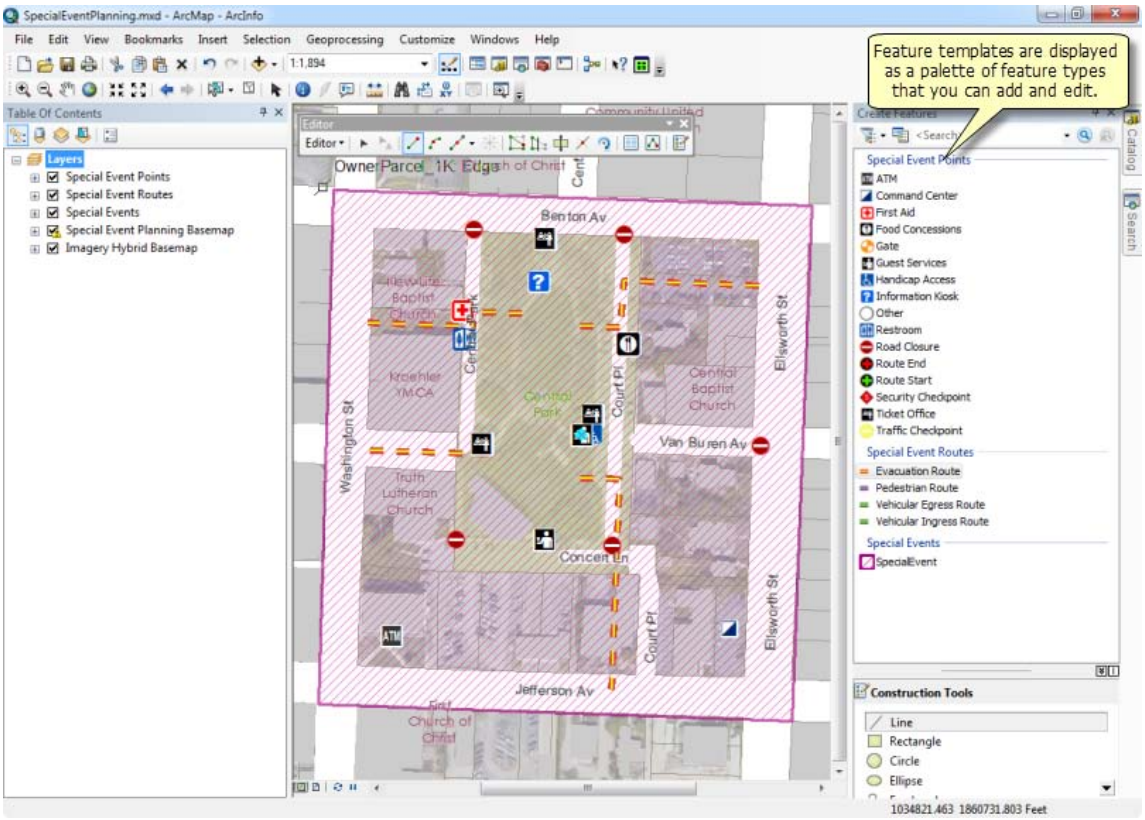


Figure 24. Use of feature templates to control editing (Source: Esri)

This strategy requires a harmonized approach for data attribution and a standard data model that can be provided to local government customers. Templates would be constructed for users to edit and manage their street centerlines as well as administrative boundaries along with tools for managing addresses. A feature editing template is shown in more detail in Figure 25.

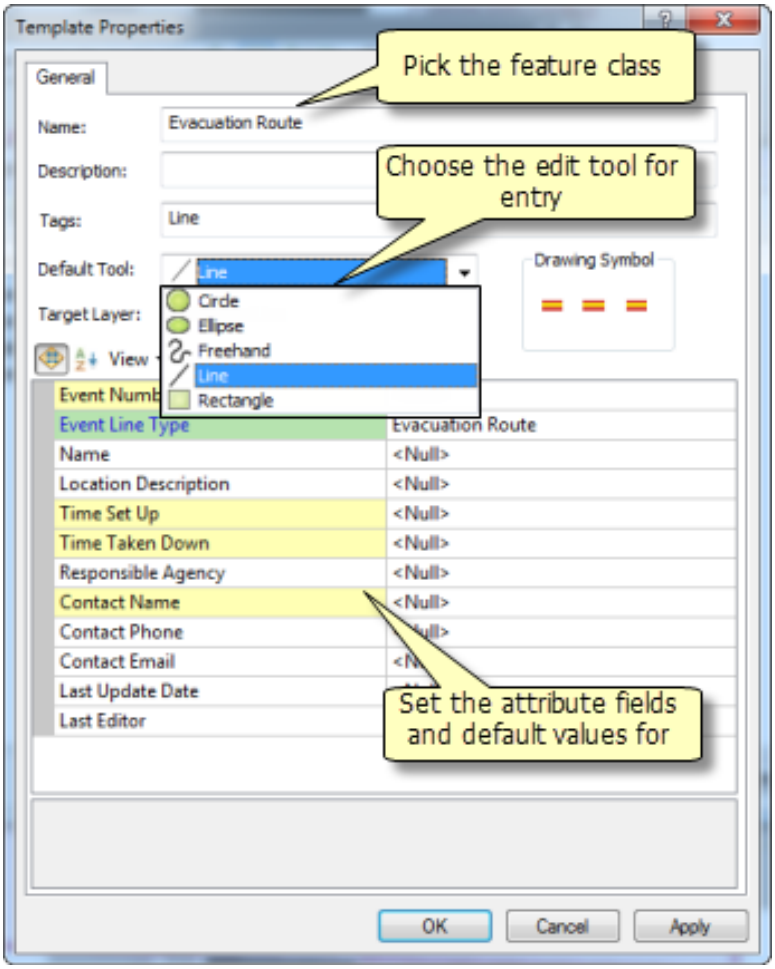


Figure 25. Using a feature template to control addition of attributes (Source: Esri)

The steps involved in this process would be as follows:

1. Set up a versioned ArcSDE geodatabase with permissions and define workflows for compiling new information along with validation and check-in procedures. Many types of users can update and contribute to shared geodatabases using a variety of procedures and workflows.
2. Identify the types of users that you want to enable editing for and which ones will contribute edits via the web. For each user type, create a map document for editing with feature editing templates (editing rules for selected feature types).
3. Publish each as an ArcGIS map service for editing along with the corresponding feature services.
4. In ArcGIS online, create and configure a web map along with its editing capabilities.
5. Share the web map with others. They can then open the web map in their browser, and use the palette of tools to edit their data.

This process is shown diagrammatically in Figure 26. As we noted in our report *Identifying the Current State and Anticipated Future Direction of Potentially Useful Developing Technologies* a similar process could be implemented using Open Source software.

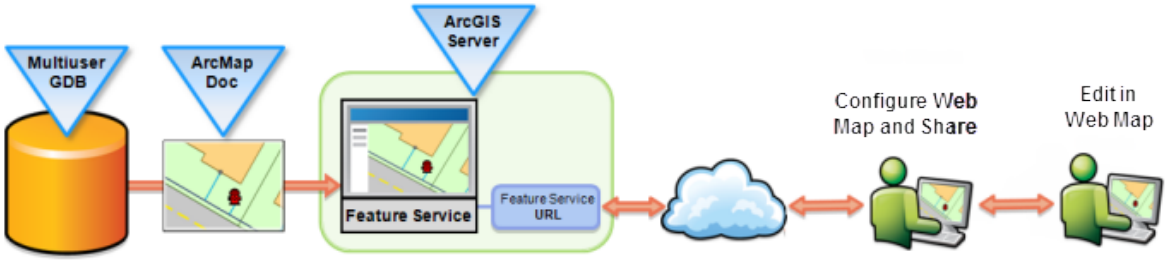


Figure 26. Summary of the web map publishing and editing process (Source: Esri)

A second approach to data exchange retains autonomous databases with each party but has shared information content whereby common data elements are identical and the data are jointly maintained by a transaction-processing scheme. As noted above, many factors (legal, institutional, and technological) can cause cooperating organizations to adopt a heterogeneous approach. In such cases, the collaborating parties can develop a strategy to support shared information content between heterogeneous, discrete, geographic databases. Provided the databases have standardized schema and feature definitions, and share a system of unique feature ids (e.g. TIGER line id), it is possible to create a spatial transaction system between heterogeneous databases. At a conceptual level, the spatial transactions are no different than the SQL transactions: update, insert, and delete. Using the road network, for example, the inset command adds a new road segment to the database, the delete command removes a non-existent road, and the update command gives a more precise geographic description to the road already in the database. These commands can operate on feature instance data (denoting the existence of a feature), feature attribute data (describing the non-spatial characteristics of the feature), and the feature geometry (describing the spatial characteristic of the feature).

By using an agreed upon GML schema (GML, an XML Grammar serves as a modeling language for geographic systems as well as an open interchange format for geographic transactions on the Internet), it should be possible for collaborators to exchange spatial transactions. Utilizing GML, commercial solutions have been developed to perform this function. One solution that issues and receives change-only or incremental updates is offered by Snowflake Software's GO Publisher / GO Loader[61]. This product has been utilized with the UK Ordnance Survey Master Map, a national database of comparable complexity with the MTdb, to handle change information. Snowflake Software is also working with the Federal Aviation Administration to use the software to create and distribute "Digital Notices to Airmen" (changes to the digital aeronautical charts). A schematic diagram illustrating the functions of the software is shown in Figure 27.

[61] See (http://www.snowflakesoftware.co.uk/products/goloader/features.htm)

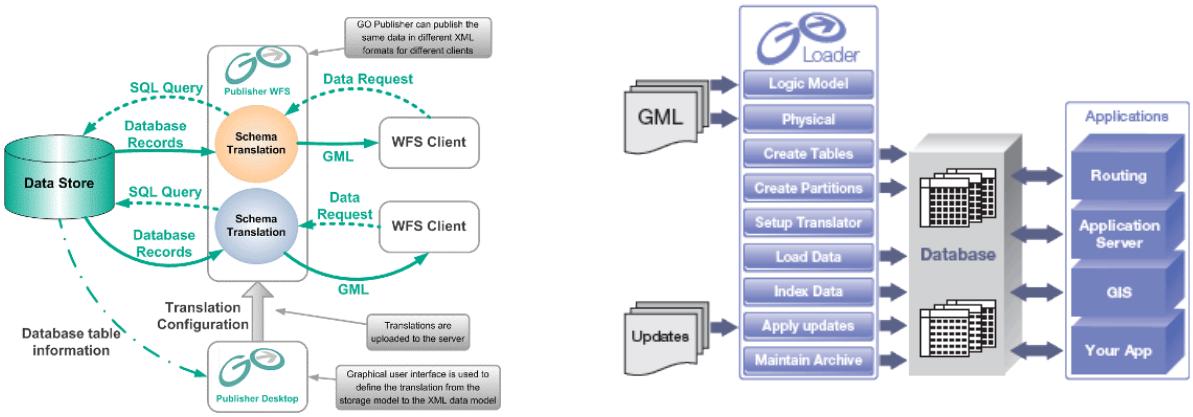


Figure 27. Functional capabilities of GO Publisher & GO Loader. Source: Snowflake Software, http://www.snowflakesoftware.com/products

The process of handling of updates across multiple, heterogeneous, geospatial databases has been termed "geosynchronization". The Open Geospatial Consortium (OGC) has recently released an engineering specification dealing with the topic of geosynchronization that is directly relevant to the Census Bureau. This specification, entitled *OWS 7 Engineering Report -- Geosynchronization service*,[62] addresses all of the concerns raised in this discussion. The scope of the problem, as characterized by the OGC is as follows:

> In order to satisfy these requirements [to deliver current, timely and verified data], data providers must collaborate with outside entities to collect new data and/or update their existing data holdings. This may, for example, mean synchronizing their data with closest-to-source providers as might be the case between municipal, state/provincial and/or federal levels of government.
>
> Regardless of the nature of the collaboration, there is a need for a service to mediate the interaction between data providers and outside entities acting as data collectors. The service must support data entry with validation, notification of changes to interested parties and allow replication of the data provider's features.
>
> A Geosynchronization service, deployed by a data provider, sits between the provider's data store(s) and data collectors. It allows data collectors to submit new data or make modifications to existing data without directly affecting the data in the provider's data store(s) until validation has been applied thus ensuring that the data published by the provider is of high quality.
>
> This candidate standard makes no assumption about the nature of the validation or the quality criteria that a provider has established; only that the Geosynchronization service allows the data to be evaluated for quality before being applied to the provider's data store(s). For example, changes submitted by trusted data collectors may be applied directly to the provider's data stores while data

[62] Open Geospatial Consortium. 2010. OWS – 7 Engineering Report – Geosynchronization service, OGC 10-069r2. Edited by Peter Vretanos. Available online at: (http://portal.opengeospatial.org/files/?artifact_id=39476)

> obtained from external sources would be subjected to more rigorous validation. Validation might involve a manual process where authorized and certified reviewers check each submitted piece of data to assess its quality. Similarly, automated heuristics might be applied to assess the quality of the data. In any case, the data is validated before being applied to the target feature types.
>
> Interested parties (or subscribers) must be able to flexibly track which changes have been submitted by data collectors, what the disposition of those changes was (i.e. whether the changes were accepted or rejected by the validation process) and also track the actual changes made to a provider's features for the purpose of replication. This flexible notification capability is achieved by allowing subscribers to specify predicates that precisely identify about which events they are interested in being notified. For example, a subscriber might register that they are interested in receiving notifications about proposed changes made within a specific geographic area.
>
> Finally, a Geosynchronization service must support replication of the data provider's features by maintaining an event channel containing a log of all changes made to the features. This channel may be used to support several modes of replication including having the Geosynchronization service read the change log and apply the changes to one or more target feature types that have been subscribed to for replication. Replication may be full or partial possibly involving a schema translation from the provider's feature schema to the schema of the target feature.[63]

Two cautionary notes need to be sounded in this discussion. First, it is incumbent upon the data providers to design the data evaluation and validation routines to check the quality of the submitted information and whether to accept or reject the changes that are proposed. Second, this geosynchronization service presumes a level of compatibility between the data schema of the data providers and the schema used by the data collectors (updaters). If the schemas are not identical, then translations between the schemas must be performed. We have been told that the Geography Division has had difficulty in obtaining information about the data schemas used by data contributors. However, with the increasing use of OGC interoperability specifications (which include XML schema encodings) and “best practice” data model templates by commercial GIS vendors, this may not be an impediment in implementing this solution. The OGC geosynchronization model is illustrated in the process wiring diagram shown in Figure 28 below.

[63] *Ibid.*, p. 15

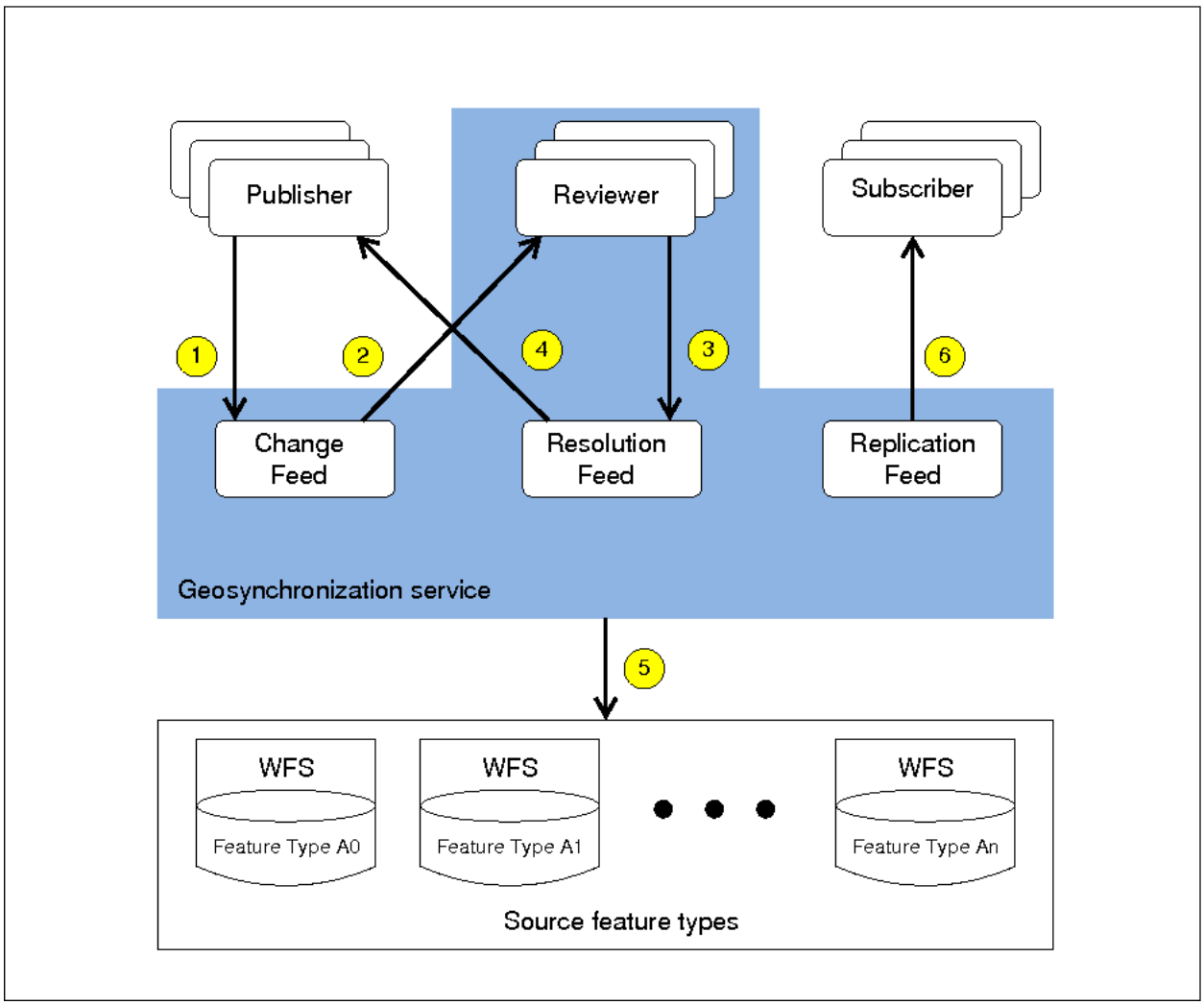


Figure 28. Geosynchronization wiring diagram[64]

Examining figure 28, with reference to the numbers within the yellow circles, the basic geosynchronization service (GSS) usage proceeds as follows:

1. A data publisher posts a proposed change to the GSS for one of the feature types within the GSS's domain of control (i.e. A0, A1… An).

2. The proposed change is read from the Change Feed and reviewed to ensure that it satisfies correctness and quality standards established by the data provider whose feature type is to be updated (i.e. one of A0, A1…An, in this diagram). The review process can be manual or automatic.

3. If the proposed change is rejected, that event is registered in the Resolution Feed and a notification is sent to all interested subscribers.

4. If the proposed change is accepted, that event is registered in the Resolution Feed…

[64] *Ibid.*, p. 43

5. …and the GSS applies the proposed change to the intended feature type.

6. …and the GSS registers all the changes made to the target feature type (i.e. A0, A1…An) in the Replication Feed. It also sends a notification to all interested subscribers that the proposed change has been accepted and that changes have been made to one of the feature types within the geosynchronization server's control domain (i.e. A0, A1…An). Changes posted to the Replication feed may also be applied, by the GSS, to any external feature type that has subscribed for synchronization. The changes are first passed through the schema translation engine, if necessary, to convert them to the schema of each target feature type. If a schema translation is required, transformation metadata is read by the GSS and the transaction is transformed to the schema of the target feature type and then the change is applied to the target feature type via WFS.[65]

To summarize, this geosynchronization service is designed to allow data collectors to propose changes to be made to a data provider's features. A change proposal can be made to create new data or to modify/delete existing data. The proposed changes are reviewed (either manually or automatically) and are either accepted or rejected. If accepted, changes are applied to the feature(s). It seems that the intent here is to provide a standardized protocol for applying incremental updates to existing geospatial databases.

However, as noted earlier, this presumes the existence of a database that has been harmonized among the data providers and data collectors (i.e. data producer – Geography Division; data contributor – local government). This presumption is often false. In many instances, the parties maintain autonomous databases, with analogous, but independently maintained data content and data elements. There is a minimal degree of commonality and cooperation between the databases. When minimal collaboration exists between the data exchange parties, they may choose to transfer complete files of data that cover a given geographic extent – a county sends Census a file containing all of its addresses or roads. Figure 29 shows one example of this in which the county of Charleston SC has collected roads based on its E911 collection and they are significantly different from the roads shown in TIGER Line 2009. The process of reconciling two such disparate datasets is often termed “conflation”.

[65] *Ibid.*, pp. 43-4.

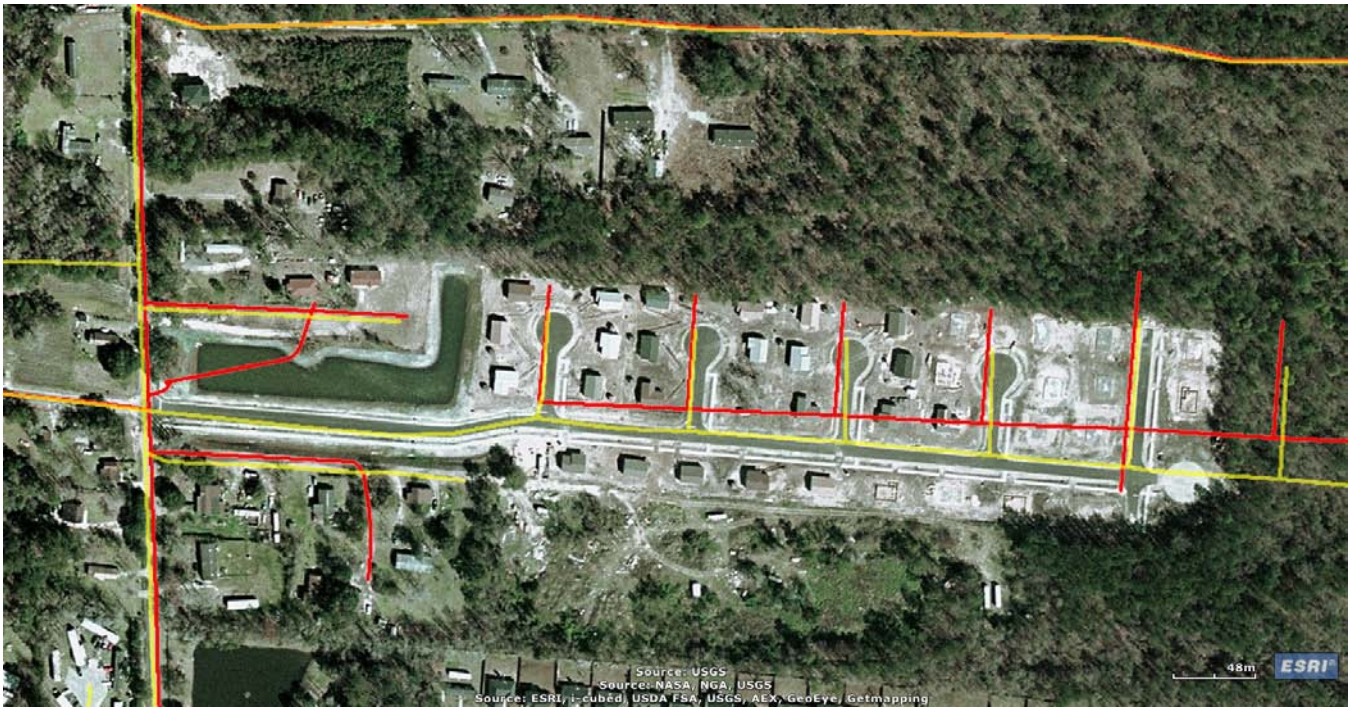


Figure 29. TIGER Line 2009 (red) and Charleston County E911 Roads (yellow)

While this mode of exchange places minimal impacts on the donors, the resources needed by the recipient (Census) to utilize the information may be quite substantial. The past experiences of the Geography Division with this technique have not been not optimal. Whole GIS files have been received from partners that usually do not match Census data and significant manual intervention was required. We will review the conflation topic and discuss recent advances that pertain to Geography Division operations in section 2b.2 below.

2.b Relevant Issues Influencing Data Exchange

When undertaking a data exchange it is important to know the circumstances of the data and the data processing environment of each of the parties involved. It is unlikely that Geography Division's data requirements will change or that the Census Bureau can alter the ongoing business practices of potential partners. Thus the Geography Division needs to understand the differences that exist with potential data partners, document these differences, and assess the impact they might have on Census processes. In cases where the level of difficulty of utilizing the contributed information exceeds its value, that source will likely not be used.

The key to establishing effective address and spatial data exchanges with partners is to fully understand and, if possible, standardize the definitions of the features, attributes, and attribute values and to agree on feature instance rules and collection thresholds. These are needed to avoid the semantic heterogeneity differences that are associated with federated geographic

data base systems. More specifically, knowledge of the following areas is needed to best utilize information obtained from a data exchange:

- Definition, Terminology, and Content: define the minimum information content of the common features, attributes, and attribute values (excluding those used for indirect positioning (see below)). The list will be refined by the development of feature delineation, representation, and collection rules.

- Feature Delineation and Representation Rules: define the data base representation of a single occurrence (instance) of each feature. This task occurs concurrently with that of defining indirect positioning standards.

- Indirect Positioning: define the means by which attributes can be attached to the data framework through indirect positioning methods such as street addresses, mileposts, and mile markers and through common non-spatial identifiers as geographic names. This task defines and sets standards for a basic set of methods to be used with the data.

- Data Collection Rules and Procedures: establish initial data collection rules, such as size thresholds for the length and/or width of feature instances.

- Data Quality: decide and adopt threshold standards for elements of data quality (positional accuracy, attribute accuracy, logical consistency, completeness, and currentness).

If these elements can be standardized, documented, and published by the Geography Division for use by local partners in the form of a “best practices” data specification/data model/data schema, it could greatly reduce the difficulties caused by collecting ostensibly the same information in variant forms.

While the various standards for spatial metadata allow such information to be included in a data exchange, often only a minimal amount of information is collected by local data producers. Often the data producers do not have the resources to completely document their data holdings. When they do, the information is often contained in a “lineage” file that may discuss the sources and processes used to create the data set. It may be incumbent upon the recipient to reconstruct the information listed above from that given in the lineage file or from ancillary descriptive information from the data producer.

When obtaining information from a data partner, the Geography Division needs to receive basic information about the quality of the data that are being provided. At a minimum, the Geography Division should obtain information about the following characteristics:

- Completeness – account for all the feature instances
- Temporal Accuracy – feature instances are up-to-date and valid
- Positional Accuracy – spatial feature instances are accurately located
- Thematic Accuracy – feature classes, addresses, and attributes are correct
- Logical Consistency – road and boundary networks are self-consistent; the relationships between roads, address points, and boundaries are consistent among entities

The amount of data quality information probably varies widely, from extensive to non-existent, among the set of potential Geography Division data partners. However this list of characteristics can be used as a template to evaluate data holdings of prospective data exchange partners and to select high value holdings for incorporation into MTdb. If the accuracy of the data quality metadata is in doubt, then a sample of the submitted data should be evaluated by the Geography Division.

While understanding the characteristics (and hence the value) of the data offered by potential partners is the key determinant, the Geography Division should be cognizant of two other factors. The first concerns the information infrastructure of the partner. If the partners operate software systems used by the Geography Division, then it may be easier to assimilate the information from the partner. However, given the capabilities of the Geography Division (noted below) to ingest and convert a wide variety of data, this issue is of lesser importance than it once was. The second issue concerns the frequency with which a partner updates their data holdings (in particular the holdings of interest to the Geography Division). For example, we know of local agencies that release new versions of their geospatial data holdings on a weekly basis. The Geography Division must determine their requirements for data currentness and use this as a factor in their benchmark in evaluating data exchange partners.

2.b.1 Assess the methodologies used by the Geography Division to receive address and spatial data

In addition to understanding the specifications and characteristics of the information content to be exchanged, there are a variety of technical issues that must be considered. Unless the information architecture of the Geography Division can support the incorporation of external data sources, our efforts to understand the data content are only of academic interest.

At the start of this study, before we more fully understood the technical infrastructure of the Geography Division, there was some concern that there were gaps in the capabilities to read and display address and spatial data. This is no longer the case. The Geography Division makes extensive use of Safe Software's FME® platform for data translation and conversion. The Geography Division also uses the ArcGIS Data Interoperability extension to perform data translation and transformation. It can also be used by ArcGIS users to directly read more than 75 spatial data formats and export to more than 50 spatial data formats.

There is still some concern about the limited ways in which updates can be propagated to the MTdb, but these will be addressed in section 2.b.3 below.

2.b.2 Assess technical solutions for adjusting (conflating) TIGER to incorporate new data provided by authoritative sources.

Geospatial data conflation is the combination of two different datasets covering the same area. The term conflation is used in the GIS industry and elsewhere for the many versions of this problem, which range from combining digitized maps and GPS as sources of street

centerline locations, to edge matching of datasets across boundaries, to combining information from different sensors in remote sensing. In general, the aim of the conflation operation is to combine the best elements of both datasets to create a resultant dataset of higher quality than either of the inputs. The conflation techniques used depend on the types of geospatial data being matched: vector to vector; vector to imagery (raster); or raster to raster. However the basic principles are to accurately match conjugate features in each dataset and then align the rest of the geospatial objects (e.g., points or lines) in both datasets by using space partitioning techniques (e.g., Delaunay triangulation) or geometric interpolation techniques.

Automated vector to vector conflation was first proposed by Saalfeld[66] and the initial focus his conflation research was to use the geometric properties of the vector files as the basis for the merger. Later research utilized additional components such as: the similarities of vector shapes, network topology characteristics, and attribute matching. Most recently, matching of the semantics of the geospatial features is being utilized as an additional component in solving the problem (see Adams, et al.[67]; Olteanu-Raimond A.M. and Mustière S.[68]).

The conflation of road networks involves matching two different versions of the same road network to determine which edges (road sections) match. Once this matching has been established, attributes can be transferred between matching road sections, and missing road sections can be added from one network to another.

A number of organizations have developed software routines that automatically establish node and edge matches between two road networks. When the automatic algorithm is unable to determine a match, manual tools are provided to create and delete matches. Commercial products include ESEA's MapMerger Professional[69] and Citygate GIS's Conflex[70], both of which are sold as ArcGIS extensions. An open-source solution, RoadMatcher[71] is also available. Figure 30 illustrates matching and adjustment steps used in the conflation process.

---

[66] Saalfeld, Alan 1988. Conflation Automated map compilation. *International Journal of Geographical Information Systems, 2*(3), 217 - 228.

[67] Benjamin Adams, Linna Li, Martin Raubal, Michael F. Goodchild. 2010. A General Framework for Conflation. GIScience 2010 Proceedings. Zurich, September 14-17, 2010. Available online at: http://www.giscience2010.org/pdfs/paper_211.pdf

[68] A.M. Olteanu-Raimond, S. Mustière. 2008, Data matching - a matter of belief, *13th International Symposium on Spatial Data Handling (SDH'08)*, 23-25 June, Montpellier (France)

[69] http://www.esea.com/mapmerger.php

[70] http://www.citygategis.com/conflex.htm

[71] http://www.vividsolutions.com/products.asp?catg=spaapp&code=roadmatcher

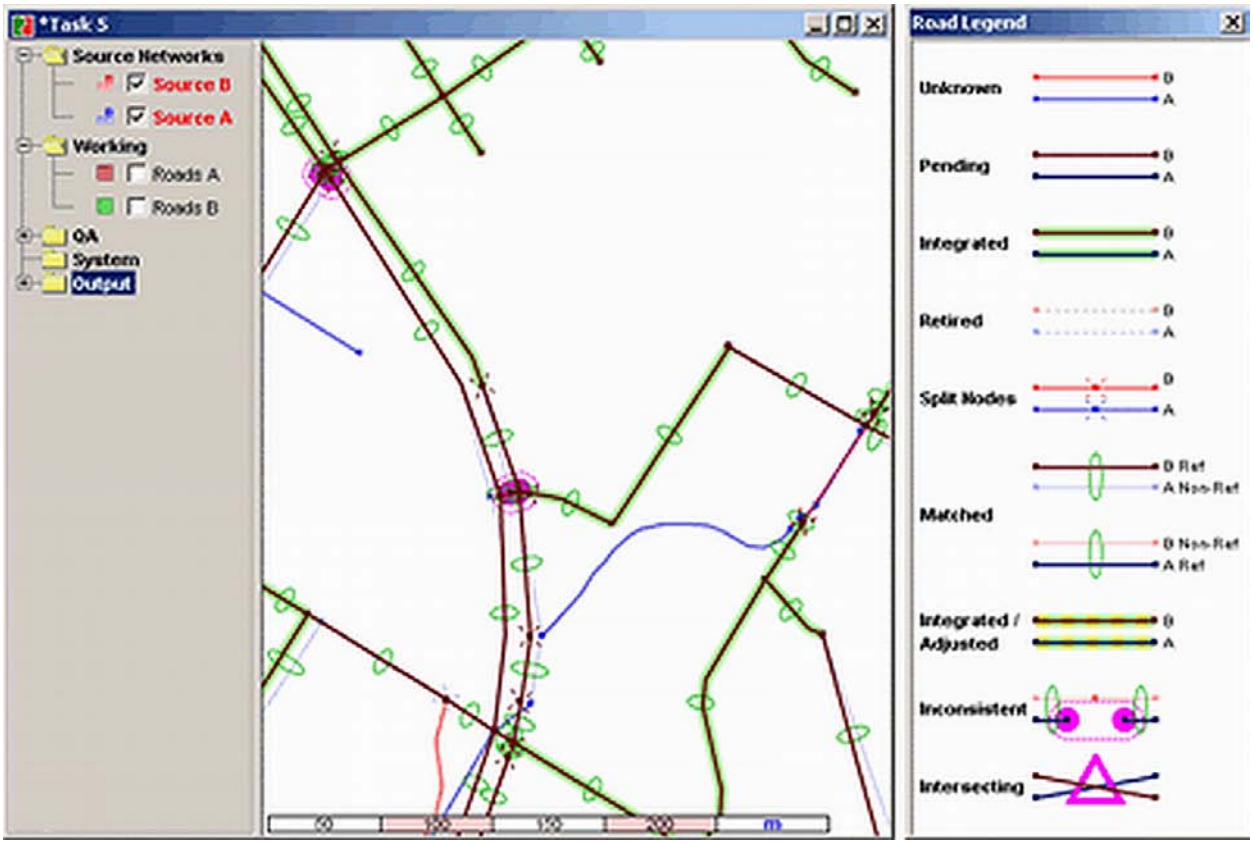


Figure 30. Illustration of conflation processes using RoadMatcher (Source: VividSolutions, http://www.vividsolutions.com/products.asp?catg=spaapp&code=roadmatcher

For simple conflation operations (cases involving two similar datasets, perhaps differing only in the temporal dimension), GIS users can create their own conflation methods by using existing GIS tools. By way of example, we have conflated two road networks – TIGER line 2009, and E-911 roads – for Charleston County SC using the tools available in ArcGIS. Figure 31 shows a small portion of the county and is used to illustrate the steps in the process. In this example, we assumed that the local road network was used to create the 2009 TIGER roads (figure a) as part of the MTAIP process. Therefore, any new segments in the more recent E911 file (figure b) would be located some significant distance away from TIGER. By segmenting the E911 roads into links (figure c) it is possible to find them based on a spatial search. In figure d, all the new links that were greater than 30 meters from the TIGER roads were selected. In order to find the links that intersect the TIGER roads another search is performed to select those segments that touch the new segments (figure e). Of course topology rules could be enforced to snap the links and generate vertices. The E911 attribute fields were renamed and formatted to agree with TIGER (figure f). This procedure was run for all of Charleston County in less than 30 minutes.

Figure 31. (a) TIGER 2009 roads roads

Figure31 (b) TIGER roads + Charlotte E911

Figure 31 (c) Road segments

Figure31 (d) 30 m buffer around TIGER roads

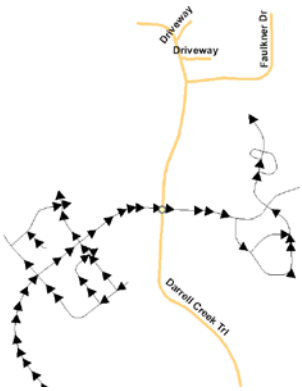


Figure 31 (e) Intersection formed between TIGER and E911 roads

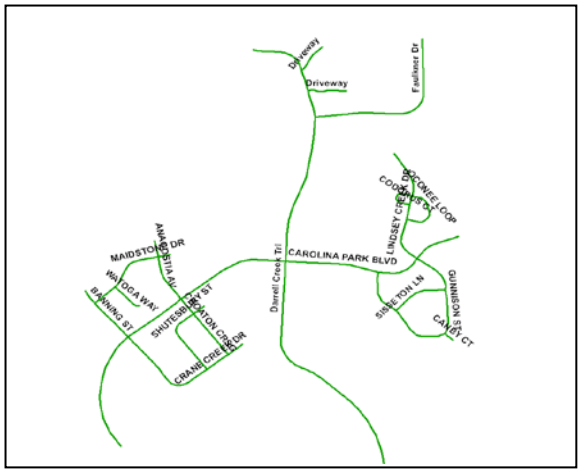


Figure 31 (f) Merged Roads and Attributes

2.b.3 The Role of Oracle Spatial

The Geography Division's operational requirements are supported by a unique, highly customized, database system; the MAF/TIGER database (MTdb). The MTdb provides a seamless national integrated spatial and address database utilizing the Oracle object-relational database management system to store all data elements. Spatial data management routines, including spatial operations and indexing, utilize Oracle Spatial functions. In addition, the Oracle Spatial Topology Data Model stores and manages geographic features and the topological data structures upon which they are built.[72] The interface to the MTdb is controlled by a set of software modules called the "Core API" that has as one component a "Business Rules Engine" containing several thousand rules and logical checks. A schematic diagram of the Core API for updating the MTdb is shown in Figure 32.

Figure 32. MAF/TIGER Core API (Source: Census Geography Division, A. Kalluri)

[72]Danso, Ama. 2010. US Census Bureau MAF/TIGER Product Database: Implementation in the Redesigned MAF/TIGER System Environment. Proceedings, ASPRS/CaGIS 2010 Fall Specialty Conference, November 15-19, 2010 Orlando, Florida

It is evident that it is not feasible for data exchanges with partners to be done directly with the MTdb. It is more practical to consider partner data exchanges occurring with a Product Database (PDB). As noted by Ama Danso:

> "The design approach adopted for the PDB includes database denormalization, use of Oracle Spatial geometry storage exclusive of the Oracle Topology Model, storage of product-specific calculations, and considerations for industry geospatial COTS tools. The goal for the PDB implementation is to create an infrastructure to easily manipulate data into more customer-oriented form with minimal processing and optimal performance….Another aspect of the PDB design is the flexibility to allow geospatial COTS tools to interface with the database structure.
>
> In particular, the use of Oracle Spatial exclusive of the Oracle Topology Model allows the PDB to be registered as an Esri Geodatabase. The registration of the PDB as an Esri Geodatabase makes data available through Esri tools which are widely used in the GIS industry to support spatial data visualization and analysis."[73]

The relationship of the PDB to the MTdb is shown diagrammatically in Figure 33.

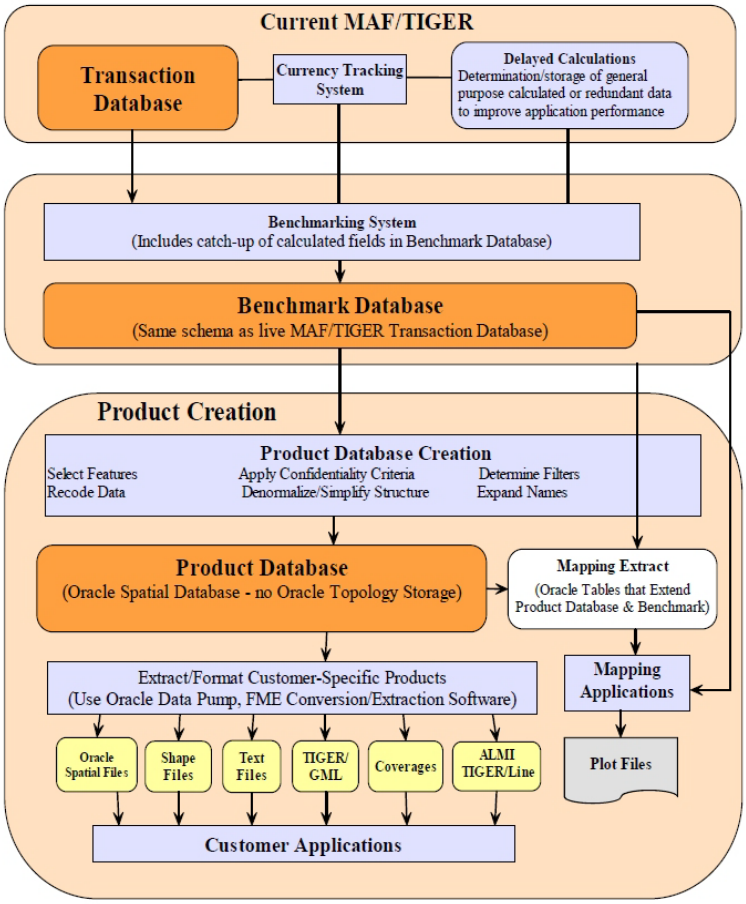


Figure 33. MAF/TIGER System Multi-level Geospatial Database Design Model (Source: Danso, Ama 2010)

[73] *Ibid.*

Technical challenges may still remain in propagating changes that occur from the interactions with the PDB (e.g. additions/deletions/modifications by partners) back to the "live" MAF/TIGER database, although it is our understanding that the Geographic Area Reconciliation Project developed some procedures that have begun to address these issues. Some of the challenges arise not from the choice of Oracle Spatial Topology Model or from the complexity of the Core API (as some have assumed), but from essentially having to synchronize multiple versions of the MTdb. Since the PDB represents the MTdb at a single point in time and the MTdb content is not locked for updates after the PDB is created, the two data holdings may lose synchronization. When changes or edits to the PDB are propagated to the MTdb, those transactions may fail because of other actions taken on the MTdb in the time between PDB creation and the execution of the updates. Further research on methods to synchronize multiple autonomous databases is warranted. It may be the case that technological advances in object-oriented databases would allow for simplification and improvements in the MTdb design and the execution of Core API functions. However, an analysis of this topic is beyond the scope of this study.

Task 3. Identify and research relevant components for the successful exchange of address and spatial data between the Census Bureau and its partners.

We have been asked to identify the relevant components for the successful exchange of address and spatial data between the Geography Division and its partners. We start with a brief examination of the concept of authority and authoritativeness followed by a discussion of the institutional framework that needs to exist to support the interchange and integration of data. Finally, we describe potential two-way transfers that could include the capability to provide both public and controlled or, perhaps, authorized access to address information by partners of the Census Bureau. In turn, we examine the possibilities related to the Census Bureau accepting unlimited and/or controlled access to address and spatial data from partners.

3.a Authority and Authoritativeness

Any form of address and spatial data exchange partnerships with the Bureau of Census must involve an assessment of the quality of the data and the establishment of a level of trust. Issues of authoritative data, authoritative sources, certified data, and data stewardship are critical for establishing the concepts of "trusted sources" and "trusted data".

The notion of authority in the world of addresses and spatial databases is a complex issue. At its base, authoritative data are data produced by an authoritative source, which is an entity authorized by a legal authority to collect specific data.[74] For example, in the United States the authoritative source for geographic place names used by Federal Agencies is the United States Board on Geographic Names (BGN), a body created in 1890, under the authority of the United States Government to maintain uniform geographic name usage throughout the Federal Government. While Google or OSM can name a place in the United States anything they like, there is an official name authorized by the BGN.

It is important to note here that numerous sources consider their data to be authoritative, but in many cases these claims are relative to specific missions or tasks, rather than being based on the authoritativeness of their data collection efforts. An example would be local taxation offices that have been authorized to assign street names within their communities in order to track the addresses and property parcels located within the jurisdiction of that community. While the local government may be the authority responsible for setting the street name for local use, the addresses of properties for purposes of postal delivery are set by the authority of United States Post Office, as is the spelling of the street name for purpose of postal operations. However, if the town has an address and spatial database management system, its rendering of the street name and preferred address are authoritative for its purposes (taxation, property mapping, E-911) and potentially in conflict with the authoritative name used by the USPS. In essence, the use of data sources requires a judgment to be made by the user as to which authority they wish to certify for their intended use of exchanged data.

[74] Stage, David. 2009. Authority and Authoritative Data: A Clarification of Terms and Concepts. Fair & Equitable, February 2009 pp13-16. See http://www.iaao.org/uploads/Stage.pdf

In general, the concept of authoritative data sources for address and spatial data rests on the belief that local sources usually provide the most accurate and up-to-date local information. However, it is clear that gathering some types of spatial data is beyond the capabilities of local governments and, in these cases, regional, state, or even national governments may become the authoritative party mandated to collect data of a specific type.

In some situations, regional, state, or national agencies need to gather and curate the "authoritative" data collected by local sources and may create a clearinghouse function to manage the distribution of these data. According to David Stage, a trusted source is one that publishes data gathered from authoritative sources. Sources that compile and publish data from numerous, local, authoritative sources are considered trusted when there is an official, documented process for compiling data from these disparate sources and resolving differences when the sources conflict.[75] Data that are specified as fit for a specific use are usually considered certified data, a classification that implies compliance with a standard setting body or that the data meets other legal requirements set for data intended for specific uses.

Organizations such as the Census Bureau that are required to or choose to work with data sources representing a local, tribal, regional, state and national sources must make a determination of the integrity of the data source with which they are working. Resolving the integrity of sources is aided when metadata at a feature level is available, since this can be used to examine the provenance of any feature that is in dispute with a record from another source.

Because the results of the Decennial Census are used for the apportionment of Congressional seats, the appropriation and distribution of Federal funds to state, tribal and local governments, state funds distribution, as well as other "official" uses, it is critical for the Census Bureau to be able to evaluate the authoritativeness of any source with which it exchanges data. In fact, although the Census Bureau operates as a federal level, as a practical matter, it must regard itself as authoritative in respect to census geography and the local addresses and spatial data which belong to and define each unit of census geography.

The role of the Census Bureau in data collection is influenced by the fact that it must consider its data to be authoritative when exchanging data with sources whose governments could benefit from the address and spatial data that they exchange with the Geography Division. As a consequence, the Geography Division must attempt to determine the authoritativeness of all data sources, as well as develop systems that can compare imported data with Geography Division data and ensure that these data meet the business rules set by the Geography Division for the use of its address and spatial data. It is for this reason that successful data exchange, in large part, requires setting a robust institutional environment.

[75] Stage, David Ibid.

3.b Partnerships and Institutional Framework

Over the past two decades as widespread use of GIS has expanded from the Federal to state and now local governments, there has been an evolution in the establishment of data sharing partnerships. A cornerstone of this evolution was the partnership agreement between the Census Bureau and the USGS to use of digital versions of the USGS topographic quadrangles as the basis for TIGER. Over time organizations moved beyond the conversion of existing maps to the development and maintenance of geospatial databases from primary image and field-based sources. Within the Federal government OMB circular A-16 established policy regarding the coordination of these data collections under the Federal Geographic Data Committee. President Clinton's 1994 Executive Order established the National Spatial Data Infrastructure (NSDI) that embraced the desire for multi-level data coordination:

> Building an effective NSDI will require a well-coordinated effort among Federal, tribal, State, local government, and academic institutions, as well as a broad array of private sector geographic, statistical, demographic, and other business information providers and users. Only through this cooperation will the NSDI become a reality.[76]

Following the initial creation of TIGER, the Census Bureau recognized that it needed to embrace the concept of the NSDI to maintain the data. In many ways, LUCA is the best example of the NSDI in action. Through this program a "coalition of willing partners" provides high resolution, current data to report additions, deletions, and corrections to TIGER. Nevertheless, even though there obvious incentives for participation in LUCA, only 8,185 of the local governments who were invited actually returned any data (approximately 21 percent). Even with the incentives associated with accurate Decennial Census tabulation, it will be a real challenge for local governments to share their data resources. In other words, the LUCA framework may not be the type of partnership that will enable the GSS Initiative

Since 1990, with core support from the Census Bureau, the Mapping Science Committee (MSC) of the National Research Council has provided external evaluation, guidance and recommendations on several aspects of the NSDI (NRC 190, 1993,1994,1995,2001, 2003 and 2004). Many of these efforts have focused on the interrelationship of the components of the NSDI and the improvement of partnerships (Figure 34).

[76] See http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf

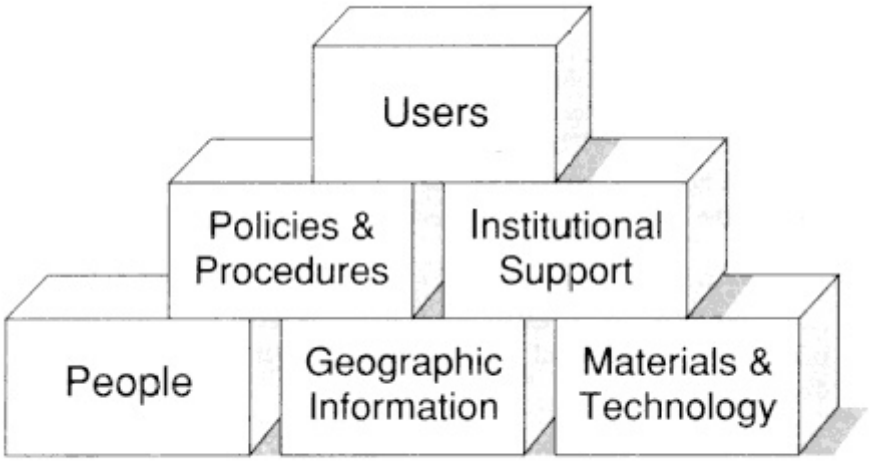


Figure 1. NSDI building blocks.

Figure 34. NSDI Building Blocks (NRC, 1994)

The importance of incentives in a successful partnership was addressed in the NRC report *Promoting the National Spatial Data Infrastructure Through Partnerships*. (NRC, 1994). The report provided a status report on state - federal partnerships and concluded that:

- Viable partnerships will require focal points within the federal government
- Clear guidelines for cost sharing and partnerships need to be developed
- Involve states in standards setting
- Incentives are needed to encourage partnerships to maximize use an benefits to the broader user community
- The FGDC should investigate the extent to which federal procurement rules are an impediment to the foundation of spatial data partnerships.

A companion report published in 1995, *A Data Foundation for the National Spatial Data Infrastructure,* identified specific data themes that are the foundation of a successful national geospatial database. The report focused on the transportation theme and offered specific recommendations regarding TIGER:

> "One of the more prevalent forms of transportation data are street centerline spatial data (SCSD), are basically computerized street maps, where streets are represented as centerlines to which attributes of the streets are appended. Almost three decades of practice have proven the value of differentiating between the left and right sides of each street segment and encoding attributes to them such as street names, address ranges, ZIP codes, census and political boundaries, school districts, traffic zones, and congressional districts. Practice also demonstrates the value of including and reconciling non street features in these data bases to form topologically consistent blocks (e.g., water bodies, railroads, political boundaries). The Bureau of the Census has incorporated some of these street features into its TIGER data bases…. SCSD provide a good example of a framework spatial data theme by virtue of their extensive current use in facility site

> selection, census operations, socioeconomic planning studies, legislative redistricting, and logistical operations management. Present and near-term developments in personal computers and consumer electronics may expand SCSD use to virtually all citizens in trip planning, route guidance, and electronic atlas applications. The digital data requirements for the Intelligent Vehicle Highway System will likely dictate higher accuracy and more comprehensive attribution than most SCSD developed for GIS applications. SCSD are important because they express fundamental relationships between street addresses (the most common spatial reference for the built environment) and coordinates and other locational links ("geocodes") mentioned above. SCSD are widely used to link street-addressed data to geographic references for GIS and other desktop mapping applications."

The Mapping Science Committee offered the following observations and recommendations regarding TIGER:

> "The TIGER files, particularly the attribute data, serve as a major stepping stone toward a mature SCSD for the nation. However, TIGER files lack accurate coordinates registered to the foundation, complete street addressing, and an ongoing maintenance program. The TIGER files could be integrated with the foundation by the following actions:
>
> - improving coordinate accuracy using ortho-rectified imagery that is tied to the geodetic control network;
> - completing and improving street and address coverage in partnerships with the U.S. Postal Service, 911 emergency agencies, state and local governments, and the private sector; and
> - establishing an ongoing update facility employing local government partnerships for timely information (transactional updates) about new streets."

The report also commented on funding for TIGER improvement:

> "These deficiencies of the TIGER files are recognized by the Bureau of the Census. The importance and timeliness of the societal demand for accurate transportation data are reflected in Executive Order 12906, which calls for provision of such data in order to support the decennial census for the year 2000. Funds should be identified to allow these improvements by January 1998, which is the target date specified in the executive order."

From a historical perspective the 1995 report laid the foundation for the TIGER accuracy improvement program and the lineage of the GSS Initiative.

As noted elsewhere in this report it is interesting to note that the Department of Transportation is the steward for the transportation theme. However, the new supplemental guidance for OMB Circular A-16 expands the roles and responsibilities for participants in the stewardship of each NGDA theme to include:

> Executive NGDA Theme Champion
> NGDA Theme Lead
> NGDA Dataset Manager

Data Steward

We understand that DOT and the Census Bureau have met to discuss the role of TIGER as the official source of street center lines. The GSS Initiative would strengthen this role and firmly establish as a central part of the Framework. This has direct implications for the National Map and other themes.

General approach to data sharing

As noted above, data sharing is critical to the success of federal geospatial data programs. Any discussions of imagery, parcels, elevation, or a comprehensive GIS "for the Nation" are built on multi-tiered governmental partnerships. It must be emphasized that data sharing involves the successful transmission and utilization of data that becomes information. The framework must address the following issues:

1. Procedures to discover potential data providers
2. Partners with accurate, current and authoritative data
3. Documentation of data resources
4. Willingness to share
5. Absence of legal and financial barriers
6. Technical ability to transmit the data
7. Technical ability to receive the data
8. Ability to utilize the data in a data maintenance environment
9. Constraints imposed by Federal IT security regulations and requirements
10. Constraints imposed by data confidentiality (i.e. Title 13)

In *Promoting the National Spatial Data Infrastructure Through Partnerships* (NRC, 1994), the following key elements were identified for a successful partnership:

> **Shared Responsibilities**. The parties to a partnership will have made a formal agreement that defines each party's responsibilities in the activity, in the form of a memorandum of understanding, contract, or other binding document.
>
> **Shared Costs**. The costs of the activity will be shared between the parties according to some agreed formula.
>
> **Shared Benefits**. Each party to the activity will derive some benefit that is consistent with its mandated role as an agency. In addition, benefits from the partnership will likely accrue to parties outside the partnership and to society at large. This is especially important when the partnership is seen as contributing to the evolution of the NSDI.
>
> **Shared Control**. Decision-making control of the project will be divided between the participants.

These issues were further explored in *National Spatial Data Infrastructure Partnership Programs: Rethinking the Focus* (NRC 2001):

- Benefits of spatial data partnerships must be evaluated for the entire national community of spatial data users, not merely for the agencies participating in the partnership.
- The contribution of spatial data partnerships to the wider objectives of NSDI must be considered in its design and management.
- Data quality is an important factor in the value of any investment in spatial data. Potential users will be confident using data only if they know the data are reliable.
- Stewardship is a key concern in reaping the benefits of investment in any spatial data partnership; the agency closest to the source of the data is likely the organization best able to maintain the data.
- An essential element of the NSDI partnership model must be a commitment to support the partnerships as part of an ongoing program. Long-term commitments will help insure data are maintained and that mutual trust in the ability of the partners to meet respective needs will be achieved.

It should be noted that adoption of metadata standards, the FGDC/ NSGIC Fifty States initiative, Geospatial One Stop, the NSGIC GIS Inventory, and software tools have addressed many of these issues and actually eliminated others. At the same time, Internet based data exchange mechanisms, such as FTP have made it possible for willing partners to simply "throw the goodies over the fence" and allow others to discover and acquire them without any formal agreements or financial transactions. As noted elsewhere, some organizations publish new versions of their data on a weekly basis or send an email alert when new versions of data are available. We believe that through peer group pressure and the need to demonstrate the value of their local programs an increasing number of governmental agencies are embracing data sharing to meet the demands of an increasingly geospatially savvy public and to counter the perception that Google and Microsoft are the best source of address and spatial data.

Marketplace

While many of the technical obstacles to data sharing have been eliminated, an entangled network of institutional and financial obstacles will impact the success of GSS Initiative in many parts of the nation. For example, while we highlighted the success of IndianaMap, it should be noted that seven of the 92 counties are not participants. The NRC published an extensive review of institutional, legal, and financial obstacles to data exchange in *Licensing Geographic Data and Services* (NRC, 2004). The report discusses the concept of a data market place (figure 35) and reviews issues regarding copyright and FOIA. The authors note:

> "Compared to civilian federal agencies, state and local government units (1) are more fragmented, (2) collect more domestic geographic data, and (3) have proportionately smaller data budgets. Conditions 1 and 3 increase the importance of interagency cooperation. Not surprisingly, state and local governments have developed various institutional frameworks to facilitate joint acquisition and sharing of geographic data:
>
> - *Ad Hoc Collaboration.* Geographic data swapping is an increasingly common form of local-to-local transaction. Additionally, local agencies may pool resources with federal, state, or other local agencies, or with private-sector partners to collect new data.

- *Organized Collaboration.* In principle, state and local agencies can buy and trade geographic data among themselves through "arms-length" transactions on the open market. In practice, many governments prefer to participate in regional entities where data are shared. Examples exist in Maine, Maryland, Kansas, and Illinois. Some observers claim that communal organizations are more efficient than market- and license-based mechanisms for small, tightly knit user communities. These organizations are less practical when disparate mandates and political structures divide potential collaborators.
- *Umbrella Organizations.* State and local governments increasingly rely on networks and partnerships to facilitate sharing. For example, approximately 100 state, federal, nonprofit, and academic entities in the Minneapolis/St. Paul area have created Metro GIS. Metro GIS facilitates data sharing among members, licenses data to outside users, provides a forum for exchanging best practices, and helps members coordinate their respective data collection programs.
- *Contract Work.* Some agencies perform data development and processing services under contract to other government entities. This encourages agencies to pay close attention to what their users want.
- *Agency Assessments.* Maine, Michigan, and Kentucky fund substantial geographic data development and project support through voluntary assessments on selected state agencies. This model provides a strong incentive to pay attention to user needs in order to obtain new assessments in later years."

They concluded:

> "State and local governments are both suppliers and consumers of geographic data. Not uncommonly, they rely on revenue from licensing their geographic data to recover some of their costs, which may limit federal agencies' ability to acquire unlimited rights in the data. The public disclosure requirements of federal law may inhibit state and local participation in partnerships to acquire geographic data."

They recommend the establishment of a National Commons in Geographic Information

> "Where citizens can post and acquire commons-licensed geographic data. The proposed facility would make it easier for geographic data creators (including local to federal agencies) to document, license, and deliver their datasets to a common shared pool, and also would help the broader community to find, acquire, and use such data. Participation would be voluntary."

42 LICENSING EXPERIENCES

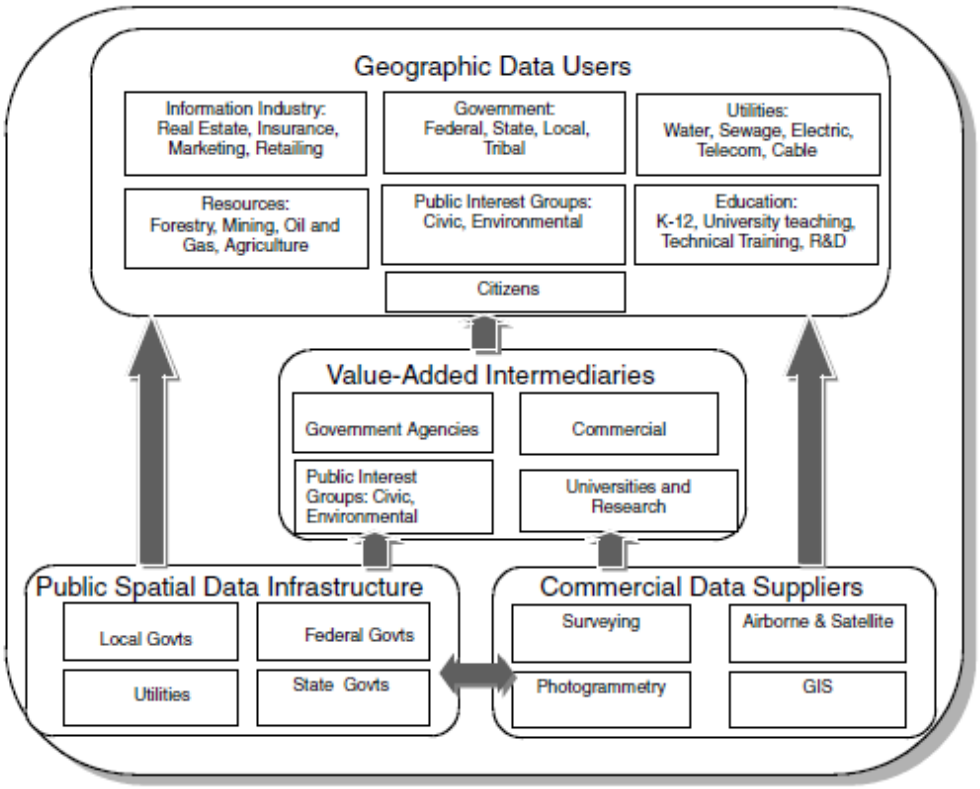


FIGURE 3-1 Geographic data marketplace. The concepts of the *geographic data market* and the *public spatial data infrastructure* are distinct. Government resources provide an infrastructure from which the marketplace for geographic data products and services emerges. Government also plays an important role inside the marketplace through its procurement of geographic data products and services. SOURCE: Adapted from: X. R. Lopez, 1996, Stimulating GIS innovation through the dissemination of geographic information, *Journal of the Urban and Regional Information Systems Association* 8(3), 24–36.

Figure 35. The market place for address and spatial data.

Title 13

A major issue relating to data exchange between the Census Bureau and its partners are the restrictions placed on sharing the MAF with local governments. In effect, the restrictions under Title 13 of the US Code of Law make the exchange of address data one directional. In 2007, the NRC published an extensive analysis of the federal need for parcel level data. *National Land*

Parcel Data: A Vision for the Future directly addressed the impact of these restrictions and offered a specific recommendation regarding the policy. The authors noted:

> "While the LUCA program allows the Census Bureau to benefit from local review, local governments that provide the feedback do not receive comparable benefits. The restrictions on use of the MAF mean that new addresses included in the MAF cannot be incorporated into local databases. This restriction means that local governments must function with what may be erroneous or missing data. This limitation could actually result in the loss of life or property and discourages many local governments from providing feedback to the bureau.
>
> The procedures that the Census Bureau must follow are designed to ensure the highest level of confidentiality for individual responses. Many communities have requested some of the information resources that the bureau uses to conduct the decennial Census. In 1982, one of these requests led to Supreme Court decision in *Baldridge v. Shapiro* (455 U.S. 345, 1982). In that case the Supreme Court held: "The master address list sought by Essex County is part of the raw census data intended by Congress to be protected under the Act [Title 13]." In that decision the Court endorsed the Census Bureau's reading of Title 13. It wrote: "The unambiguous language of the confidentiality provisions, as well as the legislative history of the Act, however, indicates that Congress plainly contemplated that raw data reported by or on behalf of individuals was to be held confidential and not available for disclosure."
>
> While the committee appreciates the bureau's need to comply with the directives of Congress and the Supreme Court it believes that the address points are actually just an enhancement to the TIGER line file based address system. It is important to note that since the 1970 decennial Census the Census Bureau has always distributed GIS files with street names and address ranges that facilitate automated address matching. In fact, it provides public domain software (Landview) that enables a user to locate an address and associate census information and other themes to that location. It could be argued that the improved point-level locations for addresses are simply an enhancement of these address matching capabilities and have nothing to do with confidential information about the residents of the dwelling units. This issue is of relevance since these expensive and valuable point-level features could provide an excellent tool for creating a national parcel database."

The committee summarized the issue as follows:

> "The Census Bureau is currently modernizing the Topologically Integrated Geographic Encoding and Referencing (TIGER) system of digital street data for the 2010 Census. It has worked to align TIGER streets and blocks to the same data used by local governments. It is also creating a point level representation of properties with associated street addresses. Several commercial companies are doing exactly the same thing. While these companies will lease their data, the Census Bureau is prohibited by Title 13 of the United States Code from sharing these data with other federal agencies or with the local governments that provided much of the information. Since addresses and their location

> are publicly available information, the ability of the Census Bureau to release just building address point locations could serve a multitude of uses and would have major economic benefits while not revealing confidential information about individuals. The availability of address points could dramatically improve emergency 911 systems across the nation and provide a starting point for parcel data in rural parts of the country."

It offered the following recommendation

> ***Congress and the Bureau of the Census should explore potential policy options, including modifications to Title 13, that would allow its digital data on building addresses and their geographical coordinates to be placed in the public domain while also maintaining important privacy protections. If publicly available, these street addresses and coordinates could be used to assist in the development of parcel data in areas where parcel data sets do not exist.***

It is noted that this recommendation has been endorsed by the NSGIC and NGAC among other notable organizations.

Summary

We believe that effective and efficient data exchange mechanisms are critical to the success of the GSS Initiative. Over the past two decades there have been major technical and institutional advancements that enable partners to discover and share geospatial data. Although the LUCA program represents a pioneering effort in the development of data exchange mechanisms, participation has been poor and new models are needed. Since the GSS Initiative does not carry the same constitutional clout as the Decennial Census, we believe that potential partners will have few incentives to participate. This calls for innovative programs that minimize the burden on local partners. In effect, if a local or state partner provides access to a current version of its data the Geography Division must be prepared to harvest the data, identify changes, and edit MTdb. Furthermore we believe the current Title 13 restrictions are an obstacle to participation in LUCA and will continue to be for the GSS Initiative. Therefore we support the recommendation of the NRC committee to re-examine the current policy and work toward a system of true bi-directional data exchange.

3.c Data Exchange Models

We note at the start of our discussion of data exchange models that it is likely that the sharing of address data would be one-way; that is, designed to provide for the export of address data to the Geography Division by its partners. Although we can envision methods that would allow the Geography Division to share its address information with approved or certified external partners for editing, the problems associated with the Census revealing address data for anything other than official purposes, is an insurmountable obstacle under Title 13. As a consequence, we will focus our comments on addressing scenarios in which partners provide address data to the Geography Division, but do not directly receive address data from the Geography Division that could be used for tasks other than preparing for the exchange of edited data. Conversely, there appear no legal limitations on the two-way exchange of non-address, spatial data and we will

focus on a variety of strategies that the Geography Division could use to exchange point, line, polygonal and attribute data with its partners.

In examining the problem of data exchange, we concluded that the Geography Division may need to consider three distinct forms of data interchange and, perhaps, three pilot programs to explore these strategies. The three exchange mechanism models that we will discuss in this section of our report are based on a common approach. Although each model ingests data in a slightly different fashion, all three models are based on the assumption that the data provided to the Census Bureau must be evaluated before they will be accepted for integration with the MTdb.

In a sense, our exchange mechanism models are pure capture mechanisms and reflect our belief that the complexity of direct exchange and integration of partner data with the MTdb, as described in section 2 of this report, make it unlikely that the Geography Division will find partners willing to exchange data in a direct manner. The systems, data schemas, approach to topology, and business rules used by the Geography Division are uniquely suited for the data handling problems that need to be resolved in order to support the strategic responsibilities of the Census Bureau. However, it is unlikely that partners willing to directly exchange data with the Census Bureau could meet the requirements posed by the Geography Division's data handling systems. Other than government agencies at the Federal level, we do not believe that states, local or tribal governments could afford the investments in technology, systems development, system design, database design and software development that have been supported by the Census Bureau to build the systems to process address and spatial data.

It is more reasonable to assume that the Geography Division should expect to capture data in diverse formats and translate and evaluate the data captured from its partners rather than directly exchange it. We envision a series of exchange mechanisms that involve the participants interacting with a Product Data Base extracted from the MTdb and accessed by the data exchange participants using platforms otherwise unrelated to the MTdb and the data handling environments normally used by the Geography Division. Once the data is exchanged, regardless of the mechanism, the processing of this data would essentially remain the same.

Our model for the data exchange model is shown in Figure 36 and described below.

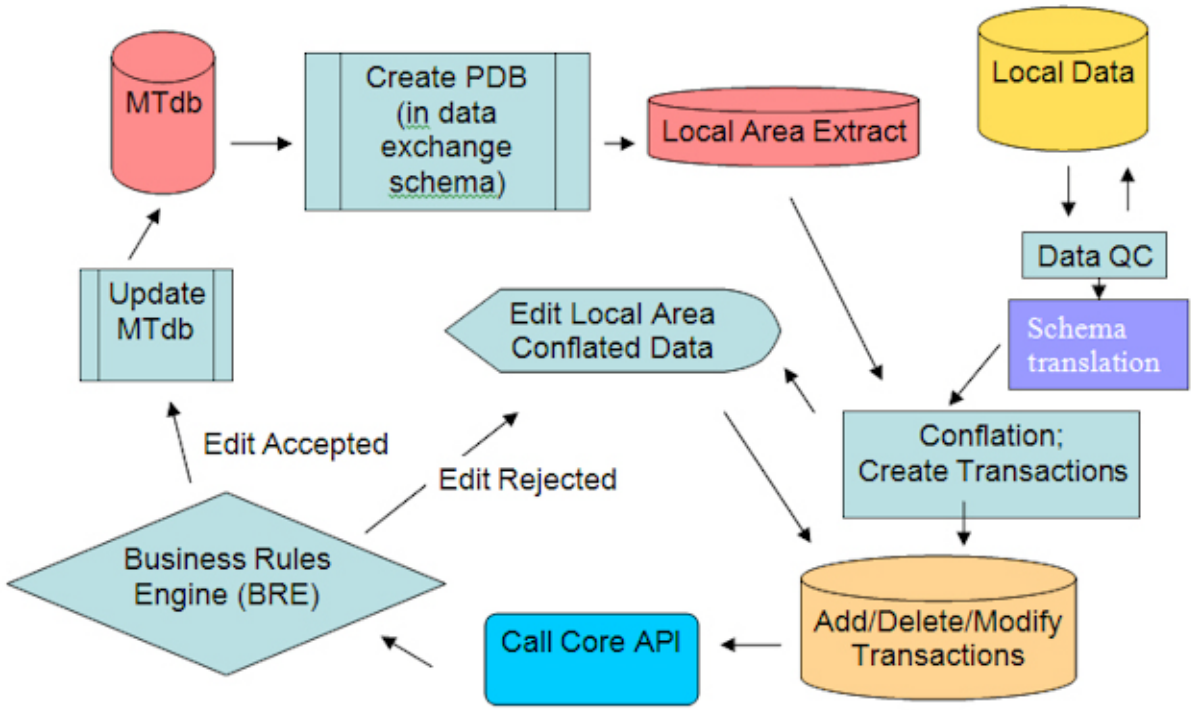


Figure 36. Generalized Data Exchange model and process

Our recommended, generalized model involves several steps:

1) From the MTdb, produce a Product Data Base (as described by Danso, 2010). The PDB should have a precise set of specifications and be well defined; the PDB should also be designed to incorporate additional fields for data quality information to be associated with each feature. Information on the currentness (date) and positional accuracy of each feature is particularly important.

2) Extract the coverage area (e.g. county) for the exchange

3) Obtain data from local authorities; collect all available specifications; schema definitions; metadata information for the dataset

4) Run the local data through data quality review software, checking for logical errors, errors in attribution, and whatever checks the incoming data schema may support. If errors are detected, Census may choose to send the data back to the provider for them to edit/correct (presumably they want correct data, so they may be willing to do this)

5) If the data schema of the local data differs from that in the PDB, then schema translation software may have to be used to make the local data compatible with the PDB.

6) Run conflation software that matches the PDB extract with the local data. The result will be a file of adds/delete/modify. Some results may not be able to be resolved automatically and will require manual editing for resolution.

7) Add/delete/modify actions submitted to Core API (in a process similar to BFUS). Most errors should have been corrected in step 4, where the data quality software can perform many of the logical checks performed by the BRE

8) If some transactions fail in the BRE, then those need to be examined in an editing session and new transactions are generated for submission to the BRE.

9) Updates are incorporated into the MTdb.

10) New edition of PDB can be generated for use by other Census operations (e.g. product generation). Cycle repeats.

Having established a proposed environment for processing the exchanged data, let's look at the platforms that we recommend as candidates for exchanging or acquiring data from partners who want to exchange data. We describe three potential models and recommend that the Census consider each for testing as a pilot project. Note that regardless of the form of implementation, the input data is sent to Census for its consideration in adding it to the MTdb, rejecting it, or marking it for follow-up research.

Model 1. Asynchronous Harvesting

We suggest the adoption of a harvesting model based on our belief that direct data exchange is too complex technologically, too time consuming, and too expensive for most partners to participate in a manner that would support the GSS goal of continuous updating. In this model, the Geography Division harvests address and spatial data from state, regional, local or tribal partners and partners harvest data from a PDB derived from the MTdb on an "as needed" basis. (See Figure 37 for general details of this model.)

In effect, this model assumes that the Geography Division will have to harvest the data that it is interested in examining for inclusion in the MTdb, but it will have to do so based on the terms of the partners and not require any special handling on their part. Instead, the Geography Division will harvest data in the form provided by state, local and tribal governments based on the update cycles of the local sources. Many communities are now providing RSS feeds to advertise that updates have been posted to their database and are now available for downloading.

While the harvesting model will require monitoring data sources and the availability of updated data by the Geography Division, doing so could result in an increased frequency of updates to the MTdb from reliable sources.

Note that the harvest model is time-sliced with the frame on the right (representing a future time when the partners can harvest TIGER) occurs after the Geography Division has harvested the

partners data, processed it and decided whether or not to include in the MTdb. Downloading the Geography Division's data will allow the data partners to see which data have been accepted by the Census Bureau and could be used to provide an automated dialog about rejected data, allowing clarifications of to be submitted, if either the partner or Census Bureau wishes to pursue issues of this type. We suggest the addition of a Wiki for comments as described by Hall et. al. (2009) that could be used as a method for authorized partners to add notes or comments on specific addresses or geographical features.

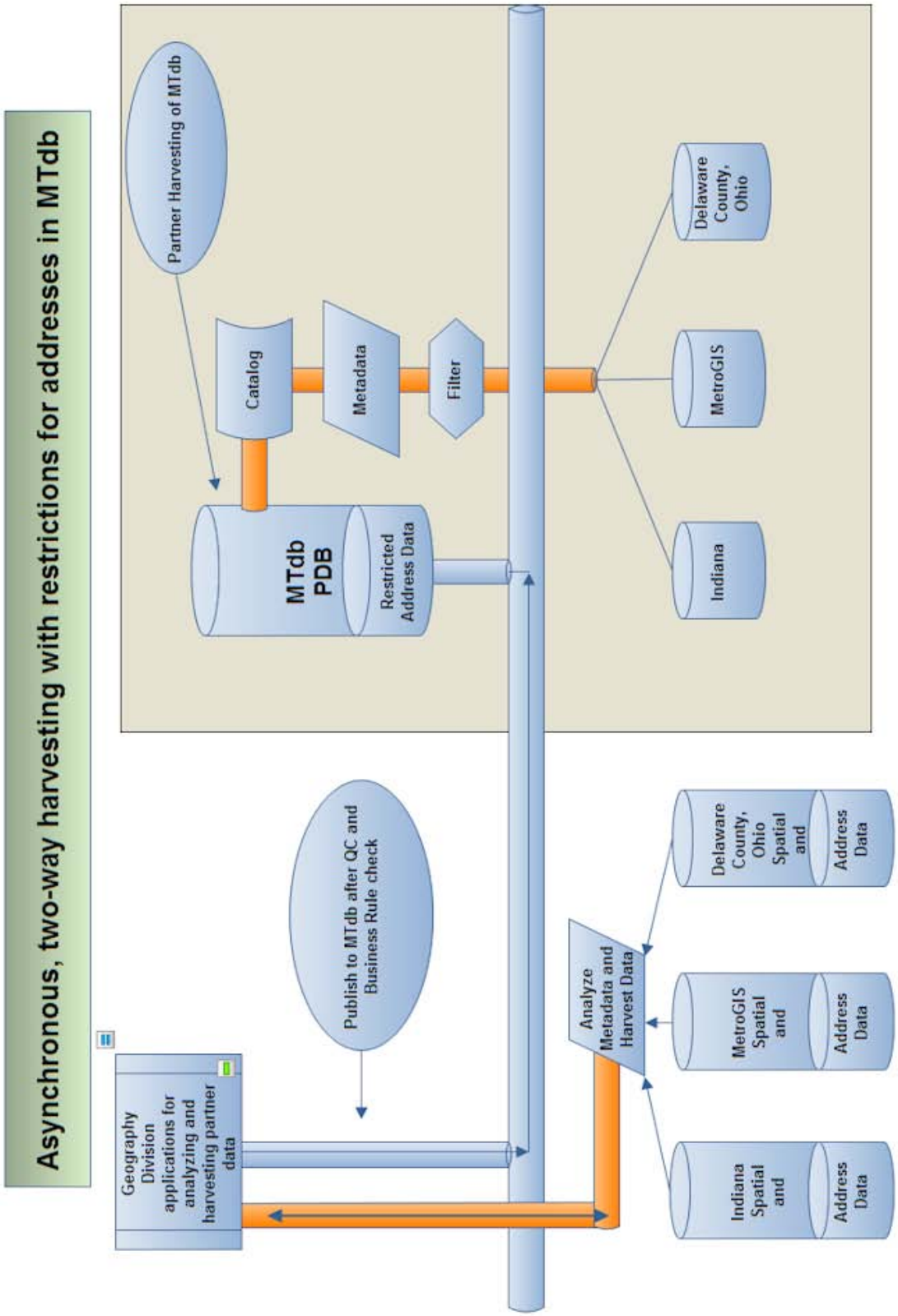


Figure 37. Example of a two-way harvesting model with participant harvesting delayed, allowing Census to integrate the data it collected from its data partners.

Model 2. OSM-Like One-way Exchange With Credentialing

We believe that the Census Bureau, due to legislative requirements of the Census Address List Improvement Act of 1994, will need to provide an exchange process that mirrors LUCA 2010. Our second model performs that function in an online environment. In our model (see Figure 38) the online exchange is one-way (participants provide their data to Census Bureau) and allows credentialed participants to access their "piece" of a TIGER PDB using a web browser. Although certified users (those who have agreed to observe Title 13 restrictions regarding the confidentiality of the data) will be able to download the address data for purposes of editing, they will not be allowed to use the data for any other purpose, as was the case in LUCA 2010.

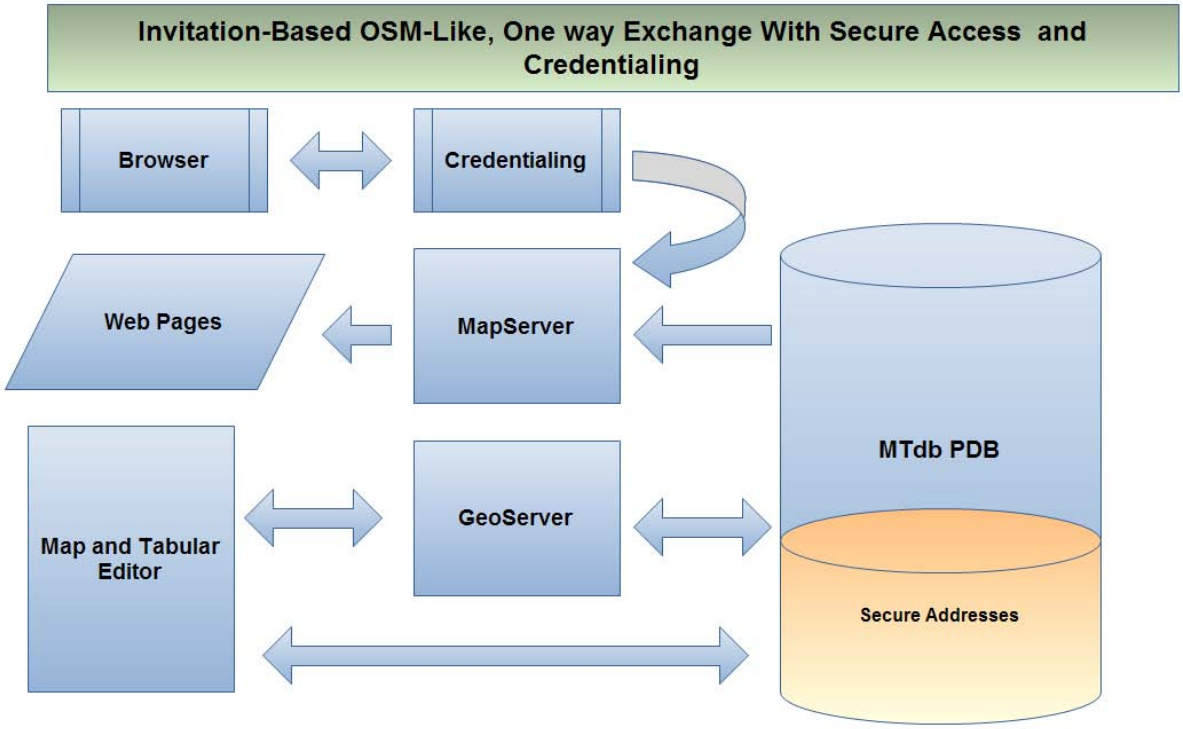


Figure 38. Based partially on the OpenStreetMap model of data exchange and integration, this model requires credentialing to ensure that parties see and can alter only the address and spatial data that applies to the geographic area they represent. Editing address data requires agreement with the limitations posed by Title 13.

The process shown in our model requires the Geography Division to provide an extract of MTdb in the form of a PDB as the core database to which all participants will react. Each participant providing acceptable tokens for credentialing would be allowed either to edit the data online or to download it and edit the shapefiles with their GIS and upload their changed data to the PDB. The online editor provides for map editing (add, delete, etc.) and tabular editing of the address data provided by the Geography Division. The map editing would also include a provision allowing the participants to provide address points. It is our belief that the address points provided by the participants would not be entitled to protection under Title 13.

Model 2 requires the implementation of a common technology and the selected MTdb content represented in the PDB needs to be defined by a template (that starts with collection rules and includes feature definitions, feature instance rules, etc.). Local authorities can add/edit data

(using web based editing tools) in the PDB, but their contributions must conform to the rules imposed by the schema template. This forces the data that is provided to be consistent with all other data. The model requires that the Geography Division establish a data standard whose terms, definitions and formats are published and known by the exchange participants. The new data contributed by the partners to the PDB would need to be processed through the generalized model for data exchange and processing that we described previously. (See Figure 36.)

The model would appear to serve for the yearly BAS submissions and allows the participants to provide address data at these times if they elect to do so. However, once a decade the platform would be used for a LUCA-like address and data exchange, although this process could also be conducted through a version of the MAF/TIGER partnership software used in 2008 and updated in 2010 for BAS for those who do not want to deal with the online model described here.

Model 3. Community–Based, OSM-like Data Exchange

If the Census chose to build Model 2, it would, in the process, have developed the technology stack to create a crowdsourced version of the model (absent the credentialing). (See Figure 39.)

The technology stack could be built using Esri software, or a combination of the open source technology that is currently available. Our recommendation would be to build the system using an OSM-like technology stack, which would have the benefits of being open and not require Esri products on the user side of the transactions, although there are definite benefits to an association with Esri. For an example of a possible technology stack to support this model, see Table 10. MapQuest in its beta sites uses the following for its technology stack: Mapnik, TileCache, Potlach 2, and Nominatim.[77]

The model allows contributors to add, delete, or edit data and to extract large chunks of TIGER (NOT MAF) from a PDB for mapping or other purposes. Input could be in the form of on-screen digitizing or by import of GPS tracks. The inclusion of "Planet TIGER PDB" is a nod to the fact that users may want to download large files and it is easier to manage the system if this database is separated from the editing operations.

[77] Valuable information about MapQuest's work with OSM can be found in the Developer Blog http://devblog.mapquest.com/author/apegg/

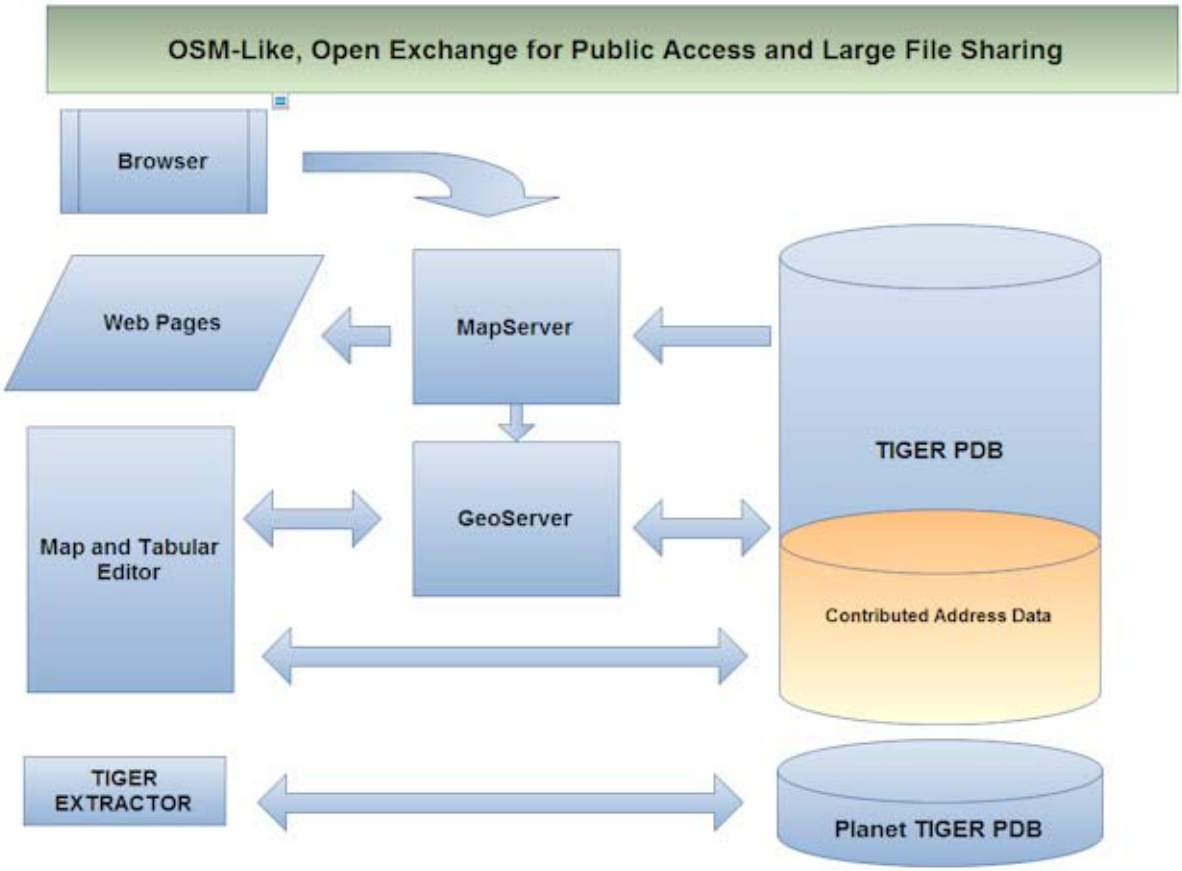


Figure 39. The model above is a rendition of the architecture of crowdsourced map databases. In this case, we provide a "Planet TIGER PDB to provide for large file sharing by the public. MAF data are not part of this model, although address data contributed by the public users might pass the Title 13 barrier.

Table 10. Example Open Source Technology Stack for Mapping Systems

Web Front-end	Open Layers
Web Service	WMS/WFS
Map Server	GeoServer
Web Server	Tomcat/Java
Database	PostGIS/MySQL
os	Linux
Edit	uDig

Summary

The information described in this section of the report has served to identify the relevant components required for the successful exchange and integration of address and spatial data between the Census Bureau and its partners. Exchange and integration are complex tasks but technology is emerging that is beginning to simplify the exchange portion of this equation. In turn, the development of geoportals based on the emerging open technology is causing a sea level change in the ease and frequency with which address and spatial data can be exchanged and updated. Integration between data sets remains a difficult issue that will become increasingly important as increasing amounts of data are exchanged more frequently. However, spatial ETL tools, such as FME[78], now allow the relatively rapid conversion and integration of data in over 250 formats.

The models we have described allow Geography Division to manage its content leveraging existing architecture while separating incoming data from the MTdb until the data have been evaluated as fit for the uses required by Census. Another advantage is that the external partners will be version independent since the operating software is on the Census side of the data transfer. Further, each of these models is driven by the need to be responsive to the value of a data partner's time and efforts. The final advantage of our proposed architecture is the time and cost benefits resulting from the numerous, reusable, components shared by two OSM-like models.

A variety of models exist for the exchange of data, but Title 13 restrictions on addresses burden the Census Bureau and limit its avenues for exchanging data. We recommend the Census pursue an examination of the three generalized models for data exchange that we described, as these inelegant solutions represent mechanisms that could quickly and easily help improve the quality of the MTdb and help meet the requirements of the GSS for continuous update.

4. Data Exchange Pilots

As noted in other sections of this paper, the concept of two-way, direct exchange of address and geospatial data is likely to be an unsatisfactory strategy in the future. More and more sources are putting their data on the web and making it openly available to third parties in an attempt to meet their responsibilities to their communities and reduce the cost, complexities, and headaches associated with a direct interchange of data. Exchange models that let data producers publish and exchange data on their own terms, while allowing interested parties to collect, evaluate, and integrate these data, will become the standard for data interchange in the near future, if they are not already the dominant form of exchange. This is not meant to imply the commercial data providers, or agencies that currently generate revenues from licensing data, will not continue to rely on direct or indirect but controlled exchanges to manage access to and use of their data. However, we believe these sources will eventually abandon direct exchange unless their data is so unique, valuable, or sensitive that it cannot be shared with the general public.

[78] See Safe Software http://www.safe.com/

We recommend that the Geography Division consider the three exchange models described in the last section of this report as possible candidates for pilot projects testing data exchange.

Pilot 1. Implementation of the Harvesting Model.

We believe that implementing a pilot project based on the "harvesting model (See Figure 37) would require little in the way of technology development if it utilized the Generalized Data Exchange model described in Section 3.c of this report (see Figure 36). In effect, the Geography Division's existing technology appears adequate to integrate information that was harvested from sources on the web.

What would be required is the development of a strategy to discover information about web services that publish authoritative or trusted address and spatial data. We have provided reviews of data sources for address and spatial data in this report, as well as in Deliverable 2 on addresses and addressing. We recommend that state-based geographic portals described in section 1 of this report as a good starting place for this effort. Other sources of local geographic information are certainly familiar to the staff of the Geography Division based on their experience with LUCA and BAS, as well as the sources that provided information on addresses and spatial data for the MTAIP. Other sources could be found using standard web services or developing or adapting web catalog servers and semantic web technology tuned to the needs of the Census.

Implementing this model could be done without a specific pilot project partner since web harvesting of data relies on the contributors making their data freely available for others to process, requiring no negotiation, memorandums of understanding, or legal entanglements. However, we suggest that the State of Indiana, MetroGIS, or Delaware County, Ohio (groups we have described in detail in Section 1 of this report) would be useful sources of address and geographic information with which to test the architecture with which the Geography Division intends to harvest address and spatial data. In addition, the Geography Division should consider using this model to harvest data from OSM (through PlanetOSM), as well as from OpenAddress.org.

Our proposed timeline for implementing the Harvesting Model is provided in Figure 40.

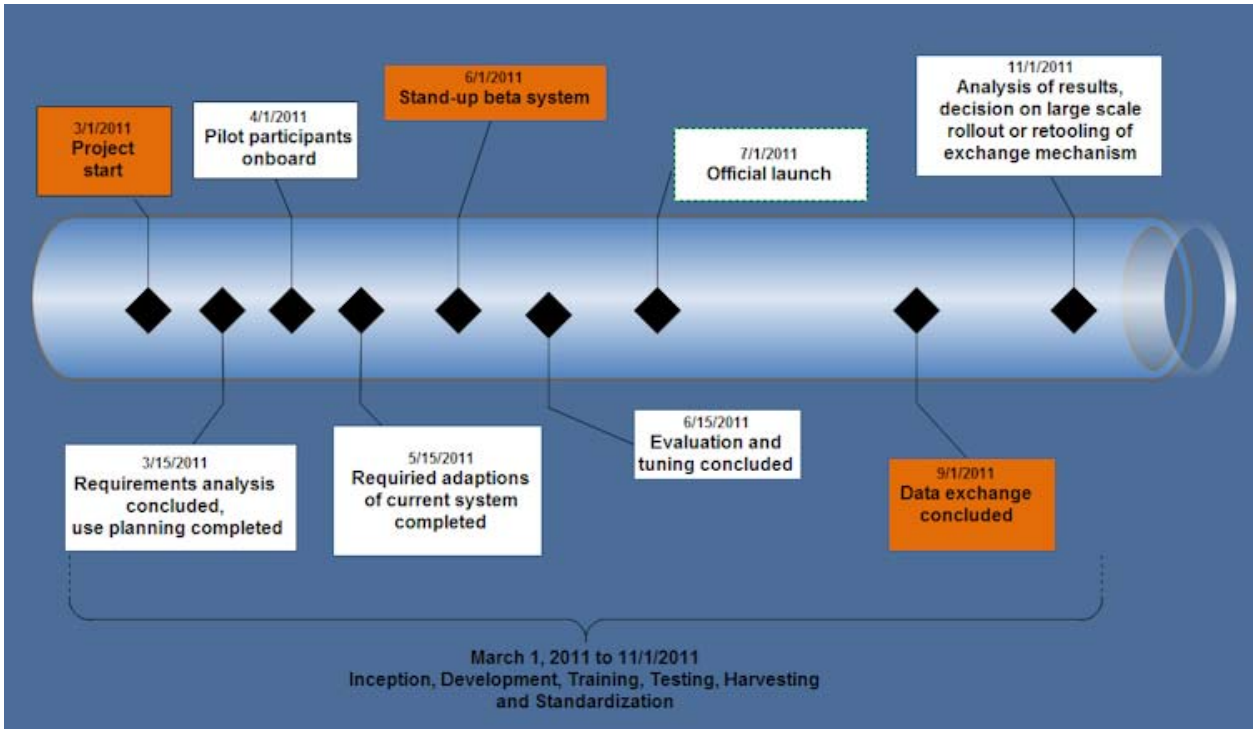


Figure 40. The Harvesting Model timeline is short since their amount of software development should be minimal. In turn, we believe that the know-how, tools and procedures needed to ingest, evaluate and integrate the data discovered already exist within the Geography Division.

Although the schedule is aggressive, we believe that the challenges in implementing this model are minimal, while the payoff could be significant in terms of its benefit for continuous updating.

Pilot Projects 2 and 3. Implementation of the OSM-like Models

The second and third exchange models we proposed (OSM-like, one way exchange with secure access and credentialing , and OSM-like open exchange for public access and large file sharing) both require similar infrastructures, but serve significantly different purposes and would require two additional pilot projects to determine their worth to the Geography Division's address and map updating goals.

In a basic sense development of the OSM-like models follow the diagram below (see Figure 41)

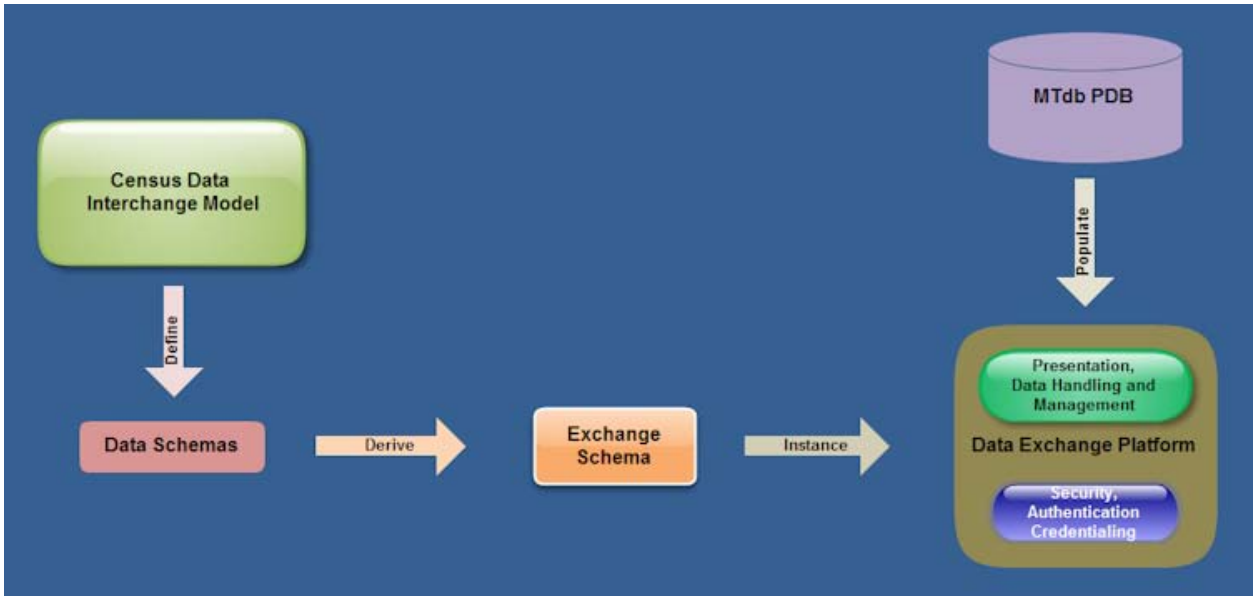


Figure 41. The OSM-like models require the Geography Division to define the data and exchange schemas, security, authentication, credential and the display and handling of address and spatial information.

The OSM-like models are enticing in that they provide easy accessibility, serve data whose quality and integrity are known and controlled by the Census, have a low Total Cost of Ownership and can be scaled with modest addition of equipment, especially if built with open source software.

Conversely, the Geography Division would have to spend some time developing the use models for these system, collecting requirements, developing the infrastructure, standing-up the system, testing it, evaluating the results and migrating to improved performance in the future. A number of the pre-planning issues that need to be considered are shown in Figure 42.

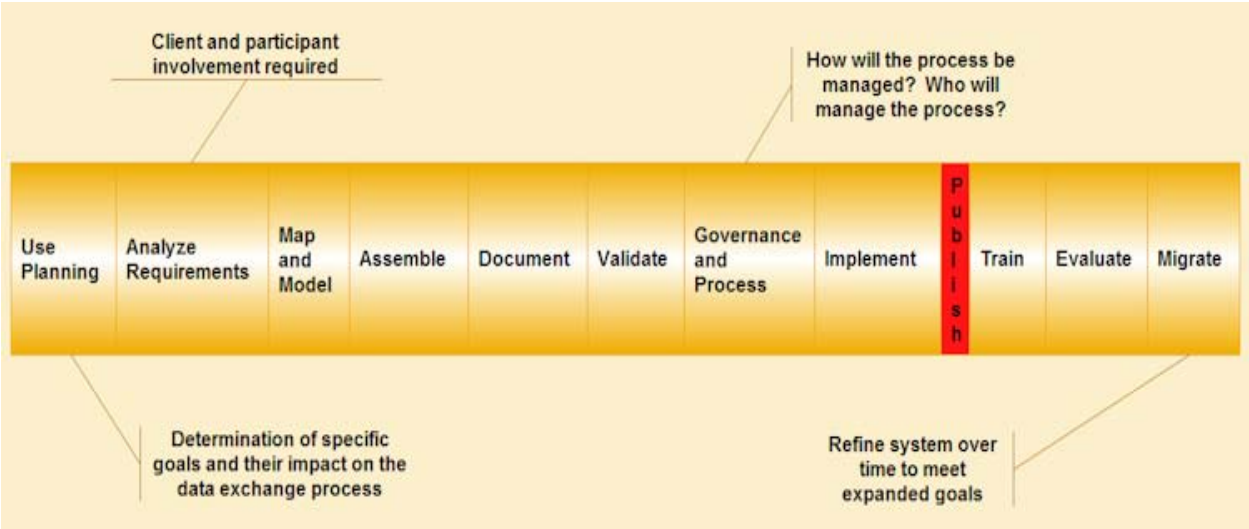


Figure 42. Critical success factors for the development of the OSM-like exchange models.

We think that conceptualizing, implementing, testing, and launching the OSM-like, LUCA-like system for exchanging addresses and spatial data from an MTdb-based PDB for community response will take approximately 20 months to accomplish, but with a modest effort and capable resources that time could be cut to 12 to 15 months. Our recommendations for a development timeline are shown in Figure 43.

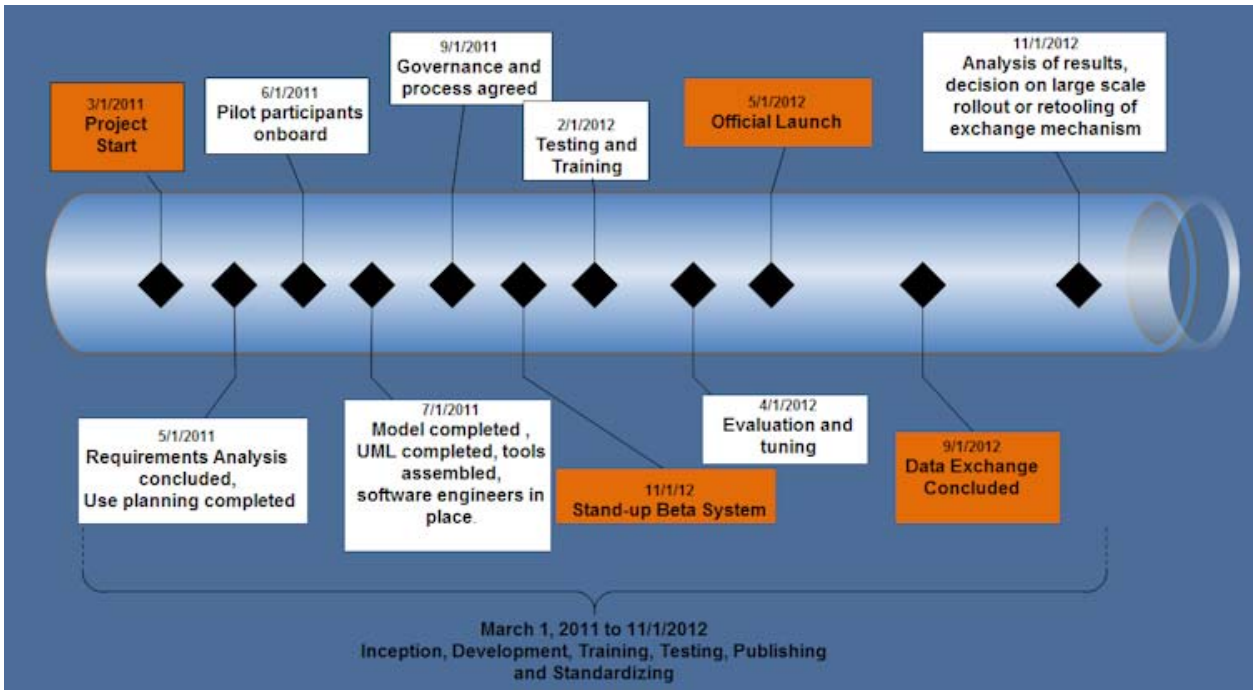


Figure 43. Timeline for developing the OSM-like, LUCA-like one-way, secure access, credentialed system.

The development of this system will be more difficult than the harvesting system due to the complex nature of the interaction with data partners, as well as the need for a secure, compartmentalized system that allows users to see and interact only with data for the geography they are credentialed to inspect, edit, as well as possibly download and upload to the system.

We recommend the Census approach MetroGIS and determine if some of its member organizations representing local governments might we willing to participate in a pilot project based on this model. In addition, we recommend that Delaware County, Ohio be considered for participation. Both candidates appear to run top-notch shops, update their data continuously and are interested the working with the Census on issues such as this.

Implementing a pilot project for Model 3 (OSM-like, open exchange for public access and large file sharing) should be considered an add-on that could be derived from the pilot project for model 2. The development effort could be carried out in parallel or sequentially, depending on the interest level for collecting crowdsourced data. The only additional development required would be providing the ability for the participants to download a TIGER-like PDB (possibly with the address data that has been contributed through crowdsourcing). We have not developed a specific timeline for this effort, but note that it could be the basis for prototyping Pilot Project 2, although it does not require credentialing or the secure system required by that project.

Summary

We recommend the Geography Division undertake three pilot projects, each examining a slightly different approach to the collection of address and spatial data. The harvesting model is one that we feel will have the greatest potential for contributing to the continuous update of the MTdb and satisfying this goal of the GSS. While it is possible that the crowdsourcing pilot project could be of considerable valuable for updating the MTdb, it is unclear to us that the Census website would attract enough traffic to benefit from the error correction mechanism of crowdsourcing equating more eyes with less error. We note that using an OSM-like model to solicit address and spatial data for use when the Census is officially required to request information from local governments (as in LUCA or BAS), may be an approach that increases the likelihood of participation while reducing the amount of error in the responses. In essence, the system will restrict the partners actions to those specified by the Census and thus potentially reduce the need for further processing of the contributed data in the MTdb.

References

Adams, Benjamin, Linna Li, Martin Raubal, Michael F. Goodchild. 2010. A General Framework for Conflation. GIScience 2010 Proceedings. Zurich, September 14-17, 2010. Available online at: http://www.giscience2010.org/pdfs/paper_211.pdf

Craig, William J., 2006, A Master Address File For State and Local Government, URISA, 2006 Conference Proceedings, see www.nsgic.org/committees1/showDoc.cfm?docID=99

Craig, William J. 2010. Email Communication December 12

Danso, Ama. 2010. US Census Bureau MAF/TIGER Product Database: Implementation in the Redesigned MAF/TIGER System Environment. Proceedings, ASPRS/CaGIS 2010 Fall Specialty Conference, November 15-19, 2010 Orlando, Florida

Elhami, Shoreh, 2008, Census 2010 – Local Update of Census Addresses (LUCA) Program in Delaware County, Ohio, Presentation at URISA 2008 Conference, New Orleans, LA.

Federal Geographic Data Committee, 2010. Modernization Roadmap for the Geospatial Platform, Version 3.0 http://www.geoplatform.gov/documents/ModernRoadmapGeoPlatform_V3.pdf

Federal Geographic Data Committee, 2010. Fifty States and Equivalent Entities Involved and Contributing to the NSDI, Federal Geographic Data Committee (FGDC) Future Directions http://www.fgdc.gov/policyandplanning/future-directions/action-plans/FD_PART_Fifty_States_Contributing_NSDI_Final_Action_Plan_v9.pdf

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

Johnson, Randy. 2010. Overview; MetroGIS Spatial Data Infrastructure (SDI)-CREATING PUBLIC VALUE

Kundra, Vivek. 2010. Issuance of OMB Circular A-16 Supplemental Guidance. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2011/m11-03.pdf

Lewis, Steve. 2010. Strategic Planning for Transportation for the Nation (TFTN) PowerPoint presentation to the National Geospatial Advisory Committee, September 23, 2010.

McClure, David. 2010. Creating a Cloud Based Geospatial Platform, PowerPoint presentation to the National Geospatial Advisory Committee, December 6, 2010.

National Research Council, 1990. Spatial Data Needs: The Future of the National Mapping Program, Washington, D.C.: National Academy Press

National Research Council 1993. Toward a Coordinated Spatial Data Infrastructure for the Nation. Washington, D.C., National Academy Press.

National Research Council 1994. Promoting the National Spatial Data Infrastructure Through Partnerships. Washington, D.C., National Academy Press.

National Research Council 1995. A Data Foundation for the National Spatial Data Infrastructure. Washington, D.C., National Academy Press.

National Research Council 2001. National Spatial Data Infrastructure Partnership Programs: Rethinking the Focus. Washington, D.C., National Academy Press.

National Research Council 2003. Weaving a National Map: Review of the U.S. Geological Survey Concept of the National Map. Washington, D.C., National Academy Press.

National Research Council (2004). Licensing Geographic Data and Services Washington, D.C., National Academy Press.

National States Geographic Information Council, 2009. NSGIC FGDC Future Directions – fifty states Initiative. http://www.nsgic.org/resources/fifty_states_initiative_handout.pdf

National States Geographic Information Council (NSGIC), 2005, Fifty States and Equivalent Entities Involved in Contributing to the NSDI, http://www.fgdc.gov/policyandplanning/future-directions/action-plans/FD_PART_Fifty_States_Contributing_NSDI_Final_Action_Plan_v9.pdf

NAVTEQ, 2006, NAVTEQ's NAVSTREETS Street Data Reference Manual, Chicago, Il: NAVTEQ v.2.0 see http://faculty.unlv.edu/jensen/gisdata/navteq/NAVSTREETS_Street_Data_Reference_Manual_Q2-2006.pdf

Nebert, Doug. 2010. Deploying Federal Geospatial Services in the Cloud: Federal Geographic Data Committee (FGDC) and GSA GeoCloud Sandbox Initiative. PowerPoint presentation to the National Geospatial Advisory Committee, December 6, 2010.

Olteanu-Raimond, A.M., Mustière S. 2008. Data matching - a matter of belief, 13*th* International Symposium on Spatial Data Handling (SDH'08)*, 23-25 June, Montpellier (France)*

*Open Geospatial Consortium. 2010. OWS – 7 Engineering Report – Geosynchronization service, OGC 10-069r2. Edited by Vretanos, Peter A. Available online at: (*http://portal.opengeospatial.org/files/?artifact_id=39476*)*

*Robinson, Milo. 2008. A History of Spatial Data Coordination (*http://www.fgdc.gov/ngac *)*

Saalfeld, Alan. 1988. Conflation Automated map compilation. International Journal of Geographical Information Systems, 2(3), 217 - 228.

Salvo, Joseph, 2010 Counting Every Addressing the Census: Joseph Salvo at Pew Research Center, New York City; Pew Research Center. See http://census.pewsocialtrends.org/2010/counting-every-address-in-the-census-joseph-salvo-at-pew-research-center

Siderelis, Karen. 2010. Geospatial Platform Update PowerPoint presentation to the National Geospatial Advisory Committee, December 6, 2010.

Siderelis, Karen. and DeLoatch, Ivan. 2010. Federal Geographic Data Committee Update NGAC Meeting PowerPoint presentation to the National Geospatial Advisory Committee, December 6, 2010.

Stage, David. 2009 Authority and Authoritative Data: A Clarification of Terms and Concepts" Fair & Equitable, February 2009 pp13-16. See http://www.iaao.org/uploads/Stage.pdf

Sweet, Thomas O. 2009. Pennsylvania Federated GIS Enables Data Sharing across County Lines, ArcNews, Fall 2009. Available online at: http://www.esri.com/news/arcnews/fall09articles/pennsylvania-gis.html

Tele Atlas, 2007, Tele Atlas Multinet, Eindhoven, The Netherlands; Tele Atlas

TFTN-Strategic Planning Group, 2010, Conference Notes –What is TFTN and what does it mean? Charleston, West Virginia: Konag Technology Solutions (unofficial meeting notes). See http://www.slideshare.net/KoniagTFTN/gist-notes-public

U.S. Census Bureau http://www.census.gov/geo/www/tiger/tgrshp2009/TGRSHP09AF.pdf MAF/TIGER Feature Class Codes PDF 2009 US Census

U.S. Census Bureau, 2010, Boundary and Annexation Survey Respondent Guide: MAF/TIGER Partnership software. Suitland, Maryland: Geography Division, U.S. Bureau of the Census (PDF). See http://www.census.gov/geo/www/bas/bas11/11_RespondentGuide_MTPS.pdf

U.S. Census Bureau, 2010, Boundary and Annexation Survey Respondent Guide: Digital BAS, Suitland, Maryland; Geography Division, U.S. Bureau of the Census PDF. See http://www.census.gov/geo/www/bas/bas11/11_RespondentGuide_DigitalBAS_Local.pdf

U.S. Census Bureau, 2008, 2010 LUCA Promotional Workshop, Suitland , Maryland: U.S. Bureau of the Census. See http://www.census.gov/geo/www/luca2010/luca_promo_work.pdf

U.S. Census Bureau, 2008, LUCA 2010 Feedback User Guide Supplement, Suitland, Maryland: U.S. Bureau of the Census. See http://www.census.gov/geo/www/luca2010/user_guide_suppl.pdf

U.S. Census Bureau, 2008, FAQ 77. Converting Microsoft Excel 2002 to Pipe-Deliminted ASCII Text Files. Suitland, Maryland: U.S. Bureau of the Census, (online FAQ). See http://www.census.gov/geo/www/luca2010/luca_faq/FAQ_77_for_2010_LUCA.pdf

U.S. GAO, 2007, 2010 Census: Census Bureau Has Improved the Local Update of Census Addresses Program, but Challenges Remain, Washington, DC; US. Government Accounting Office. See http://www.gao.gov/new.items/d07736.pdf

Urban-Mathieu, Brigitta. 2010. Email communication December 12

Weait, Richard, 2010, Introduction to OpenStreetMap Fosscon 2010. See http://www.slideshare.net/rweait/introduction-to-openstreetmap-fosscon2010