

**U.S. Census Bureau
Geography Division
Final Report**



**GSS-I
Address Summit Pilot:
FGDC Address Standard and
Implementation**

March 29, 2013

Approval Log

This Pilot Project Plan has been reviewed and approved for use.

Date	Version	Name & Area Represented	Signature	Description of Major Changes
	V1.0	Timothy F. Trainor Chief, Geography Division		
	V1.0	Deirdre Dalpiaz Bishop Mentor, Geographic Operations Advisor		
	V1.0	Paul Riley Mentor, Office of the Geographic Advisor		
	V1.0	Michael Ratcliffe Mentor, Geocartographic Products and Criteria Assistant Division Chief		
	V1.0			

Table of Contents

1.	Overview	2
2.	Introduction	3
3.	Objectives	4
4.	Methodology	4
5.	Application, Results, and Findings.....	6
5.1	Learning Task.....	6
5.1.1	Overview of the Address Standard.....	6
5.1.2	Mandatory Components	6
5.1.3	Spatial Focus Training	7
5.1.4	Changes to the Original Plan.....	7
5.1.5	Overview of the Training	7
5.1.6	Results from the Spatial Focus Training	8
5.1.7	In-House Training Materials	9
5.2	Assessment Task.....	13
5.2.1	Training, Tools, and Best Practices.....	13
5.2.2	The Address Standard Implementation Questionnaire.....	16
5.3	Testing and Evaluation Tasks	20
5.3.1	Internal Evaluation of the City of Dallas Texas Address Data	21
5.3.2	Pre-training Test and Evaluation.....	21
5.3.3	Post-training Test and Evaluation	22
5.4	Conclusions	23
5.4.1	Best Practices for Implementing the Address Standard	23
5.4.2	Phased Approach for Implementing the Address Standard	26
5.4.3	Geography Division considerations for implementing the Address Standard	29
6.	Recommendations	31
6.1	Recommendations	32
6.2	Discussion of the Recommendations.....	33
7.	Acknowledgements.....	37
8.	Attachment A: Terminology and Acronyms.....	39
9.	Attachment B: References.....	40
10.	Appendices	40

1. Overview

The Geographic Support System Initiative (GSS-I) is an integrated program of improved address coverage, continual spatial feature updates, and enhanced quality assessment and measurement. Its goal is to allow for a targeted, rather than full, address canvassing during 2019 in preparation for the 2020 Census. A pivotal piece of the GSS-I will be working with federal, state, local, and tribal governments, as well as other key stakeholders, to create an address list that is suitable for this purpose.

In preparation, the Census Bureau's Geography Division (GEO) hosted a Census Address Summit in September 2011. Forty-four external experts in the fields of address list development, maintenance, and sharing attended the summit.

The goals of the Address Summit were as follows:

1. To educate Census Bureau partners about the GSS-I and the benefits of conducting a targeted address canvassing.
2. To gain a common understanding regarding the definition of an address.
3. To learn how Census Bureau partners are collecting, utilizing, and maintaining addresses.
 - i. What industry standards are they following?
 - ii. What are their best practices?
 - iii. What are their major challenges?
 - iv. What are their current practices for data sharing?
4. To brainstorm about potential pilot projects that will contribute to the improved quality of the Census Bureau's Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Data Base (MTDB).

During the Address Summit, attendees proposed six pilot projects. Geography Division determined that five pilots were feasible and began development in January 2012. One pilot was determined to be duplicative of other efforts occurring within the division and therefore, was not included. The goal of each pilot project follows:

1. *Address Authority Outreach and Support for Data Sharing Efforts Pilot* – to research and develop an approach for identifying and creating an inventory of address authorities which facilitates address data sharing activities and provides guidance on overcoming barriers (legal/policy) at the local level.
2. *Federal Geographic Data Committee (FGDC) Address Standard and Implementation Pilot* – to educate local authorities on the benefits, use, and implementation of the FGDC's United States Thoroughfare, Landmark, and Postal Address Data Standard (the FGDC Address Standard).

3. *Federal/State/Tribal/Local Address Management Coordination Pilot* – to create a formalized model to allow for the development, maintenance, and bi-directional (state-local, state-federal, and tribal-federal) sharing of high quality multiple use address data.
4. *Data Sharing - Local, State, U.S. Postal Service (USPS), and Census Pilot* – to create an address data exchange model that will allow for address data sharing between local governments, state governments, the USPS, and the Census Bureau. It will provide a business process that increases the accuracy and coverage of local government address lists, while streamlining the process of sharing those externally.
5. *Hidden/Hard to Capture Addresses Pilot* – to determine how to capture hidden and hard to capture addresses in the Master Address File and make them useful for enumeration purposes.

This report focuses specifically on the FGDC Address Standard and Implementation Pilot.

2. Introduction

Addresses are critical information for governments, emergency response, mapping, navigation and many other purposes. Over many decades, thousands of local jurisdictions have evolved a diverse set of address formats and types that now pose complex geo-processing and modeling issues. Government agencies struggle with these issues as they seek to integrate mission-critical files into master address repositories. As a result, the Federal Geographic Data Committee (FGDC) has sponsored the creation of a comprehensive Address Standard. The objectives of the FGDC United States Thoroughfare, Landmark, and Postal Address Data Standard (the Address Standard) include, but are not limited to:

1. Providing one standard that meets diverse address data management requirements.
2. Defining the elements needed to compose addresses.
3. Defining the attributes needed for address documentation, mapping, and quality testing.
4. Providing a systematic classification of US addresses.
5. Defining tests and procedures for address data quality testing.
6. Supporting the seamless exchange of address information by defining Extensible Markup Language (XML) models for address elements and integrating the information into a single XML Schema Document.

As part of the GSS-I to improve the coverage of the address data for the 2020 Census, the GEO instituted the FGDC Address Standard and Implementation Pilot Project. The goal of the pilot project is to create a program to educate and encourage implementation of the Address Standard by local, county, state, and tribal government partners. The addressing community can benefit greatly from implementing the Address Standard. Addressing authorities would benefit from an improved data exchange process as the Address Standard is designed to resolve common issues that arise during the transfer of data. In addition, as addressing authorities begin utilizing the data

quality measures, they should realize higher quality data in the form of reduced duplicate addresses, data entry errors, etc.

This pilot project includes GEO staff from the Address Products Management Branch (APMB), Geographic Project Management Branch (GPMB), Geographic Process & Quality Management Branch (GPQMB), and the National/State Geographic Partnerships Branch (NSGPB), in addition to an external partner, the City of Dallas, Texas.

3. Objectives

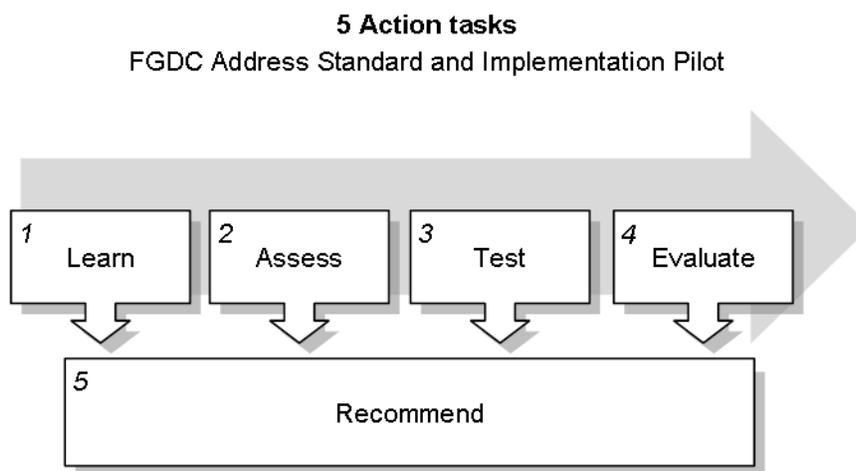
The goal of the pilot project is to create a program to educate and encourage implementation of the Address Standard by local, county, state, and tribal government partners. The objectives listed below can help achieve this goal:

1. Provide local governments with the knowledge needed to understand the Address Standard in order to make informed decisions regarding its use and implementation.
2. Improve the level of expertise within the pilot team (and/or other GEO staff and partners) for critical elements of the Address Standard.
3. Identify common concerns and issues across a variety of jurisdictions in order to develop coherent responses/solutions in the future.

4. Methodology

This project was developed around five general action-oriented tasks: *Learn*, *Assess*, *Test*, *Evaluate*, and *Recommend*. Figure 1 provides a high-level workflow diagram of the pilot team's methodology.

Figure 1: The Five Action Tasks



Learning about and understanding the Address Standard occurred through internal research as well as through training by external experts. A Statement of Work (SOW) was written requesting a contractual agreement with an industry leader to conduct a two-day workshop at Census Bureau headquarters on the Address Standard. In preparation for the workshop, the pilot team developed in-house training materials to assist staff in understanding the Address Standard.

In addition to the learning task, the pilot team *assessed* existing tools and best practices related to the Address Standard and compiled a list of known implementers of the standard. The Implementation Questionnaire, which the pilot team distributed to various governments to garner thorough and consistent information on their implementation of an address standard, served as a major component to this task.

Another component of the project involved an evaluation of the quality of sample address data that the pilot team received from the external partner. This was envisioned as a multi-step process requiring the *testing* of data before and after use of best practices and tools for implementing the Address Standard. Tied to the data testing was the *evaluation* of the results of the tests. This pilot utilized existing matching and geocoding software and evaluation methods developed by the GEO to conduct these steps.

Finally, based on the findings from the *learning, assessing, testing, and evaluation* tasks, the pilot team made *recommendations for* how government partners can begin to implement the Address Standard.

The remaining structure of this report closely follows the workflow from Figure 1. Chapter 5 includes the first four tasks (*learn, assess, test, evaluate*), and provides the application of and the results from each task as well as conclusions. Chapter 5 also includes a section on topics for future discussion that informs the pilot recommendations. Chapter 6 contains the pilot recommendations based on the results/findings from Chapter 5.

This report refers to the FGDC Address Standard as *the Address Standard*. The pilot team references another FGDC standard, the Content Standard for Digital Geospatial Metadata, which this report refers to as the *CSDGM*.

5. Application, Results, and Findings

5.1 Learning Task

The pilot team studied the Address Standard to gain an understanding of the topics discussed in the standard as well as expectations for address managers. After an initial review of the Address Standard, the team identified the *numbered thoroughfare address class* as the focus of the pilot project since this class of address is comparable to the typical *city-style* address used by the Census Bureau and partners.

As part of the learning task, the pilot team developed in-house training materials to assist pilot team members in understanding the primary topics discussed in the Address Standard. The pilot team also invited outside experts to conduct a workshop on the Address Standard for Census Bureau staff at headquarters and the regional offices, as well as invited external partners participating in the other address pilots.

5.1.1 Overview of the Address Standard

The Address Standard bills itself as one standard with four parts. The four parts are: *address content*, *classification*, *data quality*, and *data exchange*. The address content part of the Address Standard specifies the critical address data elements and attributes that are stored in an address table. The classification part codifies the 11 standard-supported classes of addresses. An example includes the *numbered thoroughfare address class*, which is analogous to the city-style address used by the Census Bureau. This class of address requires at the very least an address number, street name, place name, and state name. The data quality part provides 39 measures of data quality, with instructions on how to prepare the data and examples of structured query language (SQL) code for running the quality measures within a relational database structure. The data exchange part of the Address Standard provides steps for preparing the address data and metadata file into an exchange package as well as steps for the recipient in handling the data after receiving the exchange package. In addition to these four parts, the Address Standard contains a section that discusses the concept of an Address Reference System (ARS), which is a framework for local governments to assign addresses. The ARS is critical for local governments as this contains rules for address assignment and plays a role in identifying anomalies in the address records.

5.1.2 Mandatory Components

The Address Standard deliberately avoids a database model, which would have established rules for table structures and relationships within the address database. The Address Standard acknowledges that there exist many ways to manage these tasks, and therefore it is up to the discretion of the local address authorities on how to manage their respective databases. In terms of abbreviations, the standard only allows for two-letter abbreviations for states, the District of Columbia, Puerto Rico, and the Island Areas, as well as special delivery codes used by the United States Postal Service (USPS), such as PO Box and Rural Route (RR). Any other

descriptors, such as street directionals or street types, must not be abbreviated. The standard contains a few mandatory address elements, such as the *AddressID*, but most core sets of address elements may be considered as a best practice rather than a firm requirement. This report discusses these issues in greater detail in the following sections with a summary of best practices in Section 5.4.

5.1.3 Spatial Focus Training

The pilot team invited Spatial Focus, Inc. to conduct a two-day onsite training on the Address Standard and standard-supported tools. Spatial Focus is a leading industry consultant on addresses and played a major role in working with the FGDC to develop the Address Standard. In addition, Spatial Focus developed an open-source address tool as part of an FGDC Cooperative Agreements Program (CAP) grant awarded in conjunction with the work drafting the Address Standard. The purpose of the tool is to assist addressing managers in quality assessment and to facilitate the exchange of address data.

5.1.4 Changes to the Original Plan

The pilot team planned for the Spatial Focus-led training to be held in late May, which was listed as a project milestone on the address pilot's original project plan. The training date was delayed to August 28-29th due to space availability at the Census Bureau Headquarters building and length of time required to release the Statement of Work (SOW) for Spatial Focus to hold the training at the Census Bureau. In addition, the address data and exchange tool (Address Standard Tool) was not ready to demonstrate until late August.

5.1.5 Overview of the Training

The pilot team worked with staff from Spatial Focus to discuss the topics for the two-day training. On the first day of the training, Spatial Focus instructors provided an overview of the Address Standard, discussed common address problems resolved by the Address Standard, provided an overview of the data quality measures defined in the standard, and demonstrated use cases. The second day of the training focused primarily on applications of the address quality and data exchange tool. Appendix C provides a detailed agenda of this training. In addition, the training was captured as a video recording that shows the presentation slides in conjunction with the instructor's narrative. The pilot team plans to post the recording to the GSS-I external webpage. Prior to the training dates, the pilot team requested the other pilot teams to announce the Address Standard training to their external partners. In addition, the pilot team compiled a list of contacts within the GEO to distribute the announcement for the training. The pilot team also worked with the Field Division (FLD) to advertise the trainings to Census Bureau staff at the regional offices (ROs). External partners and the geographers at the ROs were able to access the training live through a webinar platform. In total, the pilot team invited 64 participants to the training.

5.1.6 Results from the Spatial Focus Training

During the first day of the training, 24 participants logged into the webinar. There were 20 participants on the second day. The actual number of participants may be greater since the pilot team asked the ROs to share connections when possible. Although the pilot team did not take attendance in the training room, the estimated attendance varied from 20-30 people each day. Five external partners attended from: Montana, North Dakota, North Carolina, New York, and Georgia. Following the training, the pilot team distributed evaluation forms (see Appendix B) and received 18 completed forms. The pilot team received one form for each RO that chose to respond. Of the 18 responses, 10 were from HQ employees, five from the ROs, and three from the external partners. The first 10 questions on the evaluation form provided multiple-choice questions regarding the characteristics of the participant and their assessment on the overall quality of the training. See Appendix F for a complete breakdown of the responses to questions one through 10. Eight respondents indicated that they had some prior knowledge of the Address Standard, whereas six respondents indicated they had no prior knowledge. Four respondents did not respond to this question.

The majority of respondents indicated that the training was well executed. Twelve responded that the training met their expectations; 15 indicated having a better understanding of the Address Standard after attending the training and 15 indicated that the knowledge gained from the training will be applicable to their work. With regard to the effectiveness of the instruction done by Spatial Focus, 15 out of 18 responded that the quality of the instruction was good and 14 out of 18 responded that the presentations were easy to follow.

5.1.6.1 External partner evaluation of the training

All three external partners that responded had prior experience with the Address Standard. None of the respondents expressed difficulty in following the discussion throughout the training. However, one respondent felt that even though the data quality discussion on the second day of the training was easy to understand conceptually, the SQL discussion was too detailed for an introductory-level training. In terms of being able to implement the Address Standard, one of the respondents expressed concern that the address quality and data exchange tool will not work under the Esri ArcSDE environment. Attendees were also concerned that many governments still use ArcSDE as well as other Esri software to manage address data. Currently, the address tool does not support these proprietary environments.

5.1.6.2 HQ employee evaluation of the training

Of the 10 HQ respondents, only three had prior experience with the Address Standard. Unlike the external partners, many HQ respondents had difficulty with the data quality discussion, which assumed previous knowledge of SQL and Java. Seven respondents indicated that the SQL discussion was difficult to follow. Respondents indicated that barriers to implementation of the Address Standard included lack of documentation for the address tool, concerns over a lack of support for governments that use Esri software to maintain address data, and difficulty in using the quality and data exchange tool out-of-the-box. The majority of respondents indicated that the address tool may have some use for the GEO despite the difficulties and barriers. Examples

include using the tool for data sharing between the GEO and government partners, running quality measures on address data received, and interest in seeing future documentation on the address tool.

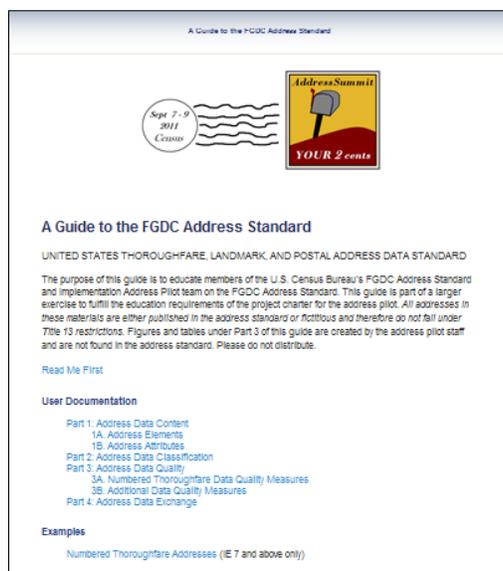
5.1.6.3 RO evaluation of the training

Five ROs provided consolidated responses. Two ROs indicated having prior experience with the Address Standard. Similar to the HQ employees, several had difficulty in following the data quality discussion. Four responded that SQL discussion was too complex or difficult to follow. None of the respondents identified any potential barriers to implementing the standard at the Census Bureau. One respondent, however, wished to have seen more discussion on tools that can be used by the GEO to collect address data from government partners. Another respondent hoped that the GEO would be able to develop a similar training for its government partners. Another respondent wanted to see more information on how RO geographers can use Address Standard Tool in the future.

5.1.7 In-House Training Materials

The delay in scheduling the Spatial Focus training (described in Section 5.1.4) provided the pilot team additional time to study the Address Standard and create an in-house guide to the standard to be made available to Census Bureau staff. The purpose of the guide was to educate team members on the key concepts of the Address Standard in advance of the Spatial Focus training. The guide provides a summary of the four parts to the standard: *address content*, *classification*, *data quality*, and *data transfer*. The guide is designed to look like a webpage and contains hyperlinks to help the user navigate between sections (see Figure 2).

Figure 2: Main page for the in-house guide to the Address Standard



For the parts on address content and classification, the Address Standard provides a detailed account of each address element and address attribute as well as the classification of address types. The in-house guide provides a quick lookup tool to locate each element, attribute, and classification as well as examples taken directly from the Address Standard. The parts on data quality and data exchange provide detailed requirements on how to develop tests to assess address quality and formalize steps for formatting the address dataset into a standards-compliant exchange file.

Since the Address Standard relies heavily on SQL examples for conducting the data quality measures, the guide provides visual examples of what the SQL query is achieving. One example is the Address Number Fishbones measure, which identifies address points that may not be appropriately located relative to the street centerline.

Figure 3: Example SQL Code for the Address Number Fishbones Measure

```
//Identify addresses with intersecting fishbones

INSERT INTO FishboneAnomalies
(
    AddressID,
    AddressNumber,
    CompleteStreetName,
    Anomaly
)

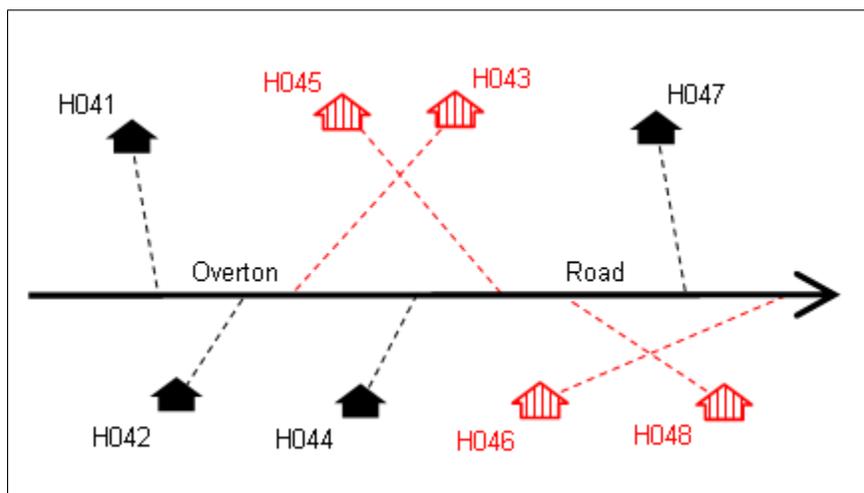
SELECT
    a.AddressID,
    a.AddressNumber,
    a.CompleteStreetName,
    'Fishbone touches'::TEXT as "Anomaly"

FROM
    Fishbones a
    INNER JOIN Fishbones b ON TOUCHES(a.Geometry, b.Geometry)

WHERE
    a.AddressID != b.AddressID
```

Whereas the Address Standard uses SQL code to communicate how the quality measures are to be conducted (see Figure 3), the in-house guide uses visual examples to help communicate the concepts behind these quality measures to those unfamiliar with SQL. Figure 4, taken from the in-house guide, visually describes the SQL code for the Address Number Fishbones measure shown in Figure 3.

Figure 4: Addresses with Intersecting Fishbones



The example SQL code from Figure 3 conducts one component of the Address Number Fishbones Measure data quality test in which the *fishbone* geometries (represented as the dashed lines in Figure 4) are constructed in order to assess address point locations in reference to the address range. Fishbones connect the address point to the corresponding point along the street centerline to which the address geocodes. Fishbones that intersect are an indicator of house numbers that are out of sequence or address points that potentially have incorrect XY coordinates. The address labels from Figure 4 reference the unique ID value for each address point. The SQL query reads the entire address file and generates a table called FishboneAnomalies (see Table 1), which contain address records with intersecting fishbones.

Table 1: FishboneAnomalies table

id	AddressID	AddressNumber ¹	CompleteStreetName	Anomaly
0	H043	<###>	Overton Road	Fishbone touches
1	H045	<###>	Overton Road	Fishbone touches
2	H046	<###>	Overton Road	Fishbone touches
3	H048	<###>	Overton Road	Fishbone touches

¹Address numbers suppressed for this example.

The in-house guide describes eight of the 39 quality measures described in the Address Standard. Prior to the training, the pilot team identified these eight quality measures as being useful for assessing the quality of numbered thoroughfare (city-style) addresses. In addition, the guide contains examples of how numbered thoroughfare addresses are to be parsed into their standard address elements (see Figure 5) in order to stage the address data for conducting the data quality measures and creating the data exchange file.

Figure 5: Address element examples from the in-house guide

Address Element Examples | Numbered Thoroughfare Addresses - Windows Internet Explorer

1 South Avenue C, [Redacted] [Redacted]
 A12 Calle B, [Redacted] [Redacted]
 12 Boulevard of the [Redacted] [Redacted]

Complete Landmark Name & Complete Place Name			
Landmark Name 1	Landmark Name 2	Place Name 1	Place Name 2

Complete Address Number		
Address Number Prefix	Address Number	Address Number Suffix
	1	

Complete Street Name							
Pre Modifier	Pre Directional	Pre Type	Separator	Street Name	Post Type	Post Directional	Post Modifier
	South	Avenue		C			

Complete Subaddress					
Subaddress Element 1		Subaddress Element 2		Subaddress Element 3	
Type	Identifier	Type	Identifier	Type	Identifier

Complete Place Name	
Place Name 1	Place Name 1
[Redacted]	

State Name	Zip Code	Zip Plus 4	Country Name
[Redacted]	[Redacted]		

Note: All address examples are taken from the FGDC Address Standard.
 The address elements are defined under Part 1 of the FGDC Address Standard.

5.1.7.1 Spatial Focus Assessment on the In-House Materials

Spatial Focus had the opportunity to review the in-house guide and provide professional feedback to the pilot team. During the review, Spatial Focus pointed out that the in-house guide did not include sections on the Address Reference System (ARS) and the XML schema definition (XSD) under the data exchange part of the Address Standard. The pilot team did not include the ARS in the guide since this concept did not apply to the work done by GEO at the time, and the ARS can vary depending on locale. However, the ARS would be important at the local addressing authority level, which is responsible for the creation of addresses. In the future, though, the ARS provided by government partners may play a larger role within the GEO, which may aid GEO staff in understanding local address assignments and anomalies in the data. The XSD from the Address Standard was too cumbersome for the pilot team to include it in the guide. The pilot team recommends that users refer to the Address Standard, which does provide a well-developed XSD.

For the address content and classification pages of the guide, Spatial Focus indicated that the full notes found in the Address Standard must be included in all the elements and classification tables. For the data quality page, Spatial Focus wanted more data quality measures to be included. The Address Standard has a total of 39 tests, of which the pilot team selected eight to include in the guide. These eight tests are applicable to the numbered thoroughfare (city-style)

addresses. Spatial Focus proposed that the in-house guide should be more comprehensive in covering the topics from the original Address Standard. The pilot team's intention, however, was to create a primer that provides a simple introduction to the overall standard that would be more approachable to a larger audience, particularly those users that do not have the database management background to fully comprehend the technical aspects of the standard.

The pilot team received and approved the suggestions provided by Spatial Focus, but due to the limited time of the pilot project, any updates suggested by Spatial Focus to the guide will be implemented at a later time. If the guide were to be distributed to the public, potentially all 39 tests would need to be included.

5.2 Assessment Task

For the assessment task, the pilot team researched existing resources and used the results of the research to formulate and distribute a questionnaire designed to gather information about how state and local entities were implementing address standards. The research consisted of a review of websites of organizations such as the Urban and Regional Information Systems Association (URISA), the National States Geographic Information Council (NSGIC), and the National Emergency Number Association (NENA). In addition, staff from the GEO attended address-related conferences such as the URISA/NENA Addressing Conference, the NSGIC annual conference, and the URISA annual conference, which offered opportunities to network with state and local partners about their addressing programs.

As a result of attending the NENA development conference to support the Address Summit pilot projects, the Census Bureau participated in the development of the new version of the NENA Next Generation 9-1-1 (NG9-1-1) United States Civic Location Data Exchange Format (CLDXF) Standard. This offered an opportunity for the team to learn how the NENA CLDXF Standard relates to the Address Standard. Whereas the FGDC created one address standard that meets diverse address data management requirements for a variety of purposes, the CLDXF Standard targets the exchange of fixed addresses and mobile location information from Public Safety Answering Points (PSAPs) for 9-1-1 calls. The CLDXF Standard is a very specific data exchange standard that is not intended to support address data management or standardization of addresses. While the two standards were designed to support different objectives, there are set of core address data elements that are common to both standards.

5.2.1 Training, Tools, and Best Practices

The pilot team's research of existing training, implementation, and best practices for the Address Standard resulted in a list of governments engaged in various stages of implementing an address standard (see Appendix B). This information came from discussions with Spatial Focus as well as a review of the NSGIC Geospatial Maturity Assessment database:

<http://www.nsgic.org/geospatial-maturity-assessment>. The pilot team used this list to identify potential recipients of the Implementation Questionnaire. With respect to best practices, the pilot team was unable to find publically available information on the implementation activities of several entities known to have implemented the Address Standard.

5.2.1.1 Available training

During the research on existing training materials, the pilot team found few existing trainings on the Address Standard. The State of Louisiana conducted a workshop on the Address Standard, and URISA/NENA hosted trainings for their Addressing Conference. These were both URISA-sponsored training classes and materials were available only to attendees. The classes are offered for a fee and are focused on addressing techniques for local authorities rather than implementation of the Address Standard. Census Bureau staff had participated in a one-hour URISA sponsored Address Standard webinar led by Martha Wells of Spatial Focus, which preceded the Census Bureau release of a SOW to Spatial Focus for the two-day onsite training discussed in Section 5.1.

Other than several overview presentations by Spatial Focus on the Address Standard and the tools proposed for a FGDC CAP grant study, the pilot team did not find comprehensive and freely available training materials on the Address Standard. Existing training materials are high-level overviews of the topics prepared for various audiences or available from URISA for a fee. The few Address Standard-supported tools identified by the pilot team were those developed by Spatial Focus.

5.2.1.2 Available tools

The first tool is an electronic version of the Address Standard, which employs hyperlinked documents for better navigation through the large amount of information (the Address Standard, when printed, is almost 550 pages). This tool offers search capabilities and the ability to collapse and expand sections of the Address Standard providing easy access to the content of the standard. The webpage may be viewed on Android mobile devices. The tool can be downloaded at: <http://www.spatialfocus.com/wiki/bin/view/SpatialFocus/SearchableFGDCAddressStandard>.

The second tool is the Address Quality and Data Exchange Tool (Address Standard Tool). This tool is currently a prototype as opposed to a fully mature product. The software for this tool is open source, written in Java, with the H2 database engine embedded. The only software needed to use the tool is Java. The tool is freely available and runs on Linux as well as Windows. The source code files are stored at github (www.github.com tag line: "Social Coding"), which is a popular hosting environment for supporting collaborative software development. Spatial Focus envisions further development of the tool to occur, in part, through collaboration by the address data community as an open source project. Spatial Focus will continue to support and lead the development of the tool.

The purpose of the prototype Address Standard Tool is to: 1) convert a user's address data file into an XML-based exchange file, and 2) load the user's address data into the embedded database to prepare it for the 39 potential SQL-based quality measures. The quality measures are very comprehensive and require specific data inputs. Most notably, the data must have unique identifiers for each address record. This unique ID is referred to as the AddressID in the Address Standard. In addition, not all of the measures can be applied to any address data file. For example, many of the advanced quality measures require a road network. Running the tool requires experience with SQL, and thorough knowledge of the user's address data.

The tool is very flexible for a couple of reasons. For one, the user maps their address data components to the Address Standard elements via configuration tables. This task saves the user from altering the structure of their local address database. In addition, the user selects which data quality measures to execute. These measures can be executed in intervals and over time as the user gains experience with the Address Standard and as the quality of their address data changes over time.

While these are good features for assessing the quality of local address data, the tool presents challenges to many users. Since the developers designed the tool to be customized by the user, the out-of-the-box functionality is limited. By default, the tool simply creates the user's source address data as an XML exchange file and reports the number of addresses in the file, but does not perform any basic quality measures. Utilization of the quality assessment measures requires additional SQL coding by the user. The Address Standard defines the required address data elements and the expected output of the data quality measures in addition to providing stub SQL code for all the quality measures. However, the user must possess a thorough understanding of their local address data prior to using the tool as well as an intermediate-to-high-level SQL expertise to execute most of the data quality measures. Spatial Focus is currently enhancing the tool to make it easier to use by incorporating a graphical user interface (GUI). The GUI will minimize the need for SQL expertise, making it easier for the average local government to use and adopt.

Spatial Focus plans to release the new GUI version of the tool in 2013. Currently the tool runs from the command line and the user must manually configure the tool to link the address components to the Address Standard elements. In addition, the documentation that comes with the prototype tool is not detailed enough for most users. To compensate, Census Bureau staff created a user's guide for internal use that may be shared with the public in the future. Spatial Focus is currently creating full documentation for the GUI version of the tool and the supporting code.

By design, the capabilities of the Address Standard tool can be extended in a number of ways. For example, the tool supports the use of other RDBMSs for those users who want to leverage their database system instead of the embedded RDBMS. However, the SQL utilized by the tool is the OGC Simple Features for SQL (SFSQL) which requires a RDBMS that supports specific spatial capabilities. Because of this there are some incompatibilities between the tool and some RDBMSs such as Oracle. Incompatibilities also exist between the tool and ArcSDE. Workarounds may exist, but the examples in the Address Standard do not provide a ready solution.

Spatial Focus has also identified the following open source software as complimentary to the Address Standard Tool:

GDAL – an extract, transform and load (ETL) package to facilitate loading data
PostgreSQL - is an object-relational database management system (ORDBMS)
PostGIS - adds support for geographic objects to the PostgreSQL
QGIS - a spatial client for PostGIS.

Currently there is no documentation on how to extend the capabilities of the tool using these software packages. Spatial Focus envisions this as a possible outcome from the collaborative development environment they have established.

While it is fairly easy to use the tool to create an address data exchange file, the pilot team is concerned that the address expertise and database query skills required to implement the data quality measures with the existing prototype tool will be an obstacle for many governments. In order to use the Address Standard Tool for any meaningful assessment of the quality of the data, the address provider must have staff experienced in SQL. Given the intermediate- to-high level SQL knowledge required to implement the data quality measures, the pilot team expects that standard address data will come mainly from governments that have a technologically sophisticated addressing department, such as state, regional, county, and larger municipalities that maintain an enterprise address database.

For governments that maintain address data by spreadsheet, shapefile, or geodatabase, the data quality measures from the Address Standard may not be straightforward. Also, the number of data quality measures will be limited, especially if the entity does not maintain address feature geometries along with the address data. On the other hand, many local and state governments have not developed any address quality testing procedures and the tool offers some very practical and useful measures that can be used to improve address data quality.

The address quality and data exchange tool prototype has great potential, especially if it can gain support from a community of users and developers, but its current prototype functionality may not be enough to encourage implementation of the Address Standard by government partners. Spatial Focus does intend to continue to improve and extend the tool and it is beneficial to have an open source solution that is not dependent upon any specific operating system or GIS software. Commercial software such as Esri ArcGIS is a popular platform used by many governments to manage address data. Many government users may find utility in an address standard tool that can conduct the quality measures within the Esri environment. This, however, would sacrifice many benefits offered by an open source approach.

5.2.2 The Address Standard Implementation Questionnaire

Another component of the assessment task included designing a questionnaire to collect information about the implementation of an address standard by tribal, state, and local entities. The pilot team used information collected from the questionnaire to inform its recommendations for implementing the Address Standard. The pilot team's approach to the design of the questionnaire follows these three principles:

- *Simplicity* – the burden on the respondent should be minimal. For that reason, the questions provided different choices for the respondent's answers (there is only one question that asked for explanation).
- *Thoroughness*– although simplicity was the main goal, the questions provided the means for a thorough response, which the pilot team used to assess the implementation of an address

standard by the respondent. The questionnaire provided an “Explanation of the Questions” section in case the respondent needed additional explanation when completing the questionnaire.

- *Electronic (email survey)* – In order to have a quick turnaround, the pilot team sent the questionnaire via email and responses were sent back to the pilot team via email or fax.

5.2.2.1 Development of the questionnaire

The design and content of the questionnaire went through several changes. As Figure 6 shows below, after reviews by the GEO management and the other pilot teams, the latest design provided a less intimidating and a better look and feel, than the more official looking original version (left hand side). This was in line with the principle of simplicity.

Figure 6: Changes to the design of the Implementation Questionnaire

The Implementation Questionnaire was subject to the OMB clearance for the GSS-I. GEO received OMB clearance and proceeded to collect information under clearance number **OMB NO. 0607-0795**. The team included the OMB control number (**OMB NO. 0607-0795**) and the OMB burden statement on the questionnaire to meet the OMB requirement. Appendix G contains a copy of the questionnaire.

During the selection process (discussed in Section 5.2.1), the pilot team reviewed a list of entities that are in the process of implementing or have implemented an address standard. The pilot team began with 25 entities: 18 local and seven state governments. The pilot team then narrowed down the list to those entities that have NSGIC state representatives or are currently working as partners with the other GSS-I address pilot projects. This final list yielded 12 entities: seven state and five local governments. The pilot team used this list to send out the questionnaire with the following procedures:

-
- Inform NSGIC prior to sending out the questionnaire –a copy of the letter that the GEO emailed to the president of NSGIC is included in Appendix I.
 - Allow flexibility for response submission – respondents may submit their completed questionnaire by email or by fax
 - Allow two weeks for respondents to complete and submit the questionnaire.

5.2.2.2 Results from the questionnaire

Ten of the twelve respondents provided responses to the questionnaire. In assessing the results, the pilot team focused on four areas within the questionnaire:

1. *Use of an address standard* - Half of the respondents indicated that they follow a particular standard; half do not.
2. *Use of which address standard* - Of the five respondents that indicated that their entity is using an address standard, two are using the Address Standard only, while one is using the NENA Standard only, one is using the USPS standard only, and one is using a combination of FGDC and NENA standards.
3. *Statewide address database* - The other area of the questionnaire focuses on the availability of a statewide address database. The majority of the respondents (six) indicated that no such database exists in their state. Three of the respondents indicated that their state maintains a state-level address database, and only one respondent did not know of any state-level address database.
4. *Authoritative and centralized address database* - Half of the respondents did not have a centralized and authoritative database, but did have plans to implement one in the future. Three respondents did not have such a database and did not have plans to implement one. Only one respondent indicated having such a database.

Below is a table summary of the responses and the respective counts for these key questions.

Table 2: Implementation Questionnaire Responses

Question	Response	Count
<i>Does your organization follow a particular standard?</i>	<i>Yes</i>	<i>5</i>
	<i>No</i>	<i>5</i>
<i>Which Address Standard does your organization follow?</i>	<i>FGDC</i>	<i>2</i>
	<i>NENA</i>	<i>1</i>
	<i>FGDC/ NENA Combination</i>	<i>1</i>
	<i>USPS</i>	<i>1</i>
	<i>None</i>	<i>5</i>
<i>Does your state have a statewide address database?</i>	<i>No</i>	<i>6</i>
	<i>Yes</i>	<i>3</i>
	<i>Don't Know</i>	<i>1</i>
<i>Is address data from an authoritative and centralized address database?</i>	<i>No, but there are plans to implement one</i>	<i>5</i>
	<i>No. We maintain address data for our own use</i>	<i>3</i>
	<i>Yes</i>	<i>1</i>
	<i>N/A</i>	<i>1</i>

Although the limited results indicate use of address standards, many of the respondents, including those organizations that are considering adopting an address standard, indicated that full implementation faces significant limitations such as:

- Lack of training
- Computer code/software limitations
- Limited knowledge of standards and resources to implement them
- Conflicting agency needs or organizational reluctance
- Cost/funding

This report revisits some of these issues in the remaining sections. The pilot team's approach for developing a comprehensive training plan which also includes a guide and workbook, considers the current lack of training and limited knowledge of data standards. In addition, this report's

recommendation to establish a user community and a supporting website with a frequently asked questions (FAQ) section can help answer questions related to best practices, funding, and internal organization issues (see Chapter 6).

Despite these limitations, some respondents indicated progress toward a centralized or statewide address database. The emergence of NextGen911 is also mentioned as a catalyst for adoption of an address standard or centralized address database where necessary. Please refer to Appendix H for a complete list of responses to the Implementation Questionnaire.

5.3 Testing and Evaluation Tasks

One of the tasks of the pilot project was to test an address dataset prior to and after gaining an understanding of the Address Standard, tools, and best practices as described in the learn and assess tasks of this report. In March of 2012, the pilot team's external partner delivered an address point geodatabase file (.gdb) to the Census Bureau pilot team members as well as to Spatial Focus for testing purposes. The file contained a sample of approximately 50,000 addresses from the City of Dallas, TX. The address data contained attribution stored in a table as well as spatial data stored as point coordinates. In addition to the address point file, the external partner provided a street centerline file for the City of Dallas, TX in anticipation of conducting a subset of the data quality measures that check address data against the street centerlines.

During the data collection and preparation steps of the testing and evaluation tasks, the GEO management and pilot team members raised concerns over potential Title 13 implications with regard to providing feedback to the external partner on observations made from the test address data. This concern was relevant to the other GSS-I pilot teams as well. The pilot team initially proceeded with these tasks under the assumption that the data testing did not violate Title 13 restrictions, since the Census Bureau pilot team members would not provide record-level feedback to the external partner. However, the original project plan did include matching the address data received from the external partner to the Census Bureau's MTDB discussed in the next section, which raised the question of *when* the data become Title 13 restricted. In order to avoid any violations to Title 13, both real and perceived, the pilot team had to reassess the data testing and evaluation tasks.

The timeframe for releasing the Spatial Focus address tool created another challenge for the pilot. The original release was planned for early spring of 2012, but due to a deadline extension that FGDC granted Spatial Focus and federal procedures for releasing software within the Census Bureau, the pilot team did not have the tools available for testing until late summer of 2012. The pilot team did perform one basic test of the Dallas address data prior to the training (see Section 5.3.2) and the external partner performed additional tests on the address data after the training (see Section 5.3.3).

Given the challenges stated above, the team developed a contingency plan for the testing portion of the pilot project. This involved the external partner independently using the address tool to evaluate the City of Dallas address file outside of the Census. This restructured process removed any Title 13 concerns. Our partner provided thorough feedback on this exercise, which is summarized in Section 5.3.5.

5.3.1 Internal Evaluation of the City of Dallas Texas Address Data

One of the workflow steps outlined in the pilot project plan was to identify existing software and methodology available within the GEO for testing the external partner's address data. The evaluation of the sample address data follows these two components: (1) compliance with the Address Standard and (2) data conformance to the MTDB.

The pilot team used the Spatial Focus tools to evaluate the compliance of the external partner's address data with the Address Standard. Due to the lack of documentation for the address tool and the delay in receiving the address tools mentioned in Section 5.1, the pilot team could not identify the data evaluation factors. The team initially assumed that the address tool would assist Census Bureau staff in evaluating the presence or absence of address elements and validate the format and types of required address elements.

For testing the data for conformance with the MTDB, the pilot team identified the Geographic Support System Matching and Geocoding Software (GSSMG) and associated evaluation methods as an appropriate tool for the internal testing and evaluation of the address data. The GEO subject matter experts and programmers developed the GSSMG to support the collection and evaluation of external address files and corresponding spatial information, such as address point coordinates. In addition to the GSSMG, another GSS-I address pilot team (Federal/State/Tribal/Local Address Management Coordination Pilot) was developing recommended methods to the GEO for providing feedback on how volunteered address data compares with the MTDB to external partners. The Address Standard pilot team considered utilizing these existing resources to establish a *baseline* for evaluating the effectiveness of the Spatial Focus onsite training on the Address Standard and demonstration of Address Standard Tool. Concerns about Title 13 implications as well as missing information from the address data restricted the pilot team's ability to leverage these existing in-house resources. As a result, the external partner's data was not tested for conformance to the MTDB.

5.3.2 Pre-training Test and Evaluation

After an interactive review of the external partner's address data, the pilot team found that ZIP Codes were missing from the file. ZIP Codes are critical pieces of information needed to effectively evaluate a file with the GSSMG software. Addresses without ZIP Codes fail to match to the MTDB and yield inaccurate results. Due to the lack of ZIP Codes and the Title 13 concerns mentioned above, the external partner's data was not evaluated with the GSSMG. However, one unanticipated benefit of this project was that the external partner was made aware by Spatial Focus of products available through the USPS that could help provide accurate ZIP Codes. While further discussion and support on this topic was outside the scope of the pilot project, the external partner is excited to pursue this in the near future, as a lack of ZIP Codes has been a persistent address related problem for the external partner for quite some time.

The pilot team, with help from Spatial Focus, conducted a "dry run" of the Spatial Focus address tool prior to the onsite training. During this preliminary demonstration of the software, the pilot team ran the *UniquenessMeasure* on the *AddressID* field. While narrow in scope, the dry run

identified that the external partner's data lacked unique identifiers (AddressIDs) for each address record, a necessary attribute for conducting the data quality measures.

5.3.3 Post-training Test and Evaluation

For the post-training data test and evaluation, the pilot team's external partner examined their address data in relation to the data quality measures defined in the Address Standard. The external partner found a few quality measures that could be conducted on the current address dataset. These quality measures include:

1. The XY Coordinate Completeness Measure – identified addresses that did not have XY coordinate information stored in the appropriate columns. The external partner was able to update the addresses identified by this data quality measure.
2. Related Element Uniqueness Measure – checked for uniqueness of address number, subaddress type (suite), and complete street name. This measure identified a large number of duplicates, which the external partner traced back to a past data conversion.
3. Tabular Domain Measure – checked the street type values against the USPS address standard. This measure identified several discrepancies attributed as anomalies—elements that do not follow a standard but nevertheless remain valid for the city government's addressing purposes. The external partner ran this measure to check *CompleteStreetNames* against the authoritative street centerline file for the City of Dallas, TX, as well as for *StreetPreDirection* and *StreetPostDirection* values. The measure identified several discrepancies, which the external partner corrected.

Since unique IDs had not been established and maintained for the City of Dallas, TX address dataset and they are a prerequisite for using the Spatial Focus address quality and data exchange tool, the external partner ran the quality measures listed above without the use of the tool. This was done by constructing and executing SQL queries, modeled after queries documented in the Address Standard, directly against the database.

The external partner was unable to conduct other data quality measures (e.g., Fishbone or Range Domain measures) as the current address data are stored in a standalone dataset, not a relational database. Many of the data quality measures found in the Address Standard require a link between address point features and related features, such as street centerlines and other addressable features such as structures. While the necessary columns to create these linkages do exist in the current dataset, the external partner currently does not maintain such linkages. In addition, as discussed previously, the lack of a unique ID prevented further use of the Address Standard quality measures. Based on the experience gained from this exercise, the external partner has identified a solution for using unique IDs for Dallas' address data, and plans are under way to implement this solution within the year.

While some specific measures are not easily implemented, the external partner feels that the data quality measures tested from the Address Standard were extremely helpful in identifying issues with the current address data. Based on the lessons learned from this address pilot, the external

partner will begin steps for assigning unique IDs to the Dallas address data, linking address points and street centerline data, and resolving duplicate records and incorrect values in the data. While Dallas may not be able to exchange data with the Census Bureau today, as the data would score low on the data quality measures, the external partner feels they have been given tools that will allow them to clean their data and prepare for data exchange within the next few years.

The experience of our external partner outlined above, as well as feedback from the training, made it clear that developing a starting point for implementing the Address Standard would be beneficial. Both our external partner and participants in the training expressed concerns that the sheer volume of information in the FGDC Address Standard might be overwhelming at first to addressing authorities interested in implementing it. Specifically, concern on how to get started with the quality measures was expressed. This led the pilot team to the concept of identifying a core set of quality measures that met the following criteria:

1. The potential improvement to the data must be high;
2. The complexity of running the measures must be relatively low;
3. There should be a limited need for data in addition to the address dataset itself.

The core set of quality measures identified can be found in Table 4 below. Detailed descriptions and instructions on running them can be found in *Part 3: Address Data Quality* of the Address Standard.

The GEO can benefit greatly from the *data quality* and *data exchange* parts of the Address Standard. As local addressing authorities begin utilizing the data quality measures, the GEO should realize higher quality data in the form of reduced duplicate addresses, data entry errors, etc. Similarly, as government partners are implementing the data exchange protocols outlined in the standard, the GEO would benefit from an improved data exchange process. The Address Standard seeks to resolve common issues that arise during the transfer of data including but not limited to: the loss of data, formatting errors, and ambiguous field labels. Understanding how the *data quality* part of the Address Standard could be utilized to help assess MTDB data, as well as determining how the GEO's current quality testing processes align with the Address Standard, would improve the level of expertise within the GEO and ultimately aid government partners in their understanding and implementation of the Address Standard. This will require input from both address and spatial experts in the GEO as many of the data quality measures are based on relationships between addresses and spatial features.

5.4 Conclusions

5.4.1 Best Practices for Implementing the Address Standard

A number of best practices emerged from our learn, assess, test, and evaluate activities. These best practices encompass mandatory and optimum address elements per the standard in addition to practices that lead to improved data quality. State and national level data repositories will also benefit from implementing these practices. They are listed in the table below.

Table 3: Best Practices for Implementing the Address Standard

No.	Description
1	<i>The address data file must include all required address elements based on the class of addresses.</i>
2	<i>The address data file must include a unique identification number/code (AddressID) for each record.</i>
3	<p><i>The address data file should include the following additional attributes:</i></p> <ul style="list-style-type: none"> • <i>Address Classification</i> • <i>Address Feature Type</i> • <i>Address Lifecycle Status</i> • <i>Address Official Status</i> • <i>Address Direct Source</i> • <i>Address Authority</i> • <i>Address Anomaly Status</i>
4	<i>Include XY coordinates for addresses in the address data file for the applicable area and remove/correct outliers.</i>
5	<i>In addition to the required attributes provided in the address data file (described in recommendation 2), the data provider should also furnish FGDC CSDGM metadata that contains information on: copyright, use restrictions, contact information, data lineage, known data quality, and a description of the represented geographic area.</i>
6	<i>Do not abbreviate street types, directionals, landmark names, and community names.</i>
7	<i>Store address information in its most basic form (simple element).</i>
8	<i>Store address data in a normalized relational data base management system (RDBMS).</i>

1. *The address data file must include all required address elements based on the standard-supported class of addresses.* Similar to how the GSS-I optimum and minimum address submission guidelines specify the mandatory address elements of an address for city-style addresses, the *address classification* part of the Address Standard requires a minimum set of address elements for the numbered thoroughfare class of addresses. The GEO should ensure that staff understands these two requirements and that the information being conveyed to government partners is consistent.
2. *The address data file must include a unique identification number/code (AddressID) for each record.* The *address ID* is the unique identifier assigned to an address by the address

authority. The *address ID* is needed for all the address quality measures and is key to supporting state or national level aggregations of addresses into consolidated databases.

3. *The address data file should include the following additional attributes:*

- a. The *address classification*: assign address classification to each address based on syntax (i.e. address elements and their order). Classifying addresses by their syntax, that is, their address elements and the order in which the elements are arranged, allows users of the standard to focus on record structures and avoids making any assumptions about what the address might identify on the ground. Record structure often is all that is known about addresses in a given file and is needed to hold and exchange addresses. In addition, classifying addresses allows for important quality control measures to be taken, particularly data checks that ensure all necessary elements are present for the specified class (eg. the numbered thoroughfare address class).
- b. The *address feature type(s)*: establish and maintain address feature type (e.g. building, parcel, power pole, etc.) for each address. While classifying addresses by feature type, that is, a category of real world phenomena with common properties whose location is specified by an address, is not recommended in the Address Standard, the Address Feature Type attribute should be assigned to an address record when possible. Many organizations' business processes require knowing what is found (or expected to be found) at a given address. For example, the Census Bureau may need to know if an address is associated with a residential structure, whereas emergency service organizations need to know if the address is for property access, building entrance, etc..
- c. The *address lifecycle status*: establish and maintain the address lifecycle status (i.e. potential, proposed, active, retired) for each address. The lifecycle status of an address is a critical attribute for a host of processes including data assessment, land use planning, accurate field work, and efficient delivery services. A locality should have conditions describing a given lifecycle which should be considered in their quality control program.
- d. The *address official status* indicates whether address components exist as they were established by an official or unofficial entity or is a known error. Examples of values for this address attribute are: "official", "alias", "unofficial", etc.
- e. The *address direct source* is the source from which the data provider obtained the address, or with which the data provider validated the address.
- f. The *address authority* contains the name of the authority (e.g. municipality, county) that created or has jurisdiction over the creation, alteration, or retirement of an address.

-
- g. The *address anomaly status* is a flag or explanatory note for an address that is not correct according to the local Address Reference System (ARS) that governs it, but is nonetheless a valid address. This is valuable for the GEO as it may qualify addresses that might otherwise violate address number parity or sequence.
 4. *Include XY coordinates for addresses in the address data file.* The XY coordinates are needed for most of the address quality measures.
 5. *The data provider should furnish FGDC CSDGM metadata to accompany the address file.* This practice must be followed by federal agencies. The Census Bureau should require address data providers to include a metadata file during address submission in order to inform the GEO process for address assessment and to promote more accurate conclusions on how including partnership data will improve existing address coverage in certain geographic areas in the MTDB.
 6. *Do not abbreviate street types, directionals, landmark names, and community names.* Data managers should avoid using abbreviations as this introduces ambiguity in the address data. Some abbreviations are open to interpretation based on who originated the abbreviation. The most common examples of abbreviated address elements are street types and directionals. Common street types recognized by the USPS such as *St*, *Ave*, and *Ext* should be spelled out as Street, Avenue, and Extension respectively. Directionals such as *NW* or *S* should be stored as *Northwest* and *South*.
 7. *Store address information in its most basic form (simple element).* The Address Data Content part of the Address Standard establishes the data elements used to construct addresses. In practice, the address elements are represented as fields in a table within the address database. The Address Standard contains two types of address elements: complex and simple. A complex element is an address element that is the concatenation of data from two or more elements. For example, the complete street name, such as *MAIN STREET* is a complex element. Complex elements may be parsed into simple elements, which in this case contain the two simple elements *MAIN* and *STREET*.
 8. *Store address data in a normalized relational data base management system (RDBMS).* This recommendation follows general best practices for managing large data sets. Also the data quality measures from the Address Standard are best run from an RDBMS.

5.4.2 Phased Approach for Implementing the Address Standard

As mentioned in the previous sections, the external partner was unable to run many of the data quality measures from the Address Standard for the testing and evaluation tasks. A major challenge recognized during the pilot project was that the external partner's address database does not store critical information required to run many of the data quality measures (e.g., ZIP Codes, *AddressIDs*, joins between address tables and street feature tables, etc.). Based on the results from this exercise, the pilot team anticipates that missing information will be a barrier to governments seeking to adopt the Address Standard and recommends developing a coherent response in the future.

As a first step to developing a response to this issue, the pilot team proposes a phased approach, which implements best practices and a *core* set of data quality measures from the Address Standard. The intention of the core set of measures is to provide governments with a starting point for implementing the data quality part of the Address Standard, and for identifying problems with critical information needed to run additional data quality measures from the Address Standard. An example of the phased approach is outlined in the table below.

Table 4: An example of a phased approach for governments to implement the Address Standard¹

(**) denotes an optional/advanced implementation step for governments with linkages between address points and street features).

Phase 1: Prepare for Implementation of the Address Standard	
Stage	Activity
Stage 1: Planning	Join the Address Standard Implementation Working Group (<i>proposed</i>).
	Get training on the Address Standard and the Spatial Focus Address Quality and Data Exchange Tool (<i>through URISA, from existing training materials from the Census Bureau, or other means</i>).
	Obtain existing Address Standard tools and materials: <ul style="list-style-type: none"> • Census Bureau hosted Address Standard Workshop materials (workbook and training session presentation recording); • Spatial Focus Searchable Standard; • Spatial Focus Address Quality and Data Exchange Tool; • Address Quality and Data Exchange Tool User's Guide
Stage 2: Initial Data Preparation	Ensure that all address records in the local data set contain unique identifiers (Address ID). If not, create a unique identifier for each record.
	Review and consider implementation of best practices as described in Section 5.4.1 of this report.
	** Join address point attribution to street centerlines: <ul style="list-style-type: none"> • ** Identify a street centerline dataset; • ** Link address points to street centerline using a geocoding process.

¹ This approach applies to data sets containing numbered thoroughfare (city-style) addresses.

Phase 2: Evaluate the address data with a core set of data quality measures	
Stage	Activity
Stage 1: Conduct <i>core</i> data quality measures	Data Type Measure- check that elements agree with the specified data type for the columns they populate.
	Pattern Sequence Measure – check that complex elements such as Complete Address Number and Complete Street Name are composed of their simple elements in the correct sequence.
	Uniqueness Measure – check that no duplicate values exist in the Domain table, in preparation for conducting the Tabular Domain Measure.
	Tabular Domain Measure – check stored address element values against a Domain table of acceptable values. This measure identifies erroneous values caused by previous data entry errors or new values that have not been standardized/approved by the addressing authority.
	XY Coordinate Completeness Measure – check addresses for complete and relevant X/Y or latitude/longitude coordinate pairs.
	Related Element Uniqueness Measure – check for uniqueness of the values related to a given element. For example, check the uniqueness of Complete Address Numbers along a given Complete Street Name.
	** Fishbone Measure – check for the spatial sequence of address point locations along the street centerline.
	** Range Domain Measure – check for address numbers that do not fall within the address range for the associated street centerline.
Stage 2: Prepare Data for Exchange	Include the <i>address direct source</i> , <i>address authority</i> , and <i>address anomaly status</i> for each record in the address file.
	Create a CSDGM-compliant metadata file containing the copyright information, use restrictions, contact information, data lineage, data quality, and description of the represented geographic area.
**Phase 3: Implement additional data quality measures	
<i>The GEO or government partners will need to identify other data quality measures and implementation steps for Phase 3 as part of a future project.</i>	

5.4.3 Geography Division considerations for implementing the Address Standard

This section provides conclusions pertaining to implementing the Address Standard within the Geography Division. These conclusions relate to the GEO's mission to improve the MTDB in preparation for the 2020 Census and the GSS-I.

5.4.3.1 Address programs

If participating governments implement the Address Standard to support their day-to-day operations, the GEO will need to accept data in the Address Standard exchange format. The Address Standard is designed to reconcile the problem of data exchange between multiple users. This problem involves receiving address data under a source schema and recreating the source data as accurately as possible. The problems of data exchange include potential loss of data during transit, encoding problems, formatting, and compatibility, which often occur as data are stored in various formats such as XLS, SHP, GDB, KML, etc. The Address Standard specifies the use of XML as the file format for the exchange of address data and metadata between users and the pilot team recommends that the GEO develop software that can accept/ingest partner address data in the standard-supported XML format.

5.4.3.2 Optimum and minimum data submission guidelines

In 2012, the GEO released draft documents of *optimum and minimum address data submission guidelines* and *optimum metadata content guidelines* in support of the GSS-I. These documents outline the address data elements and metadata that the Census Bureau identifies as optimal components in address datasets.

The *FGDC Address Standard Crosswalk* document, *produced by the GEO*, complements the optimum and minimum address data submission guidelines. The crosswalk lists the common terminology used to describe the elements of address data (Census Address Element) used by the Census Bureau and then references the name and description of the equivalent address element found in the Address Standard.

The responsible staff from the GEO should review the findings from this final report and update, if necessary, the optimum and minimum address submission guidelines, the optimum metadata content guidelines, and the FGDC Address Standard Crosswalk document to maintain a consistent reading of the components of the Address Standard. In order to comply with the Address Standard, the GEO should add address authority related attributes (required attributes listed in Table 5) to the optimal address guidelines and metadata guidelines.

5.4.3.3 Metadata

The pilot team also suggests that the GEO develop software that can accept/ingest address attributes, and accept/ingest information from any CSDGM-compliant metadata file that accompanies a data submission. The GEO should also consider creating a schema to validate

incoming metadata for minimum and optimum address elements per the guidelines. In addition, staff should develop examples of CSDGM-compliant metadata files for the partnership programs.

5.4.3.4 Data quality measures for partner data

The pilot team recommends that GEO consider applying a core set of data quality measures, such as the ones listed in Table 4 under Section 5.4.2, to partner address data submissions. The results from this process could feed into quality indicators (QIs) to support the assessment for targeted address canvassing.

If the GEO does implement the application of some Address Standard quality measures on partner data submissions, the GEO should consider providing a metadata file with quality measure results to the data submitter as a feedback product. This would be an effective gesture to lead by example with standards implementation and may assist government partners with developing and updating their own standards-compliant metadata and data exchange files.

Similarly, whenever the GEO receives Address Standard related data quality information from government partners, it will need to know what these measures are testing and how to interpret the results. If a government partner submits an address file that contains a quality measure at an 80% level of conformance, does such information add any meaning to assessing the quality and confidence from the standpoint of the GEO? Policy and direction would be required in order to assess the meaning of these data quality reports. The data quality reports from the metadata file may be one of the metrics that the GEO uses to establish quality indicators of address data. The following examples show how the data quality measure reports are recorded in the metadata file:

Example 1: Completeness Test Report

```
Tested Address Completeness Measure at 87% Conformance
```

Example 2: Logical Consistency Test Reports

```
Tested AddressLeftRightMeasure at 85% Conformance
```

```
Tested AddressNumberFishbonesMeasure at 80% Conformance
```

Example 3: Positional Accuracy Test Report

```
Tested XYCoordinateSpatialMeasure at 90% Conformance
```

Example 1 shows the results of the address completeness measure. Example 2 shows the results from a few measures used to test the logical consistency of the address data. Example 3 shows the results from one of the tests that measures the spatial accuracy of the address coordinate information.

5.4.3.5 Street centerline file

Many of the logical consistency tests in the Address Standard require a street centerline file containing street names and address ranges as one of the input data. This is particularly useful in identifying inconsistencies in the address number parity and ordering in the address data file. The source street centerline file could conceivably be received as part of the data exchange package from the government partner. However, the GEO could use the TIGER feature data as the control when recreating these quality measures after receiving the address data from a government partner. If the address data is deemed to be of sufficient quality, some of the tests could be used to check the TIGER feature data against the address file to identify discrepancies in address ranges and feature names.

The Address Standard describes a procedure for geocoding the address points to the centerline of the street feature as a means for linking address information with the street features. This process could be accomplished using TIGER feature data assuming that there are no major discrepancies between the source's address street name and the feature name and that the Census Bureau address range is joined to the TIGER feature. If the street centerline is to be used to assess the quality of address data, or to assist Census staff in identifying potential errors in the data, the GEO must draft step-by-step instructions on how to link the source address data to the TIGER feature using some geocoding process.

5.4.3.6 The Census Bureau's role as the Address Standard maintenance authority

The Census Bureau is the maintenance authority of the Address Standard. The primary responsibility of the maintenance authority is to initiate the public review of the standard at least once every five years. The FGDC endorsed the Address Standard in February 2011 and the review is scheduled to occur in 2016. The Census Bureau is also expected to assist users with implementation strategies, including user guides, training, public relations, workshops, etc. The Census Bureau could promote the Address Standard by mirroring the FGDC's Metadata Working Group and the National Oceanic and Atmospheric Administration's (NOAA) activities in promoting the International Organization for Standardization (ISO) metadata standard (ISO 19115- Metadata). Their work was oriented toward other federal agencies but it is an effective model: foster a community of interest; create a web page with resources to support the community; create training materials and user's guides; conduct training workshops/classes, and conduct a series of webinar events. The Census Bureau could also work to have FGDC more involved in supporting and promoting the Address Standard through the existing Standards Working Group or other subgroups. These activities would be coordinated with organizations such as URISA, NENA, and NSGIC.

6. Recommendations

Based on the results from the four tasks discussed in Section 5, the pilot team has identified 11 recommendations that fall within the scope of this address pilot.

6.1 Recommendations

The Reference Sections column in the table below links the recommendation to the related section(s) and discussion in the report.

Table 6: Address Standard Implementation Recommendations

No.	Description	Impact	Reference Sections
1	<i>Establish an Address Implementation Working Group of external stakeholders to help promote the use of the Address Standard and participate in implementation recommendation activities.</i>	<i>Government partners</i>	5.4.3.6
2	<i>Develop a comprehensive training plan/program to assist government partners in implementing the Address Standard.</i>	<i>Government partners</i>	5.2.1.1 5.2.2.2
3	<i>Develop a guide and/or workbook for the Address Standard similar to the workbook for the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) and the workbook for the ISO metadata standard</i>	<i>Government partners</i>	5.4.3.6
4	<i>Establish a GEO Working Group to plan for the internal impact of the Address Standard implementation pilot project recommendations.</i>	<i>Internal to GEO</i>	3 5.1.6.2 5.1.6.3 5.3.3 5.5.1
5	<i>Evaluate the application of the Address Standard quality measures to: 1) assess the quality of partner data and, 2) assess how current MTDB quality tests align with the quality measures.</i>	<i>Internal to GEO</i>	3 5.3 5.4
6	<i>Develop Address Standard training for GEO staff tailored to the immediate needs for the GSS-I and 2020 Census.</i>	<i>Internal to GEO</i>	5.1.6.2 5.1.6.3 5.3.3

No.	Description	Impact	Reference Sections
7	<i>Have the partnership branches review this final report and assess what parts of the standard would improve existing partnership programs.</i>	<i>Internal to GEO</i>	5.4.3
8	<i>Develop procedures for address <u>submission</u> based on issues raised by the Address Standard, including but not limited to: required/recommended address data quality measures for partners to conduct prior to submitting addresses to the GEO and instructions for creating a CSDGM metadata file to accompany the address file.</i>	<i>Government partners, Internal to GEO</i>	5.4.1 5.4.2 5.4.3.2 5.4.3.3
9	<i>Develop procedures for address <u>retrieval</u> based on issues raised by the Address Standard, including but not limited to: required/recommended data quality measures to be conducted by the GEO after receiving the data from partners, and a process for ingesting XML address files and metadata files received from partners.</i>	<i>Internal to GEO</i>	5.4.3.1 5.4.3.4
10a	<i>Conduct a pilot of the address quality and exchange tool with volunteer partners to assess the value of this tool for improving data quality testing and exchange.</i>	<i>Government partners, Internal to GEO</i>	5.2 5.3
10b	<i>Consider incorporating the address quality and data exchange tool or concepts from the tool into a future integrated data exchange system such as Community TIGER.</i>	<i>Government partners, Internal to GEO</i>	5.2.1.3
11	<i>Establish a web page for public collaboration and support of the Address Standard.</i>	<i>Government partners</i>	5.4.3.6

6.2 Discussion of the Recommendations

1. *Establish an Address Implementation Working Group to help promote the use of the Address Standard and participate in implementation recommendation activities. This group would be modeled after the FGDC Metadata Working Group and open to*

interested staff from federal agencies, government partners, and organizations to collaborate and identify implementation needs and solutions. Activities could include developing a training plan and identifying best practices. This group would serve as the working group referenced under the phased approach in Section 5.4.2. This working group would be in addition to the existing Address Standard Working Group (ASWG). The Address Standard Working Group was formed and supported by URISA to produce the Address Standard. While the ASWG is still an active group, their focus is on developing the standard and the address standard tools. There is a need for a group that would focus on developing strategies for implementing the standard. The proposed implementation working group would work in conjunction with the Census Bureau to encourage adoption of the standard. We envision that this implementation group would work closely with the ASWG and the FGDC. The implementation working group could eventually become an FGDC entity depending on the outcome of upcoming National Geospatial Advisory Committee (NGAC) recommendations related to address data.

2. *Develop a comprehensive training plan/program to assist government partners in implementing the Address Standard.* The pilot team recommends that the Census Bureau provide materials on how to implement the Address Standard for interested governments. These materials primarily target the needs of local governments for their internal needs and would not be limited to address data submissions to the Census Bureau. The intent is to promote a broader adoption of the Address Standard among governments throughout the country. At the very least, if address authorities are utilizing the quality measures from the Address Standard, the GEO should expect to see higher quality data from address submissions. The most critical step is instructing governments on how to begin the implementation process. The pilot team expects that many local address databases will have preexisting deficiencies that will preclude government partners from implementing the Address Standard. The pilot team recommends the following to steps to assist government partners:
 - a. Publicize to government partners the ZIP Code products from the U.S. Postal Service Address Information System (AIS). These products may be used to append ZIP code information to address data.
 - b. Develop one or more state and local implementation models (based on the phased approach discussed in Section 5.4.2)
 - c. Provide a preliminary training guide to the Address Standard to a subset of government partners and conduct a follow-up session to assess needs and see areas that need assistance from the Census Bureau.
3. *Develop a guide and/or workbook for the Address Standard similar to the workbook for the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) and the workbook for the ISO metadata standard.* This recommendation supports the second recommendation. In addition to a formal training program, the U.S. Census Bureau may provide a workbook, similar to existing metadata standards workbooks such as the one provided to the public by the FGDC for the Digital Geospatial Metadata Content

Standard: http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf or the [ISO metadata workbook: www.ncddc.noaa.gov/metadata-standards](http://www.ncddc.noaa.gov/metadata-standards).

4. *Establish a GEO Working Group to plan for the internal impact of the Address Standard Implementation Pilot Project recommendations.* The GEO will need to support the recommendations being made in this report. Recommendations 5 – 9 in Table 6 above will require various subject matter experts to discuss and plan tasks and develop specifications and requirements. The team expects the formation of several subgroups to work on recommendations 5 – 10. Topics include: application of the quality measures; providing additional internal training; data submission procedures; and accommodating the data exchange format.
5. *Evaluate the application of the Address Standard quality measures to: 1) assess the quality of partner data and, 2) assess how current MTDB quality tests align with the quality measures.* The Address Standard describes 39 measures of data quality. These quality measures may be helpful in determining the quality of partner address data. In addition, GEO would benefit from understanding how the current MTDB quality testing processes relate to the Address Standard quality measures.
 - a. Establish a core list of quality measures from the Address Standard that government partners are encouraged to conduct on address data prior to submitting to the GEO. This may lead to improving the initial list of data quality measures listed under Phase 2 of the phased approach to implementing the Address Standard, discussed in Section 5.4.2.
 - b. Build staff competencies in the quality measures in order to recommend implementation strategies for states/local governments.
 - c. Identify which tests in the Address Standard are currently implemented in the GEO and deliver recommendations on tests currently not implemented that have the potential to improve existing QC processes within the GEO.
6. *Develop training for GEO staff tailored to the immediate needs for the GSS-I and 2020 Census.* In addition to training for local governments, the pilot team recommends a structured training plan for internal Census staff. The major components of the Address Standard must be understood by GSS-I support staff in the relevant subject matter and programming branches in the GEO (refer to the discussions under Section 5.5).
7. *Have the partnership branches review this final report and assess what parts of the standard would improve existing partnership programs.* The Address Standard describes address data content, classification, data quality, and data exchange. These topics may overlap with requirements for existing partnership programs in the GEO. Also, the Address Standard contains steps that may improve the existing processes for the GEO partnerships. Other recommendations (see 7, 8 and 9) propose the adoption of the Address Standard for data exchange, which will involve the coordination of partnership branches. This pilot team also recommends the following:

-
- a. Develop a means of tracking partners interested in reviewing Census Bureau training materials, procedures, and documentation developed during the pilot project.
 - b. Conduct a follow-up study on respondents of the Implementation Questionnaire that had implemented the Address Standard.
 - c. Continue to gather best practices for implementing the Address Standard to share with other partners.
 - d. Evaluate the impact of the implementation of the Address Standard on the quality of data shared with the Census Bureau.
8. *Develop procedures for address submission based on issues raised in the Address Standard.* The implementation of the Address Standard will require additional steps for the existing data submission processes. The GEO will need to establish procedures for government partners to follow when submitting address data to the Census Bureau. The working group from recommendation 4 would inform the development process for these procedures.
 9. *GEO will need to define procedures for address retrieval based on issues raised in the Address Standard.* Just as the GEO needs to define procedures for government partners to submit address data (see recommendation 7), the GEO will also need internal procedures on how to process these data after receipt. The data exchange section of the Address Standard instructs users to use XML as the exchange file format and that the exchange package contains a CSDGM-compliant metadata file with information about the address authority, publication date, address ID, address source, and known address anomalies. The metadata may also contain data quality measures conducted by the government partner. The GEO will need to have procedures in place on how to handle this information. Also, the GEO will need to assess whether the metadata can contribute to the quality indicators (QI) in support of the Confidence Analysis and Tracking Tool (CATT).
 10. Spatial Focus demonstrated an address data quality and data exchange tool (Address Standard Tool) as part of the two-day workshop hosted at the Census Bureau. After the workshop, the tool experienced a few upgrades and a version with a more intuitive user interface is expected. Spatial Focus shares the source code for this tool to the public on github (www.github.com search: "FGDCAddressTool").
 - a. *Conduct a pilot of the modified address quality and exchange tool with volunteer partners.* Since the tool is available to the public, the GEO should consider conducting a pilot study with procedures for government partners to assess the quality of test data. Other recommendations for this pilot include:
 - i. Recommend and support further development of the Spatial Focus address quality and data exchange tools including detailed documentation, a user's guide,

an intuitive GUI, and possibly a web-based version of the tool. Support would be coordinated by the proposed Address Implementation Working Group from recommendation #1 and come from the working group participants and other volunteers from within the federal government, state/local governments, or the general public.

- ii. Commit a GEO programmer resource to support the internal use of the address quality and data exchange tool.
 - iii. Recommend a comparable address tool that operates within the Esri ArcGIS environment. Since many government address authorities use Esri software to maintain their data, an Esri-supported tool would assist governments in implementing Address Standard data quality measures.
- b. *Incorporate the address quality and exchange tool or concepts from the tool into a future integrated data exchange system.* The working group, as described in recommendation 4, would inform what types of data quality measures the GEO would adopt as part of the data exchange process. In addition, address data discussed in the previous sections of this report may have a role in how the GEO tracks data quality and authority. The existing address quality and data exchange tool may facilitate this process. This pilot recommends that the GEO consider using this tool or concepts from the tool as part of any planned address data exchange systems in the future.
11. *Establish a web page for collaboration and support of the Address Standard.* The U.S. Census Bureau is the maintenance authority (see section 5.5.3) of the Address Standard. This pilot deems it necessary for the Census Bureau to ensure that a web page exists that is devoted to the Address Standard, preferably within the FGDC web site. The web page would include the following:
- a. Brief introduction to the Address Standard,
 - b. A synopsis of the work the GEO has done to support the Address Standard:
 - i. Link to the Census Address Data Submission Guidelines
 - ii. Link to the Census- FGDC Address Standard Crosswalk document
 - c. Frequently Asked Questions (FAQ) section for the Address Standard, and
 - d. A link to the hardcopy address standard, the electronic searchable standard, and the prototype address quality and data exchange tool.

7. Acknowledgements

The pilot team would like to thank our external partner for volunteering his time to contribute to this project and the invaluable perspective and technical knowledge he brought to the team. Also,

thanks to our participants that responded to our Implementation Questionnaire. The information we received was informative and shed new light on what governments have experienced during their implementation of address standards. Also, applause should be in order for those pioneering governments that have already implemented the Address Standard. We hope that those governments continue to assist the Census Bureau. Thanks to staff at the Census Bureau that participated in the Address Standard Workshop and provided their feedback. Thanks to the management team for guidance on Title 13 related concerns and OMB requirements. And, special thanks to Spatial Focus, who provided additional time in assisting the pilot team understand the details of the Address Standard and the quality and data exchange tool.

8. Attachment A: Terminology and Acronyms

Term	Definition
FGDC Address Standard Address Standard	The United States Thoroughfare, Landmark, And Postal Address Data Standard endorsed by the FGDC in February 2011.
FGDC Address Standard Tool	The address quality and data exchange tool developed by Spatial Focus, Inc. as part of the FGDC CAP grant study

Acronym	Meaning
APMB	Address Programs Management Branch
CAP	Cooperative Agreements Program
CSDGM	Content Standard for Digital Geospatial Metadata
FGDC	Federal Geographic Data Committee
GEO	United States Census Bureau—Geography Division
GSS	Geographic Support System
GSS-I	Geographic Support System Initiative
GPMB	Geographic Project Management Branch
GPQMB	Geographic Process and Quality Management Branch
ISO	International Organization for Standardization
NSGPB	National/State Geographic Partnerships Branch
OMB	Office of Management and Budget
RDBMS	Relational Database Management System
SQL	Structured Query Language
USPS	United States Postal Service

9. Attachment B: References

Document Title, Author	Document #	Version, Date	Location
N/A			

10. Appendices

- A. Spatial Focus FGDC Address Standard Training Agenda
- B. Spatial Focus FGDC Address Standard Training Handout
- C. Training Evaluation Form
- D. Training Evaluation Form Summary of Responses for Questions 1 – 10
- E. The FGDC Address Standard Implementation Questionnaire
- F. The FGDC Address Standard Implementation Questionnaire Responses
- G. Spatial Focus Comments
- H. Guide for Installing and Using the FGDC Address Tool