TOWSON UNIVERSITY

Center for GIS

Structures Data Themes

For

Maryland and the National Capital Region

08HQAG0060

June 2008 – August 2009

MSGIC

Maryland State Geographic Information Committee

Towson University Center for GIS

NSDI CAP 2008 Category 5:

Building Data Stewardship for *The National Map* and the NSDI

FINAL Report

Date	October 30, 2009
Agreement Number	08HQAG0060 5/1/08 – 4/30/09 + extension to 8/1/09
Project Title	*Structures Data Themes for Maryland and the National Capital Region*
Type of Report	Final Report
Organization	Center for GIS at Towson University 8000 York Road Towson, MD 21252-0001 http://cgis.towson.edu
Project Leader	Primary Investigator Ardys Russakis, Operations Officer, Center for GIS 410-704-3887 arussakis@towson.edu Project Lead Kenneth Juengling, Project Manager, Center for GIS 410-704-3887 kjuengling@towson.edu Project Coordinator Missy Valentino, GIS Specialist III, Center for GIS 410-704-3887 mvalentino@towson.edu
Collaborating Organizations	Metropolitan Washington Council of Governments GIS Committee http://www.mwcog.org/ Barney Krucoff, GIS Manager, District of Columbia, OCTO 202-727-2277 Barney.Krucoff@dc.gov Maryland State Geographic Information Committee http://www.msgic.state.md.us/

	Mr. Kenneth M. Miller, Chair, and Maryland's Geographic Information Officer 410-260-8987 kenmiller@dnr.state.md.us Participating Counties and Jurisdictions Baltimore County, Maryland Frederick County, Maryland Montgomery County, Maryland District of Columbia Fairfax County, Virginia
USGS Geospatial Liaison	Roger A. Barlow 703-648-5189 rbarlow@usgs.gov
Data Themes	Landmark Structures

Executive Summary

The Landmark Structures data sharing project established an automated, sustainable means to collect and document landmark structures data for the Baltimore-Washington area and the National Capital Region for *The National Map*. Stakeholder jurisdictions included Baltimore, Frederick, and Montgomery Counties in Maryland; the District of Columbia; and Fairfax County, Virginia. Stakeholders agreed to adopt the USGS Best Practices Data model for the project. A Memorandum of Understanding (MOU) and an alternative Letter of Intent to share data were developed. Baltimore County and Montgomery County signed the MOU; Frederick County and the District of Columbia signed the LOI. In a pilot phase, the technical team worked closely with the Baltimore County stakeholder to develop the data sharing and update process. Examination of the data model revealed the need to develop logical steps to address each jurisdiction's unique datasets then collect and transform the datasets individually. The model for each jurisdiction can be used in the future to create an updated dataset. The final output from each jurisdiction's model was exported directly into the Structures database and manually imported into the appropriate feature class. The data and metadata were submitted to USGS on disc. Metadata was compiled and will be published to www.marylandGIS.net.

Project Narrative

A: Project Description

The *Structures Data Themes for Maryland and the National Capital Region* (NCR) project assembled and contributed FGDC landmark structures data for jurisdictions within NCR to *The National Map* through an automated, sustainable process, using a common standard. A consistent data model for the project based on the USGS model was established and agreed upon by the stakeholders.

To encourage additional regional jurisdictions to contribute data in the future, using the sustainable process, the project team worked toward building regional partnerships and also created a broad standard that allows inclusion of detailed attributes.

Tasks

The project team completed tasks in the following major categories.

Meetings, Communication, and Documents

- Scheduled and executed group meetings for effective collaboration and results. See *Table 1: Meetings and Presentations* on page 7of this report.
- Collaboratively drafted a Memorandum of Understanding for sharing data and vetted it through county administrators and/or their legal departments. Since some stakeholders preferred a less formal agreement, the project team created an alternative Letter of Intent to share data.

  - Received signed MOUs from Baltimore and Montgomery Counties. Sent three originals of the MOUs to Paul Wiese's Denver, CO address on June 19, 2009.
  - Received signed LOI from Frederick County, Maryland, and Washington DC.
- Obtained agreement on accepting the points of interest data that each jurisdiction collects as the definition of Landmark Structures for this project.

Data

- Compiled a data survey, attached to this report as *Appendix A.*
  - Sent the data survey to each stakeholder jurisdiction.
  - Received completed surveys from Washington, D.C. and Baltimore County.
- Baltimore County volunteered to serve as the pilot jurisdiction to help develop a process for transforming the data to conform to the data model. It was determined that four steps must be completed:
  1. *Develop a better understanding of the USGS Best Practices data model attribute definitions.*

     Paul Wiese, the USGS point of contact for this project, was consulted. Based on documentation provided by USGS, the team developed a data dictionary, included in this report as *Appendix B*. The data dictionary was used to map each jurisdiction's attributes to the correct fields in the final dataset.
  2. *Determine which fields in each jurisdiction's dataset would best translate to the f-code*.

     The value had to be determined for each feature in each jurisdiction's dataset using existing attributes. A different process was used for each jurisdiction to extract this information from the data.
  3. *Obtain information about each county's data, data sharing, and security policies, and also obtain a 100-character description of the dataset*.

     The data survey was sent to each jurisdiction. Washington, D.C. and Baltimore County returned completed surveys.
  4. *Develop a repeatable process to transform the jurisdiction's data into the data model.*

     The process was intended to simplify later updates and allow the data to easily migrate into the data model. The ESRI Spatial ETL tool

(Extract Transform and Load) worked best for this process and also allows the user to build a model to manipulate and transform data to a new output data source. The model can be used repeatedly until data schemas change. If data schemas change, then the ETL tool would must be updated to reflect the changes.

Data Sharing Process

- Established, tested, and presented the technical process for collecting data and inputting the data into the model.
- Received data from Baltimore County, Frederick County, Montgomery County, and Washington, D.C.

Send Data to *The National Map*

- The data and metadata were copied to DVD and sent to Laurie Temple at USGS in Bozeman, Montana.

Compile and Publish Metadata

- Metadata was compiled for the data received.
- Metadata will be published to the Maryland Mapping Resource Guide at www.marylandGIS.net after USGS publishes the data to *The National Map*.

Conduct Outreach and Communication to Regional Stakeholders

- Presented Landmark Structures Project update at conferences:
    - TUgis, March 16, 2009, at Towson University in Towson, Maryland http://tugis.towson.edu
    - Pennsylvania GIS Conference, May 19, 2009. http://www.pagisconference.org/
    - ESRI International Users Conference, July 14, 2009, in San Diego California. http://www.esri.com/events/uc/index.html

Table 1: Meetings and Presentations

Meeting Date	Group/Location	Agenda
April 2008	TU-CGIS Team meetings	Plan project strategy
April 2008	Teleconference with Paul Wiese	Kick off project; Provide USGS answers to logistical and technical questions
May 2008	Teleconference with Paul Wiese to coordinate Category 5 states	States presented their project approach
June 24, 2008	First Stakeholder Meeting/MW-COG office in DC	-Project overview and details -Budget tracking -Plan meeting schedule (4 meetings) -Discuss documentation -Present SHA's Landmark Data project as example -Obtain feedback
July 2008	TU-CGIS Team meetings	Proceed with action items and planning
July 22, 2008	Stakeholder meeting/MW-COG office in DC	-Follow-up from previous meeting, and status of action items -Discuss draft of MOU -Discuss data issues
July 29, 2008	Stakeholder meeting/MW-COG office in DC	-Follow-up from previous meeting, and status of action Items -Review and discuss data model -Discus symbology -Discuss ongoing action items and issues raised
October 21, 2008	TU-CGIS office with Baltimore County representatives	-Further examine and discuss USGS data model, -Review process for populating the model -Review definitions and codes (county, municipal FIPPS, NAICS, etc.)
November 25, 2008	Project team. Frederick County, Maryland. 117 East Church Street.	Laurie Temple and Paul Wiese of USGS joined the meeting by teleconference. • Update on the current status of the project • USGS representative: discuss the Best Practices Data Model and answer questions • Discuss the process followed, lessons learned, and outcome of importing Baltimore County data into the data model. • Address concerns and develop a plan to move forward with participating jurisdictions.
March 16, 2009	TUgis Conference, Towson, Maryland	Presented project background, structure, and status update on the project at the annual TUgis conference at Towson University.
May 19, 2009	PA GIS Conference, Grantville, PA	Presented background, structure, and status update on the project at the annual Pennsylvania GIS Conference in Grantville, PA.
July 14, 2009	ESRI Conference, San Diego, CA	Presented background, structure, and status update on the project at the annual ESRI International User Conference held in San Diego, CA.
October 20, 2009	Teleconference with Laurie Temple and Paul Kimsey of USGS.	Meeting to discuss the technical aspects of pushing the data to *The National Map*. The meeting resulted in the following:

		• Memorandum of Understanding (MOU/Agreements) for Baltimore and Montgomery Counties: Paul Kimsey will shepherd these documents through USGS signatory process. • Structures data delivery for project: updated data will be sent to USGS after CGIS receives signed MOUs back from USGS.

Highlights

- Worked closely with Baltimore County to develop, test, and present the data collection and transformation process.
- Obtained signed MOUs from Baltimore County and Montgomery County. Obtained signed LOIs from Washington, D.C. and Frederick County.

Challenges

- *Stakeholder collaboration*
    - Two of five stakeholders responded to the data survey.
    - One stakeholder could not obtain signed MOU or LOI.
- *Standard symbolization for data*: Data for *The National Map* are symbolized by f-type. The project team intended to create an .axl file and asked USGS for guidance relative to a standard set of symbols. The team was informed that USGS had created a structures symbol set, but the set was not yet officially approved or released by the end of this project. Symbology for structures would be tied to a new .mxd format (*not* .axl) in a beta viewer environment that USGS is currently testing. The design is intended to align with new USGS digital graphics at the 1:24,000-scale. The symbol style file has been created for some features but not yet finalized, and probably will not be released until USGS is closer to placing structures on the digital graphics sometime in 2010.
- *Identifying feature type for data*: The feature type for the data was difficult to determine due to the differences in the original datasets. The jurisdiction with the most difficulty was the District of Columbia, which does not maintain feature codes. The project team worked with the existing attributes of the data to develop feature codes; however, the stakeholder decided that the method could add error to the dataset and that a general code for all of the data would be best.

Accomplishments

- Data survey, MOU, and LOI were drafted, edited, and completed.
- Data model, data process, and outcomes were discussed at stakeholder meetings.
- Stakeholders agreed to adopt the *USGS Best Practices Structure Data Model* to accommodate data in its existing form, yet be inclusive for future growth.
- Baltimore County worked with CGIS as the pilot jurisdiction to import data into the adopted data model.
- Formal data sharing agreements were made with 4 of the 5 participating jurisdictions.

- Completed a Data Processing white paper (included in this report as *Appendix C*) describing the technical process used to transform each jurisdictions' data to conform to the USGS Best Practices Data Model.
- Completed metadata for the final dataset and will publish the metadata to www.marylandGIS.net, which is available to GeoSpatial One Stop.
- Sent the data on DVD to USGS to be published to *The National Map*.
- Completed a table containing metadata details about each jurisdiction's datasets.

Impediments to participating in *The National Map* can include concerns about data security and retaining data ownership; whether or not policies and procedures are established that allow the jurisdiction to share data; and whether or not a jurisdiction collects the data under consideration. The project attempted to help jurisdictions overcome impediments by:

- Obtaining consensus on a method for jurisdictions to state their intention and commitment to share data by collaboratively creating and then vetting an MOU through the jurisdictions' legal departments, and by creating an LOI for jurisdictions that preferred a less formal sharing agreement.
- Creating and documenting a process for sharing data that all project stakeholders can follow. The documented process may help overcome the presumption that data sharing will unnecessarily increase a jurisdiction's workload.
- Defining the data themes and attributes to be shared in a way that considers the differences among jurisdictions relative to their datasets.
- Obtaining agreement on data model, data standards, symbology, and other technical considerations.
- Determining layers of jurisdictional bureaucracy (if any) that must be satisfied before data sharing can happen.

Recommendations for Project Success

The project team acknowledged the challenges for ongoing participation in any effort that requires adoption of procedures and policies that are not currently built into the existing business processes of an organization, as well as the need for further funding.

Technical

The project team agreed to use the USGS Best Practices Data Model for Structures and also frequently consulted USGS for guidance on technical matters. The team also considers the data processing white paper to be a good reference for participating jurisdictions that may need technical guidance. USGS can assist by offering more technical information and guidance early in the process.

Institutional

The project team recommends that an MOU and/or Letter of Intent be considered as an initial step for any organization undertaking a data sharing effort. The team also recommends that the Data Model for this project be the "starting point" for consideration of future participants and be adopted and modified as necessary. Ongoing update of the data should be considered a necessary component of future efforts.

Organizational

Each partner organization must make a commitment to the project's success and to sustaining the data update and sharing process.

B: Data content provided to The National Map

Data Themes

The project assembled structures data themes, including the geospatial location, label, classification, and other characteristics of manmade facilities, specifically, landmark buildings that would appear on USGS quadrangles, such as schools, post offices, courthouses, fire stations, and railroad stations. The team researched The National Model, The Fire Data Model, and the USGS Best Practices Data Model, as well as existing applicable datasets, such as Maryland's State Highway Administration (SHA) Points of interest, Maryland's Department of Natural Resources (DNR) Lands, and Homeland Security Infrastructure Program (HSIP) documentation. Stakeholders also adopted the description of the data theme, as follows

"*The structures data theme is comprised of manmade features important to planners, land managers, utility companies, and the general public for a broad range of analyses and applications. This theme is key for the locations of critical structures, which are of vital interest to emergency responders. The data include those from Federal partners including agencies of the Departments of Homeland Security and Defense, and State and local government agencies*."

The final dataset consists of point locations, provided by each jurisdiction, of landmark structures as defined above.

Use Restrictions

Jurisdictions collect and maintain structures data for multiple purposes and needs. When considering which model to adopt, stakeholders had to ensure that no jurisdiction would be

prevented from participating by being required to maintain information that they currently do not collect, or to share data that they are not able to make publically available due to homeland security issues. The model also had to allow for growth as jurisdictions begin to maintain more information and comply with Federal government requirements for homeland security purposes. The USGS Best Practices Data Model includes all categories of structures listed in **Table 2**. Jurisdictions determined the categories of data and attributes to make publically available. In the future, the same model may be used to create a regional dataset for homeland security purposes.

Table 2 USGS Best Practices Data Model Structures Categories

USGS Structures Categories
Agriculture, Food, and Livestock
Banking and Finance
General Building
Commercial and retail
Education
Emergency Response and Law Enforcement
Energy
Government and Military
Health and Medical
Industrial
Information & Communication
Mail and Shipping
Public Attractions and Landmark Buildings
Transportation facilities
Water Supply and Treatment
Weather

Geospatial One-Stop

Metadata was compiled and will be published to www.marylandGIS.net, which is harvested by Geospatial One-Stop, after USGS publishes the data to *The National Map*.

Status of maintaining, updating and serving data to The National Map

This project leverages existing data collection and maintenance activities at the local level for participating stakeholders. As each stakeholder continues to maintain local data according to local business needs, participation in sharing and serving the data will continue on a voluntary basis. A process is in place for stakeholders to use existing policies and procedures to participate in ongoing data sharing and submission to *The National Map*.

User Requirements for National Level Spatial Data

The final dataset meets three levels of requirements.

1) Local Data Requirement – The most spatially accurate, up-to-date data is maintained by local jurisdictions for decision making and public safety.
2) Regional Data Requirement – Combining the local jurisdiction data into a regional data set supports public safety and decision making across local and state borders.
3) National Data Requirement – The national dataset provides spatially accurate data to support public safety and security, emergency, and other planning, as well as natural resource management, environmental stewardship, human services, and other uses.

C: Operational capability to maintain and update data

The technical functionality of the Spatial ETL tool enables each jurisdiction to conform their data to the USGS Best Practices data model by using one tool, which can also be utilized to update the data. Currently, USGS cannot receive automated updates but prefers instead to receive data and updates on CD.

D: Issues, Difficulties, and Challenges

Technical

- A Points of Interest dataset is currently made publically available by the Maryland State Highway Administration (SHA). SHA's data includes landmark features; however, the data are not detailed. The data that will be collected through the landmark structures project includes all available detailed attributes and accurate spatial data that are maintained by local jurisdictions.
- Developing a solution for overlap areas (polygon buildings, for example) was not included in this project and was considered by the project team to be outside the scope.
- The project team developed the following technical process for receiving jurisdictions' data and sharing the data to *The National Map,* as articulated in the MOU:

  "JURISDICTION will provide TU-CGIS with data. TU-CGIS will utilize Environmental Systems Research Institute's (ESRI) Data Interoperability extension to perform an Extract Transform and Load (ETL) process. This process will extract the data from the original format, transform that information to fit the regional data model, and load the data into the regional dataset."

Institutional

Each stakeholder jurisdiction must agree to share data. Stakeholders collaborated on a memorandum of understanding (MOU) and a letter of intent (LOI). The MOU was reviewed and approved by two jurisdictions' administrative and/or legal departments. Three jurisdictions preferred the LOI to the more formal MOU. Two of those jurisdictions returned a signed copy of the LOI. One jurisdiction could not obtain the appropriate signature or the data.

The MOU and LOI do not request agreement from stakeholders to accept USGS standards, which could present a local legal issue, since counties have different boundaries represented for shared administrative edges.

Organizational

During MOU development, thoughtful discussion occurred relative to details that should or should not be included in the formal agreement. Maryland participants wished to include references to current statewide activities; however, Washington, D.C and Virginia participants expressed concern that the MOU would then be too Maryland-specific. To create a less restrictive document, participants agreed to focus on details that *directly* relate to the landmark structures project. The approved MOU is intended to function as the "base" language that any jurisdiction can subsequently amend to be more specific.

E: Relationship with USGS

The project team enjoys an excellent working relationship with USGS, having been a partner on NSDI Cooperative Agreement Program grant projects since 2000. Roger Barlow, USGS Eastern Region NSDI liaison, proactively works for project success by attending project meetings, being readily available at other times, and keeping the team informed of all relative developments. The project team has also been in contact with Paul Weise and Laurie Temple relative to technical challenges and anticipated outcomes.

Towson University has not established a formal, written, ongoing agreement with USGS to provide data to *The National Map*.

Plans for Follow-on Activities

The project team will explore any opportunities that may exist to fund updates as part of other regional and/or statewide data initiatives.

On behalf of and in support of the mission of the Metropolitan Washington Council of Governments (MWCOG) GIS Committee, TU-CGIS received funding from USGS to collect,

document, assemble, and distribute the transportation data theme for the National Capital Region plus Frederick County Maryland, to *The National Map*. The similar partnerships comprise concurrent, overlapping geographies with the landmark structures data project.

G: Feedback on Cooperative Agreements Program

NSDI cooperative agreement (CAP) projects have helped Maryland build partnerships for geospatial resources throughout the state and the region and have given team members the opportunity to develop technical data sharing methods.

H: Program Strengths and Weaknesses

Strength: The program provides seed funding for collaborative geospatial data sharing projects. *Weakness*: a) Sustainability is difficult to achieve, especially when economic downturns have a negative effect on the effort to secure additional funding to maintain momentum. b) Adequate infrastructure for *The National Map* does not appear to be in place.

I: Where does the program make a difference?

The program opens communication and enables positive action relative to sharing and updating geospatial data.

J: Was the assistance you received sufficient or effective?

Paul Wiese and Laurie Temple were consistently available by teleconference and e-mail to address concerns about the data model and other pertinent technical issues, although a method for publishing data remains unresolved. USGS is changing the way data is published to *The National Map*. This process is not fully established, and USGS is not prepared for automated data sharing to *The National Map*.

K: What would you recommend doing differently? Are there factors that are missing or additional needs that should be considered?

Please refer to the response to Item **J** above, and discussion on Page 9 relative to symbology and the project team's intention to use .axl files.

M: Program Management Concerns

Concerns include USGS inability to provide the updated data model and standard structures symbology for *The National Map*. USGS also cannot process updates using a federated system (such as replication) at this time.

N: If you were to do this again, what would you do differently?

Ideally, USGS would have specific technical requirements and standards in place for *The National Map* prior to project initiation.

O: Terms and Mutual Commitment of Resources

As a result of early NSDI CAP clearinghouse node and metadata projects, Towson University began and continues to host several components of Maryland's spatial data infrastructure and Maryland's base map (MD iMap). Maryland's vision for MD iMap aligns with NSDI's vision and not only provides government agencies with a consistent cartographic experience for Maryland, but serves as the secure means through which Maryland government agencies can and will share geospatial data.

APPENDIX A

Towson University Center for GIS
NSDI Landmark Structures for *The National Map*
Project # 08HQAG0060
DATA SURVEY

Data Specifications Questions

Please complete the table below with the appropriate information about your structures data. If you have existing metadata with this information, please indicate that where applicable and attach your metadata in the reply e-mail along with the completed survey.

1) Coverage area (extent)	
2) Update frequency	
3) Data collection method	
4) Scale	
5) Attributes captured	
6) Type of structures captured	

7) From the list below, please indicate the statement(s) that best describe your organization's structures data. If your organization maintains multiple datasets please indicate all that apply.
   a) A point is placed at the structure location.
   b) A point is placed at the facility site location. (There may be multiple structures per point for large institution such as Universities)
   c) A footprint polygon is created at the structure location.

Data Policy Questions

1. Are there currently data access rules that prevent your jurisdiction from contributing to the public domain via *The National Map*?
2. Is your organization currently charging to recover the cost of collecting and maintaining structures data?
3. Is your organization currently collecting spatial data or attributes for your structures database that you are not willing to make publically available?
4. Is your organization currently creating metadata? If yes, is your organization posting it to the Geospatial One Stop (GOS) or the Maryland Mapping Resource Guide (MMRG)?
5. Please indicate any concerns you have about sharing your structures data below.

APPENDIX B Data Dictionary

Page 1

Field Name	Attribute Definition	Data Typ	Defult Value	Domain	Nulls	Precis	Scal	Leng	Definition Source
Permanent_Identifier	Unique permanent identifier assigned to every feature (40-character GUID format)	String			Yes			40	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Source_FeatureID	Unique permanent identifier that may exist in the source data.	String			Yes			40	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Source_DatasetID	Unique permanent identifier assigned to each source dataset, ties feature to its metadata (40-character GUID)	String			Yes			40	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Source_DataDesc	Free-form textual description of the source dataset where the feature came from	String			Yes			100	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Source_Originator	Name of organization that developed the data set.	String			Yes			130	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Data_Security	Security level of the data (i.e.: Top Secret, Sensitive, etc.).	Short Integer	0	Sec_Classification Domain	Yes	0			The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Distribution_Policy	Homeland Security code identifying distribution level of data (one alpha, one digit, code).	String	E4	Distribution_Policy Domain	Yes			4	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
LoadDate	Date of load or last modification of feature	Date			Yes	0	0	8	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
FType	3-digit integer value identifying feature type	Long Integer	730		Yes	0			The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
FCode	5-digit integer value identifying feature type category. (First three digits of the FCode are the same as the Ftype)	Long Integer	73002		Yes	0			The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Name	Proper name by which a feature is known	String			Yes			60	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)

Data Dictionary

Page 2

Field Name	Attribute Definition	Data Type	Defult Value	Domain	Nulls	Precision	Scale	Length	Definition Source
Name	Proper name by which a feature is known	String			Yes			60	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
PointLocation Type	Coordinate locations of a feature are plotted to the centroid of the structure. In rare cases, if after exhausting all means of verification including phone calls to the entity, you are still unable to determine the building location, select 'Approximate' or 'Unknown' from the pull down list.	Long Integer	1	Point Location Type Domain (Unknown = 0, Centroid = 1, Egress or Entrance = 2 Turn-off location = 3, Approximate = 4)	Yes	0			The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
AdminType	Identifies level of ownership. In the course of data collection and verification, if this information becomes readily available, please select the appropriate administrative type from the pull down list, else leave as 'Unknown'.	Long Integer		Owner class domain (Unknown = 0, Federal = 1, Tribal = 2, State = 3, County = 4 Municipal = 6, Private = 7,)	Yes	0			The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Address Building Name	Name of the building	String			Yes			60	
Address	Building Address	String			Yes			75	
City	City the building is located in	String			Yes			40	
State	The State the building is located in	String			Yes			2	
Zipcode	The zipcode the building is located in	String			Yes			10	
GNIS_ID	Unique identifier assigned by GNIS	String			Yes			10	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Foot_PermID	Unique permanent identifier (40-character GUID) which crossreferences the polygon footprint feature to the point feature it represents in the StructPoint feature class.	String			Yes			40	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
Complex_PermID	Unique permanent identifier (40-character GUID) which crossreferences the polygon complex feature to the point feature it represents in the StructPoint feature class.	String			Yes			40	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)
DDD_PermID – (DDD = 3 dimensional)	Unique permanent identifier (40-character GUID) which crossreferences the polygon DDD feature to the point feature it represents in the StructPoint feature class.	String			Yes			40	The National Structures Dataset (NSD) Editorial Guidelines September 12, 2008 (revised 09252008)

APPENDIX C

Data Processing Method

Background

A total of five jurisdictions participated in the project. An agreement to share structures data was reached with four jurisdictions: Baltimore County, Frederick County, and Montgomery County in Maryland, and the District of Columbia. Baltimore County, Maryland volunteered to serve as the pilot jurisdiction to help develop a process for transforming the data into the proper format to conform to the data model.

Early in the pilot phase, it was determined that the following four steps must be completed to properly conform the data to the data model.

Step 1
Develop a better understanding of the USGS Best Practices data model attribute definitions.

Action
We consulted with Paul Wiese, our USGS point of contact for this project, to gain a better understanding of the USGS Best Practices data model attribute definitions. Based on the documentation provided by the USGS the project team developed a data dictionary. The data dictionary contains Field Name, Attribute Definition, Data Type, Default Value, Domain, Nulls, Precision, Scale, Length, and Definition Source. The data dictionary was used to properly map jurisdiction's attributes to the appropriate fields in the final dataset.

Step 2
Determine which fields in each jurisdiction's dataset would best translate to the f-code.

Action
The f-code value is a 5-digit integer value identifying feature type category. The value had to be determined for each feature in each jurisdiction's dataset using existing attributes. A different process was used for each jurisdiction to extract this information from the data and is documented below.

Step 3
Answer several county-specific questions about the data to properly attribute the data.

Action

To complete the transformation process, each jurisdiction had to answer a few additional questions about their data, data sharing, and security policies and also provide a brief 100-character description of the dataset. A document was sent to each jurisdiction requesting this information. The jurisdiction would then complete and returned the document. The information was used to populate the appropriate fields in the final dataset.

Step 4
Develop a repeatable process to transform the jurisdiction's data into the data model.

Action
The repeatable process is intended to simplify later updates and allow the data to easily migrate into the data model. It was determined that the ESRI Spatial ETL tool (Extract Transform and Load) would work best for this process. This tool allows the user to build a model to manipulate and transform data to a new output data source. Once the model is built, it can be used repeatedly until data schemas change for the input or output datasets.

Once these steps were taken, the project team was able to move forward with data collection and transformation with each jurisdiction individually. The data from each jurisdiction differed in many ways, which led to unique processing steps for each dataset in the final process. The project team built a model for each jurisdiction that can be used in the future to create an updated dataset. The final output from each jurisdiction's model was exported directly into the Structures database and manually imported into the appropriate feature class.

The process employed and the final ETL models created for each jurisdiction are described on the following pages.

Baltimore County, Maryland

Baltimore County, Maryland served as the pilot jurisdiction for this project. The first step was to determine how to extract information needed to calculate the f-code for Baltimore County. Baltimore County maintains a field with codes from the North American Industry Classification System (NAICS codes). An f-code was assigned to each NAICS code within the Baltimore County dataset. This was then sent to Baltimore County for approval.

The new f-code was calculated for each NAICS code. An error was noted during the quality control check. Some features with NULL NAICS code values were assumed to be residential; however, they belonged to another category. The solution was a two-tiered approach. First, a new NAICS code value for the NULL NAICS code was calculated using the ADDRESSUSE field. Next, the appropriate f-code value was assigned to the new, updated NAICS code field.

A model was built to include this process using the ETL tool. The final model completed the process needed to transform the Baltimore County data into the appropriate format for the data model. The tools used for this process are described as follows.

1. ValueMapper – calculates a temporary NAICS code based on the ADDRESSUSE field on Baltimore County Data.
    a. Input Field – ADDRESSUSE
    b. Output Field – CodeForNull
2. NullAttributeReplacer – Replaces null values in the Code field with the NAICS code calculated from the address use field.
    a. Input Field – CodeForNull
    b. Output Field – CODE
3. ValueMapper – calculates the f-code based on the NAICS code.
    a. Input Field – CODE
    b. Output Field – NAICStoFCodeValue
4. Timestamper – Adds a field with the year, month, and day the data was added to the database.
5. ATTRIBUTECREATOR – Add fields and calculates consistent values for the fields.
    a. Field 1:
        i. New field Created: PntLocation
        ii. Consistent value calculated: 1
        iii. Value Description: Centroid
    b. Field 2:
        i. New field Created: AdminType
        ii. Consistent value calculated: 5

iii. Value Description: County Maintained

c. Field 3:
   i. New field Created: SourceOrgin
   ii. Consistent value calculated: Baltimore County, Maryland
   iii. Value Description: Data source Name
d. Field 4:
   i. New field Created: DataDescription
   ii. Consistent value calculated: The County's Facility geodatabase captures locations, addresses and names to be used for public safety and other mapping purposes.
   iii. Value Description: A description of the data from the county in 100 characters or less.
e. Field 5:
   i. New field Created: DataSecure
   ii. Consistent value calculated: 5
   iii. Value Description: unclassified
f. Field 6:
   i. New field Created: DisPolicy
   ii. Consistent value calculated: E1
   iii. Value Description: Public Domain - Free Distribution to Third Parties Via Internet
g. Field 7:
   i. New field Created: SourceDatasetID
   ii. Consistent value calculated: f8c2b226-cc21-485a-bb0d-3695e8ec4f50
   iii. Value Description: The Unique ID for the Baltimore County Dataset

6. ATTRIBUTECOPIER – creates a temp field and copies the f-code value.
   a. Input Field – NAICStoFCodeValue
   b. Output Field – FTypeTemp
7. SUBSTRINGEXTRACTOR - Extracts a substring from the source. This tool extracts the first three characters from the, Temp f-code field created above.
   a. Input Field – FTypeTemp
   b. Output Field – FType
8. UUIDGenerator - Calculates a UUID (Universally Unique IDentifier) for each incoming feature. This will be the final ID in the dataset.
   a. Output Field – _UUID
9. GOIDGenerator - Calculates a GOID (Geographic Object IDentifier) for each incoming feature, and adds it as a new attribute. This calculates the local ID field in the dataset where one is not already available.

a. Output Field – _GOID

10. ATTRIBUTERENAMER – Renames the attributes from the source dataset to match the attribute names of the final dataset.

Baltimore County Attribute	Final Dataset Attribute
ADDRESS_ID	Source_FeatureID
NAME	Name
NAME	AddressBuildingName
ADDRLABEL	Address
COUNTY_PRE	City
STATE	State
ZIP	Zipcode
NAICSroFCodeValue	Fcode
_goid	Permanent_Identifier
_timestammp	LoadDate
SourceDatasetID	Source_DatasetID
SourceOrgin	Source_Originator
PNTLocation	PointLocation_Type
AdminType	AdminType
DataDescription	Source_DataDesc
Datasecurity	Data_security
FTYPE	Ftype
DisPolicy	Distribution_Policy

Frederick County, Maryland

The second jurisdiction to contribute data was Frederick County, Maryland. The data from Frederick County was maintained in 13 datasets.

- County Lease Buildings
- County owned Buildings
- Fire Stations
- Golf Courses
- Hospitals
- Hotels
- Ice Rinks
- Library
- Places of worship
- Police Station
- Post Office
- Public Child Care
- Shopping Centers

Each of the 13 datasets was based on the categories of data; therefore, we were able to assign the appropriate f-code value to the entire datasets. The tools used to transform Frederick County, Maryland data into the appropriate format for the data model are described as follows.

1. Timestamper – Adds a field with the year, month, and day the data was added to the database.
2. ATTRIBUTECREATOR – Add fields and calculates consistent values for the fields
    a. Field 1:
        i. New field Created: PntLocation
        ii. Consistent value calculated: 1
        iii. Value Description: Centroid
    b. Field 2:
        i. New field Created: AdminType
        ii. Consistent value calculated: 5
        iii. Value Description: County Maintained
    c. Field 3:
        i. New field Created: SourceOrgin
        ii. Consistent value calculated: Frederick County, Maryland
        iii. Value Description: Data source Name

d. Field 4:
    i. New field Created: DataDescription
    ii. Consistent value calculated: The County's Facility geodatabase captures locations, addresses and names to be used for public safety and other mapping purposes.
    iii. Value Description: A description of the data from the county in 100 characters or less.
e. Field 5:
    i. New field Created: DataSecure
    ii. Consistent value calculated: 5
    iii. Value Description: unclassified
f. Field 6:
    i. New field Created: DisPolicy
    ii. Consistent value calculated: E4
    iii. Value Description: Public Domain - Free Distribution to Third Parties Via Internet
g. Field 7:
    i. New field Created: F-Code
    ii. Consistent value calculated: Unique for each dataset. New field Created: (See f-code appendix).
    iii. Value Description: The appropriate f-code for the input dataset. Matches the USGS Structures Data model.
h. Field 8:
    i. New field Created: SourceDatasetID
    ii. Consistent value calculated: e0f13696-59d3-4b0a-801f-df383028c507
    iii. Value Description: The Unique ID for the Frederick County Dataset

3. ATTRIBUTECOPIER – creates a temp field and copies the f-code value.
    i. Input Field – NAICStoFCodeValue
    ii. Output Field – FTypeTemp
4. SUBSTRINGEXTRACTOR - Extracts a substring from the source. This tool extracts the first three characters from the, Temp f-code field created above.
    i. Input Field – FTypeTemp
    ii. Output Field – FType
5. UUIDGenerator - Calculates a UUID (Universally Unique IDentifier) for each incoming feature. This will be the final ID in the dataset.
    i. Output Field – _UUID (Only used where global ID did not already exist in the dataset).

6. GOIDGenerator - Calculates a GOID (Geographic Object IDentifier) for each incoming feature, and adds it as a new attribute. This calculates the local ID field in the dataset where one is not already available.
    i. Output Field – _GOID
7. ATTRIBUTERENAMER – Renames the attributes from the source dataset to match the attribute names of the final dataset.

Frederick County Attribute	Final Dataset Attribute
AdminType	AdminType
DataDescription	Source_DataDesc
Datasecurity	Data_security
Fcode	Fcode
FTYPE	Ftype
AdminType	AdminType
DisPolicy	Distribution_Policy
GEO_ADDRES	Address
GEO_ZIP	Zipcode
PNTLocation	PointLocation_Type
_goid	Permanent_Identifier
GlobalID	Source_FeatureID
Source Orgin	Source_Originator
_timestamp	LoadDate
GEO_ADDRES	Name
GEO_ADDRES	AddressBuildingName
SourceDatasetID	Source_DatasetID

District of Columbia

The third jurisdiction we received data from was the District of Columbia. Compared to the other jurisdictions, the District of Columbia presented the greatest difficulty with determining f-codes for individual features. After we reviewed the data, we determined that the ALIAS_ID field would give us the most accurate description; however, this field did not have large category groupings but was more of a descriptor of the feature. While address alias points are the most representative dataset for landmarks in the District Columbia, the District does not maintain feature codes. Throughout the project, the project team tried to generate codes based on other attributes captured by the original dataset; however, the District of Columbia decided to keep the attribution general to minimize the chance of error. A summary of the ALIAS_ID field was extracted from the data, and the appropriate f-code value was assigned to each of the 3,944 structure points. This table was then joined back to the original data and the ETL tool was created. With this method, the update process for the jurisdiction may require some manual review of the data and an update to the table. The table would then have to be rejoined to the dataset before running the ETL tool.

The tools used to transform District of Columbia data into the appropriate format for the data model are described below.

1. Timestamper – Adds a field with the year, month, and day the data was added to the database.
2. ATTRIBUTECREATOR – Add fields and calculates consistent values for the fields
    a. Field 1:
        i. New field Created: PntLocation
        ii. Consistent value calculated: 1
        iii. Value Description: Centroid
    b. Field 2:
        i. New field Created: AdminType
        ii. Consistent value calculated: 5
        iii. Value Description: County Maintained
    c. Field 3:
        i. New field Created: SourceOrgin
        ii. Consistent value calculated: Washington DC
        iii. Value Description: Data source Name
    d. Field 4:
        i. New field Created: DataDescription

ii. Consistent value calculated: The County's Facility geodatabase captures locations, addresses and names to be used for public safety and other mapping purposes.
iii. Value Description: A description of the data from the county in 100 characters or less

e. Field 5:
i. New field Created: DataSecure
ii. Consistent value calculated: 5
iii. Value Description: unclassified

f. Field 6:
i. New field Created: DisPolicy
ii. Consistent value calculated: E4
iii. Value Description: Public Domain - Free Distribution to Third Parties Via Internet

g. Field 7:
i. New field Created: SourceDatasetID
ii. Consistent value calculated: 4201acf0-79ff-4302-9de0-1a52cc68f209
iii. Value Description: The Unique ID for the DC Dataset

h. Field 8:
i. New field Created: FCode
ii. Consistent value calculated: 79000
iii. Value Description: Building General

3. SUBSTRINGEXTRACTOR - Extracts a substring from the source. This tool extracts the first three characters from the, Temp f-code field created above.
a. Input Field – FTypeTemp
b. Output Field – FType
4. GOIDGenerator - Calculates a GOID (Geographic Object IDentifier) for each incoming feature, and adds it as a new attribute. This calculates the local ID field in the dataset where one is not already available.
5. Output Field – _GOID
6. ATTRIBUTERENAMER – Renames the attributes from the source dataset to match the attribute names of the final dataset.

DC County Attribute	Final Dataset Attribute
ADDRESSALIAS_ID	Source_FeatureID
ALIASNAME	Name
ALIASNAME	AddressBuildingName
FULLADDRESS	AddressBuildingName
CITY	City
STATE	State
ZIPCODE	Zipcode
Ffcode	Fcode
_goid	Perminant_Identifier
_timestamp	LoadDate
SourceDatasetID	Source_DatasetID
SourceOrgin	Source_Originator
PNTLocation	PointLocationType
Admintype	AdminType
DataDescription	Source_DataDesc
Datasecurity	Data_Security
FTYPEnew	Ftype
DisPolicy	Distribution+Policy

Montgomery County, Maryland

The final jurisdiction to contribute data was Montgomery County, Maryland. The data from Montgomery County was maintained in 34 datasets.

- Airfield
- Airtest
- Bus_prk
- Cemetery
- Col_univ
- Elem_sch
- Fire_fed
- Fire_sta
- Golf
- High_sch
- Hospital
- Library
- Liquir
- Lodging
- Marc
- Mogov
- Metro
- Midd_sch
- Park_fac
- Parking
- Place_name
- Poi
- Police
- Pools
- Postoffice
- Priv_sch
- Prkrid
- Rec_ctr
- Shop_ctr
- Rsc
- Spec_sch
- Sr_ctr
- Worship
- Ymca

Similar to Fredrick County, the 34 datasets were categories of data; therefore, we were able to assign the appropriate f-code value to the entire datasets. This process was completed in the Montgomery County, Maryland model. The tools used to transform Montgomery County data into the proper format for the data model are described below.

1. Timestamper – Adds a field with the year, month, and day the data was added to the database.
2. ATTRIBUTECREATOR – Add fields and calculates consistent values for the fields
    a. Field 1:
        i. New field Created: PntLocation
        ii. Consistent value calculated: 1
        iii. Value Description: Centroid
    b. Field 2:
        i. New field Created: AdminType
        ii. Consistent value calculated: 5
        iii. Value Description: County Maintained
    c. Field 3:
        i. New field Created: SourceOrgin
        ii. Consistent value calculated: Montgomery County, Maryland
        iii. Value Description: Data source Name
    d. Field 4:
        i. New field Created: DataDescription
        ii. Consistent value calculated: The County's Facility geodatabase captures locations, addresses and names to be used for public safety and other mapping purposes.
        iii. Value Description: A description of the data from the county in 100 characters or less.
    e. Field 5:
        i. New field Created: DataSecure
        ii. Consistent value calculated: 5
        iii. Value Description: unclassified
    f. Field 6:
        i. New field Created: DisPolicy
        ii. Consistent value calculated: E4
        iii. Value Description: : Public Domain - Free Distribution to Third Parties Via Internet
    g. Field 7:
        i. New field Created: F-Code

   ii. Consistent value calculated: Unique for each dataset. New field Created: (See f-code appendix)
   iii. Value Description: The appropriate f-code for the input dataset. Matches the USGS Structures Data model.
h. Field 8:
   i. New field Created: SourceDatasetID
   ii. Consistent value calculated: 8189c579-afe3-4c08-89af-c8536b519095
   iii. Value Description: The Unique ID for the Montgomery County Dataset

3. ATTRIBUTECOPIER – creates a temp field and copies the f-code value.
   a. Input Field – NAICStoFCodeValue
   b. Output Field – FTypeTemp
4. SUBSTRINGEXTRACTOR - Extracts a substring from the source. This tool extracts the first three characters from the, Temp f-code field created above.
   a. Input Field – FTypeTemp
   b. Output Field – FType
5. UUIDGenerator - Calculates a UUID (Universally Unique IDentifier) for each incoming feature. This will be the final ID in the dataset.
   a. Output Field – _UUID
6. GOIDGenerator - Calculates a GOID (Geographic Object IDentifier) for each incoming feature, and adds it as a new attribute. This calculates the local ID field in the dataset where one is not already available.
   a. Output Field – _GOID
7. ATTRIBUTERENAMER – Renames the attributes from the source dataset to match the attribute names of the final dataset.

Montgomery County Attribute	Final Dataset Attribute
AdminType	AdminType
DataDescription	Source_DataDesc
Datasecurity	Data_security
Fcode	Fcode
FTYPE	Ftype
DisPolicy	Distribution_Policy
ADDRESS	Address
ZIPCODE	Zipcode
PNTLocation	PointLocation_Type
_goid	Permanent_Identifier
_UUID	Source_FeatureID
Source Orgin	Source_Originator
_timestamp	LoadDate
NAME	Name
NAME	AddressBuildingName
SourceDatasetID	Source_DatasetID