Structures Data Collection for *The National Map* using Volunteered Geographic Information

By Barbara S. Poore, Eric B. Wolf, Erin M. Korris, Jennifer L. Walter, Greg D. Matthews

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia 2012

Contents

Figures

Tables

Executive Summary

The U.S. Geological Survey (USGS) has historically sponsored volunteered data collection projects to enhance its topographic paper and digital map products. This report describes one phase of an ongoing project to encourage volunteers to contribute data to *The National Map* using online editing tools. The USGS recruited students studying geographic information systems (GIS) at the University of Colorado Denver and the University of Denver in the spring of 2011 to add data on structures—manmade features such as schools, hospitals, and libraries— to four quadrangles covering metropolitan Denver. The USGS customized a version of the online Potlatch editor created by the OpenStreetMap project and populated it with 30 structure types drawn from the Geographic Names Information System (GNIS), a USGS database of geographic features. The students corrected the location and attributes of these points and added information on structures that were missing. There were two rounds of quality control. Student volunteers reviewed each point, and an in-house review of each point by the USGS followed.

Nine-hundred and thirty-eight structure points were initially downloaded from the USGS database. Editing and quality control resulted in 1,214 structure points that were subsequently added to *The National Map.* A post-project analysis of the data shows that after student edit and peer review, 92 percent of the points contributed by volunteers met National Map Accuracy Standards for horizontal accuracy. Lessons from this project will be applied to later phases. These include:

- simplifying editing tasks and the user interfaces,
- stressing to volunteers the importance of adding structures that are missing, and
- emphasizing the importance of conforming to editorial guidelines for formatting names and addresses of structures.

The next phase of the project will encompass the entire State of Colorado and will allow any citizen to contribute structures data. Volunteers will benefit from this project by engaging with their local geography and contributing to a national resource of topographic information that remains in the public domain for anyone to download.

Background

The U.S. Geological Survey (USGS) has historically sponsored volunteer data collection projects to enhance its topographic paper and digital map products, but these activities were suspended in 2006 due to budget concerns. Since then, new Internet technologies have made it easy for citizens to georeference many different types of information and share this information via online mapping platforms and social networking sites. These data have been referred to as volunteered geographic information (VGI).

As a result of these developments, the USGS launched a pilot program to explore the feasibility of reinstating a volunteer data-collection program for *The National Map* (*http://nationalmap.gov*). To gather best practices from mature VGI and citizen science projects, the USGS sponsored a workshop on VGI in January 2010. More information about the workshop can be found at *http://cegis.usgs.gov/vgi/*.

Later that year, the USGS organized the OpenStreetMap Collaborative Prototype (OSMCP). The OSMCP evaluated whether an existing Web-based mapping system created by the OpenStreetMap (OSM) (*http://www.openstreetmap.org*) community could be modified to satisfy USGS data collection needs. OSM is a global, open source, open access database of

geographic data produced entirely by volunteers. The OSMCP did not incorporate volunteer contributions but focused on whether OSM's software could support simultaneous collaborative editing between the USGS and a partner agency. The USGS customized a version of the OSM's Potlatch map editor that was accessible over the Internet in a standard Web browser. The Potlatch editor was hosted on a server in the U.S. Government domain (.gov). Transportation data developed by the Kansas Data Access and Support Center (DASC) were loaded into the online system and successfully edited by both the USGS and the DASC, producing data that met or exceeded requirements for integration into *The National Map*. USGS Open-File Report 2011-1136 (Wolf and others, 2011) provides additional detail on this project.

Introduction to Structures—VGI

Structures—VGI was the second phase of OSMCP, launched in the spring of 2011. The goals of the Structures—VGI project were:

- Engage volunteers to collect data for incorporation into *The National Map*.
- Develop a process to ensure that volunteered data meet or exceed quality standards of authoritative data.
- Test the ability of volunteers to effectively perform quality control on data.
- Create a workflow for VGI data collection that could be scaled up for future operations.
- Understand the quantity and quality of data collected by volunteers.
- Examine and document the potential costs and benefits of VGI to the USGS.
- Provide completed structures data to *The National Map* and contribute the structures back to the OpenStreetMap community.
- Create best practice guidelines for VGI outreach and education.
- Publish a USGS Open-File Report for the project.

To test whether volunteers could successfully collect data for use in *The National Map*, it was necessary to make it easier to edit data, engage non-professional volunteer contributors, and improve the user experience of the OSM editing software, Potlatch.

To make editing easier, the USGS shifted the target of data collection from the complex data theme of transportation to the simpler theme of structures. Structures data—point features and their locations and other attributes—represent manmade buildings such as fire stations, houses of worship, schools, post offices, and so on. These data comprise the National Structures Dataset (NSD)—one of the eight data layers of *The National Map* (*http://nationalmap.gov/structures.html*). The USGS prioritizes structures data collection for the needs of the disaster planning and response communities and for the new generation of digital topographic maps, US Topo (*http://nationalmap.gov/ustopo/*). The USGS considers a set of 30 structure types, described in detail below, as priorities for data collection by contractors and partners. Structures—VGI focused on these 30 structure types. The geographic scope of the project was limited to the four quadrangles that cover the greater Denver metropolitan area: Arvada, Commerce City, Fort Logan and Englewood (fig. 1).

The USGS engaged students enrolled in GIS courses at the University of Colorado Denver and the University of Denver who were less experienced in using GIS than the State employees from Kansas. Technical modifications to the Potlatch interface were made to simplify

the editing tasks, and guidelines to lead the volunteers through the editing process were developed.
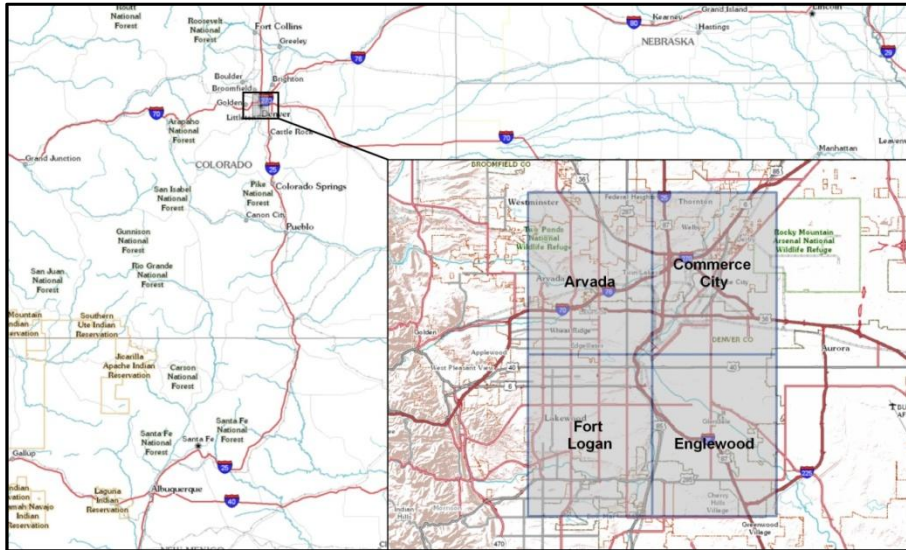
Figure 1. Four primary USGS quadrangles for the Denver metropolitan area used in Structures—VGI.

Technologies: Software and User Interface

For Structures—VGI, the OSM Potlatch editing interface was upgraded to the most recent version, Potlatch 2, (*http://wiki.openstreetmap.org/w/index.php?title=Potlatch_2&oldid=688013*). This change allowed the USGS to use Extensible Markup Language (XML) style sheets to customize the interface, incorporating USGS structure types, symbols, and standard tags and a USGS-branded custom map display and background images drawn from *The National Map*.

Project Web Site

A Google Sites Web site (fig. 2) served as the main portal through which students accessed all necessary documents and tools. These included:

- a USGS-branded local instance of OSM
- a project description
- student assignments
- a time-tracking spreadsheet
- the Potlatch 2 editor
- editing and quality control guidelines and tips
- a message board that allowed students to interact with each other, ask questions of USGS personnel, and provide feedback to USGS on potential improvements.

Figure 2. Google Sites Web site for Structures—VGI project.

Potlatch Editor

The USGS configured the Potlatch 2 editor (fig. 3) so that the resulting data would conform to USGS specifications. For editing in Structures—VGI, users could select among background layers that provided different types of maps and levels of detail (table 1). The Denver Reference Map is a general map of the Denver area on which the boundaries of the four USGS quadrangles are superimposed. This map was used in training materials and to show the status of contributed points through symbols and color coding. Four background layers were based on Web Mapping Services (WMS) provided by *The National Map:* aerial imagery (orthophotos), scanned topographic sheets (digital raster graphic, or DRG), and a base map with selected layers from *The National Map* at two different scales or zoom levels (fig. 4).
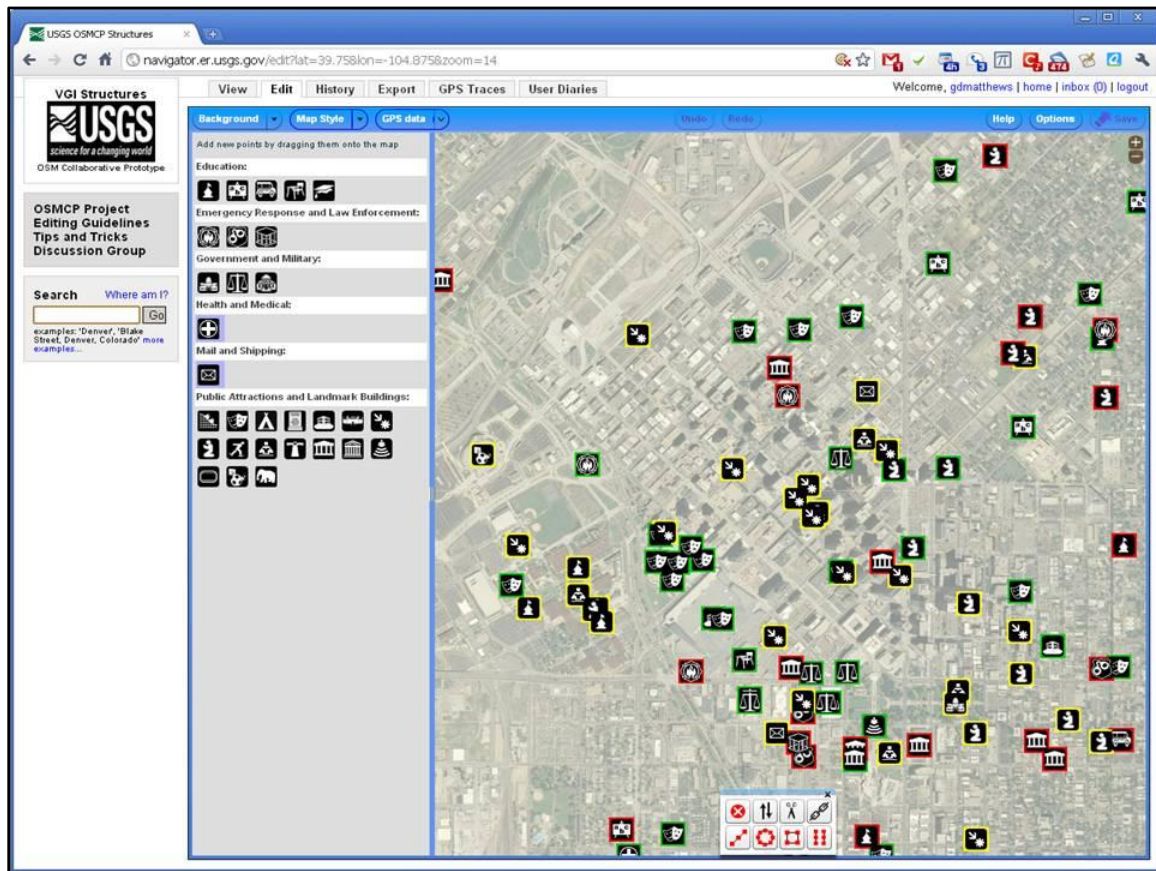
Figure 3.   Editing interface for Structures—VGI.

Table 1.   Sources for background layers in Structures—VGI.

Layer in Potlach Menu	Source
Denver Reference Map	Georeferenced TIFF installed as a Web Map Service (WMS) on the local project server
Aerial Imagery	Orthoimagery derived from the National Agricultural Imagery Program and other sources *http://nationalmap.gov/ortho.html*
TNM Large Scale	*The National Map* as a vector base map provisioned as Google-style cached tileset from *http://raster.nationalmap.gov/ArcGIS/rest/services/TNM_Large_Scale_Imagery/MapServer*
TNM Small Scale	*The National Map* as a vector base map provisioned as Google-style cached tileset from global scale to approximately 1:289,000-scale *http://raster1.nationalmap.gov/ArcGIS/rest/services/TNM_Small_Scale_Imagery/MapServer*
Classic Topos (this name chosen in interface as DRG is not well known outside technical community)	Scannned topographic map sheets (Digital raster graphic or DRG) provisioned as Google-style tileset cached on local server *http://raster.nationalmap.gov/ArcGIS/rest/services/DRG/TNM_Digital_Raster_Graphics/MapServer*
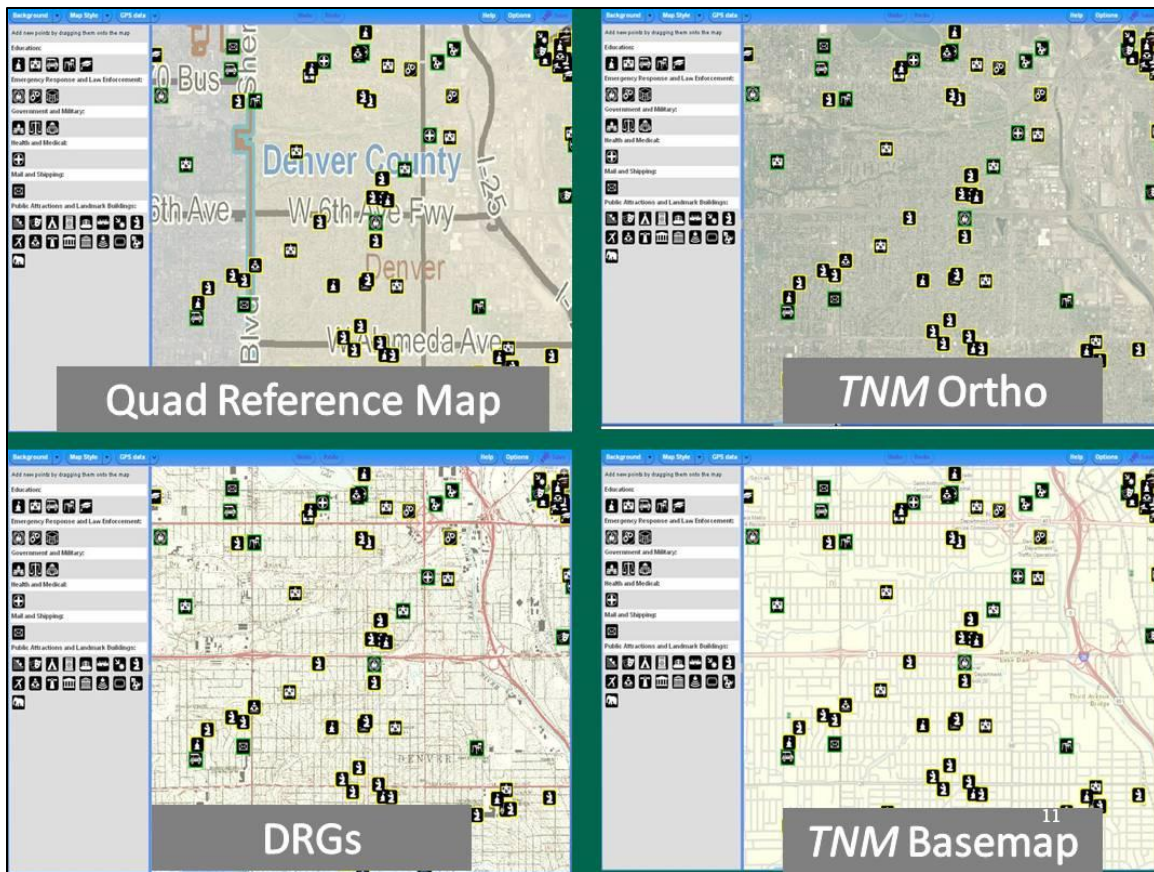
Figure 4.   Background layers in editing mode for Structures—VGI.

Data

The USGS has represented manmade and natural structures on topographic maps for 125 years. Structures were initially collected by field personnel as part of the mapping process. Later, aerial photography was used to collect information on structures for map revisions. The U.S. Board on Geographic Names (*http://geonames.usgs.gov*), whose mission is to standardize geographic names across the Federal Government, drew on structures depicted on USGS topographic maps, visitor maps from the U.S. Forest Service, and charts produced by the National Oceanic and Atmospheric Agency (NOAA) to create the automated Geographic Names Information System (GNIS, or GAZ as it is referred to within the USGS) (*http://geonames.usgs.gov/pls/gnispublic*). GNIS (GAZ) includes data on manmade structures (such as schools and fire stations) and natural features of the earth (such as volcanoes and rivers) (Orth and Payne, 1987).

In the late 1990s the USGS designed a data model for the National Structures Dataset (NSD) to formalize specifications for data on manmade structures for *The National Map*. Partnerships and contracts have been used to update the NSD from various sources:

- GNIS (GAZ)
- Critical facilities data from the Federal Emergency Management Agency (FEMA)
- Critical structures data from the Department of Homeland Security and the National Geospatial-Intelligence Agency

- State partnerships
- U.S. Department of Education.

The National Structures Dataset (NSD)

Structures are represented in the NSD by points (fig. 5). A point is collected for each unique function of a structure. Thus, in a structure that houses both a police and fire station, a point would be collected for each. The point is located by latitude and longitude, derived either from a GPS reading or by using GIS software to find the position of the structure on the digital orthophoto derived from National Agricultural Image Program (NAIP) imagery (USGS, 2012). Other characteristics of structures such as function, name, and address are stored in the database as attributes. Mandatory attributes of NSD data are shown in table 2.

Figure 5. Denver Museum of Nature and Science as represented in the National Structures Dataset; green dot shows the point, yellow indicates local roads. The source of the image is the National Agricultural Image Program (NAIP) from 2009.

Table 2. Mandatory attributes of the point feature Denver Museum of Nature and Science from the National Structures Dataset. A complete explanation follows.

Name	Point Location	FType	FCode	Address	Address Building Name	City	State	ZIP Code
Denver Museum of Nature and Science	Centroid	820	82032	2001 North Colorado Boulevard	(Mandatory if Applicable)	Denver	CO	80205

This table is drawn from in the Draft Guidelines for Contributing Structures Data to *The National Map* (USGS, 2012).

Name: Names of structures should be derived from official sources such as a city Web site for fire and police stations or the U.S. Department of Education for schools. For the Denver Museum of Nature and Science example in figure 5 and table 2, the authoritative Web sites would be the City of Denver or the museum Web site. Names and addresses of structures data must be formatted according to editorial conventions derived from the U.S. Board of Geographic Names (U.S. Board of Geographic Names, 1997) (see appendix A). For example, abbreviations are not permitted in the name field. In the quality control stages of Structures—VGI, the USGS evaluated the contributions from student editors to see how well they adhered to these editorial conventions.

Point Location: Points can be collected either at the center of the structure (centroid) which is the preferred location, the entrance or exit, the turn-off location or cross street (turn-off location), or at an approximate location in relation to the actual structure (approximate).

FType and **FCode**: Structures are assigned a feature type and a feature code—a subset of the feature type. In the case of the Denver Museum of Nature and Science, 820 is the feature type for public attractions and landmark buildings, and 82032 is the feature code for museums. A full diagram of the Best Practices Data Model for structures data and a list of all feature codes and feature types can be found as a PDF file on the *National Map* Web site at *http://services.nationalmap.gov/bestpractices/model/acrodocs/Poster_BPStructures_03_01_2006.pdf*.

Address: Street, route, or highway address representing the location of the feature. Addresses must be derived from official sources and formatted according to specific editorial conventions. Current USGS guidelines for addressing are based on the U.S. Postal Service's Postal Addressing Standard (U.S. Postal Service, 2010) (see appendix A).

Address Building Name: This attribute is mandatory only if applicable and refers to an instance in which the building has an official name such as "Murray Building."

City: The official name of the community where the structure is located.

State: The state where the structure is located expressed as the two character U.S. Postal Service abbreviation.

Zip Code: The five digit U.S. Postal Service ZIP Code associated with the street address for the physical location of the structure.

Workflow for Incorporating Structures Data into *The National Map*

The principal USGS databases for structures data, GNIS (GAZ) and NSD, are separately maintained. Contractors are used to update the GNIS (GAZ) database (left side, fig. 6). The contractors review and prepare the data. They add new features and clean up existing features. The USGS samples a subset (typically 10 percent) of the points new to GNIS (GAZ) for quality control of geometry and attributes. All new data (100 percent) are checked for spelling and punctuation according to the editorial guidelines (see sample addressing guidelines in appendix A). Metadata are updated, GNIS IDs are assigned to new features, and the data are checked into the NSD, which simultaneously updates the GNIS (GAZ).
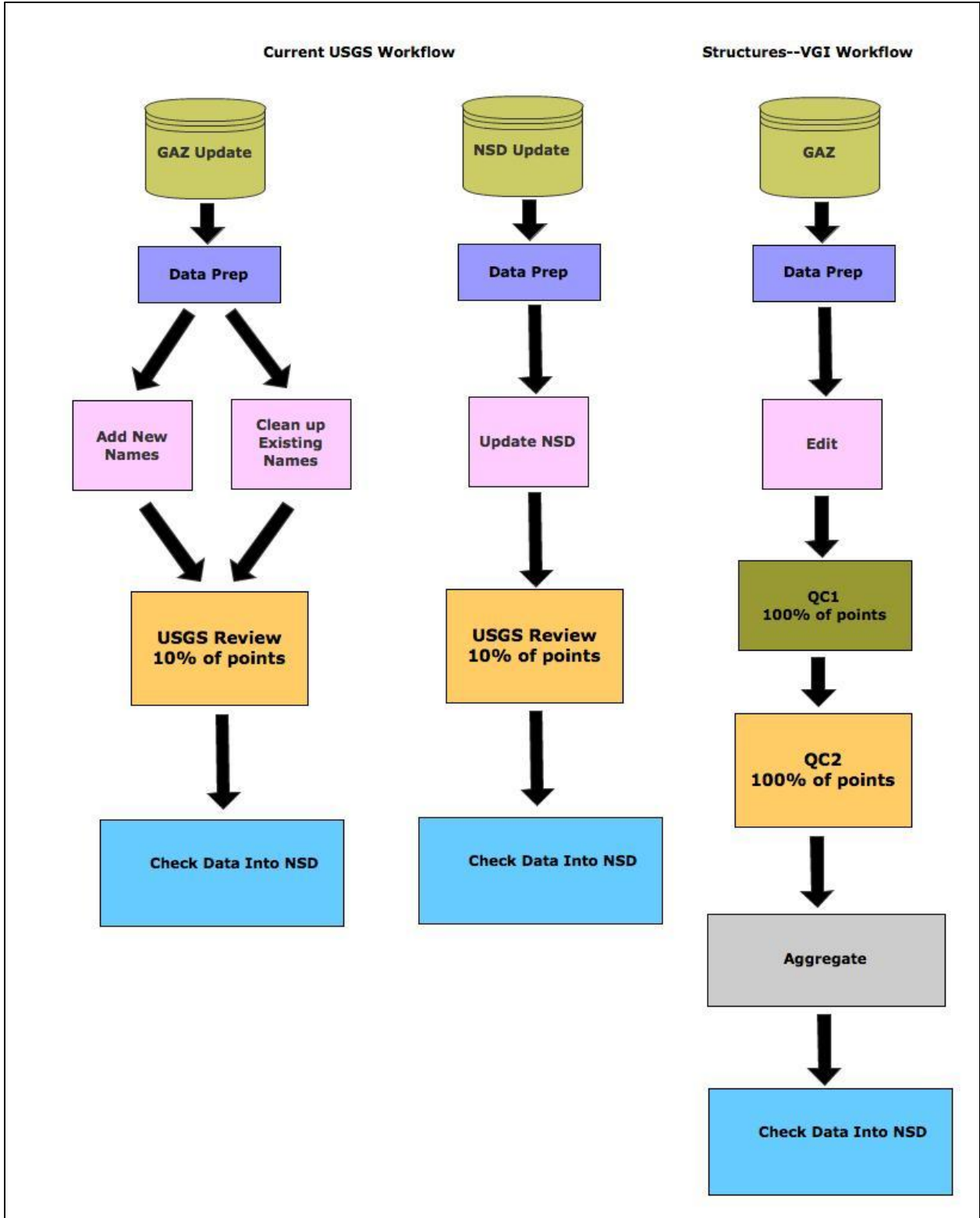
Figure 6. Overview of workflow for incorporating structures data into *The National Map.*

Updates to the NSD can be made by either contractors or partners from State and local governments (middle, fig. 6). The data are converted into NSD geodatabase format and compared with existing structures in the NSD. New points are added, and the attributes of existing points are reviewed and updated. Feature locations are verified, old points identified for deletion, and existing GAZ IDs are assigned to features, if applicable. The USGS conducts a quality review, which includes spot-checking a sample of the data (typically 10 percent) for geometry and attributes. All data (100 percent) are checked for conformance to editorial guidelines for names and addresses (see sample addressing guidelines in appendix A).

Workflow for Structures—VGI
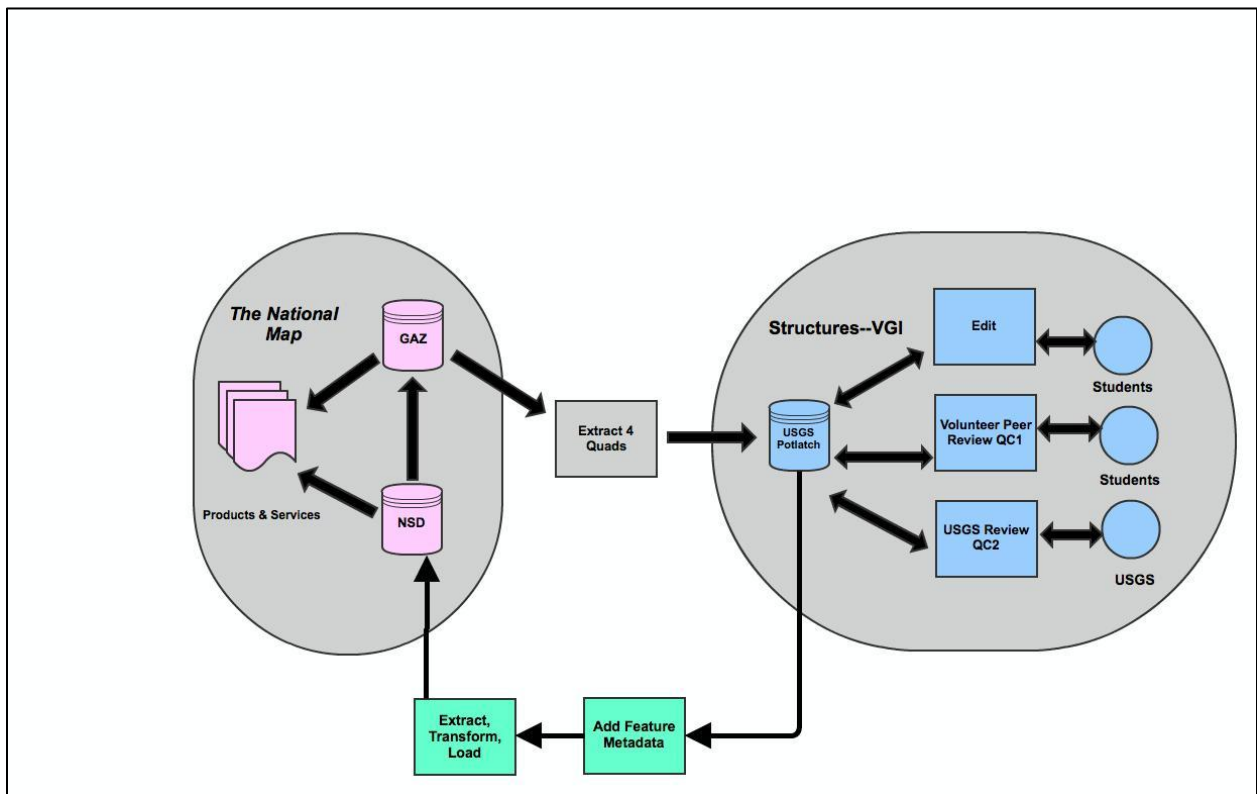
Figure 7.   Workflow for volunteer process used in Structures—VGI.

In order not to have students duplicate work on already-existing structures points, data for the four Denver quads were extracted from GNIS (GAZ) and loaded into the USGS Potlatch software (right side of fig. 6, and fig. 7). In the first phase (EDIT), the students had to research the existing points to update the locations and attributes. The students also added points not included in GAZ. Two rounds of quality control followed—the first by students (QC1), and the second by USGS (QC2). In QC1 and QC2, 100 percent of the structures were checked for positional and attribute accuracy and for conformance to editorial guidelines for names and addresses (appendix A). After these reviews, in an aggregation step, metadata that identifies the

source of each individual feature was added, and Safe Software's FME was used to exchange, transform and load (ETL) the data into the NSD. NSD data are used for products and services for *The National Map.*

Working with Student Volunteers

For the Structures—VGI project, the USGS worked with students enrolled in GIS courses at the University of Colorado Denver (UCD) and the University of Denver (DU) because they were enthusiastic about GIS and had knowledge of the local community. This decision was mutually beneficial for the students and the USGS. The students gained real world, resume-building experience and received an official certificate for their contributions to *The National Map*. This project also exposed students to the growing movements in open-source mapping and crowdsourcing (outsourcing to a distributed group of people). The USGS benefited from quality-controlled data that could become part of *The National Map*. The USGS was also able to test instructional materials and the editing interface and to assess how well crowdsourcing would work for structures data collection with relatively inexperienced users.

USGS employees made presentations during GIS classes at both universities. Presentations included descriptions of the USGS and *The National Map*, an introduction to crowdsourcing and OpenStreetMap, and an overview of the Structures VGI project and Web site. The Potlatch 2 editor was demonstrated, and methods for researching point locations and attributes were discussed. Students from UCD signed up to participate outside of class. Students from DU performed most of their edits in class with USGS personnel on hand to answer questions, but some worked on the project outside of class. DU students were given credit for their participation. A total of 44 students from both schools performed edits on the database, and a total of 41 students from both schools participated in quality control (table 3). There were a number of students, particularly at UCD, who signed up but did not participate, and there were also several students who edited data but did not participate in quality control.

Table 3.   Number of student participants in Structures—VGI.

School	Editing	Student Quality Control
University of Colorado Denver	26	23
University of Denver	18	18
Total	44	41

Volunteer Editing Process

Structure Types

To guide contributors and contractors who collect structures data for *The National Map*, the USGS has identified 30 priority structure types (USGS, 2012). The Structures—VGI project was designed so that only these 30 structure types were shown in the interface (table 4). Icons for each structure type were arranged in the left margin of the Potlatch map. Students could edit the

position of structures from the GAZ database that were pre-loaded into the system by directly moving the point on the map, or they could add structures missing from the GAZ database by dragging and dropping the symbols onto the map. Students clicked on each point to modify the attributes.

Table 4. List of 30 feature types and their definitions collected by students in Structures—VGI.

Feature	Description
School	A building or group of buildings used primarily/exclusively as a learning center for children grades pre-kindergarten through high school.
School: Elementary	A school for the first four to eight years of a child's formal education, often including kindergarten.
School: Middle School	A school between elementary school and high school, usually having three or four grades, variously including grades 5 through 9.
School: High School	A secondary school attended after middle school that usually includes grades 9 or 10 through 12.
School: University	A building or group of buildings used exclusively as an institution of higher learning that grants a degree at the completion of an extensive course of studies.
Fire Station/EMS Station	A facility housing fire-fighting equipment and/or personnel and a provider of emergency medical services.
Law Enforcement	A facility housing law enforcement equipment and personnel.
Prison/Correctional Facility	A facility for the confinement of persons convicted of crimes.
State Capitol	A building occupied by a State legislature.
Court House	A building that is home to a local court of law and sometimes the regional county government as well.
City/Town Hall	A building that contains the offices of the public officials of a city, town, or community.
Post Office	Facility operated by the United States Postal Service for the collection, processing and distribution of mail.
Hospital/Medical Center	Medical institution providing medical or surgical care.
Amusement/Water Park	A permanently located facility where there are various devices for amusement and entertainment, rides and booths.
Auditorium/Concert Hall/Theater/Opera House	A building with a stage and audience seating, used for the performing arts or meetings, not sporting events.
Campground	A developed designated site for camping in tents and/or recreational vehicles.

	Cemetary	A place or area for burying the dead.
	Convention Center	A building or complex of buildings designed and used for holding large assemblies of people.
	Fair/Exhibition/Rodeo Grounds	An area where permanent facilities exist to hold outdoor fairs, circuses, exhibitions, or rodeos.
	Historic Site/Point of Interest	Landmark of natural, cultural or historic interest.
	House of Worship	A building used for religious worship.
	Ice Arena	A facility containing a smooth expanse of ice used for skating.
	Library	A place in which literary or artistic materials, such as books, periodicals, newspapers, pamphlets, prints, records, and tapes, are kept for reading, reference, or lending.
	Lighthouse/Light	A structure on or off a coast exhibiting a major light designed to serve as an aid to navigation.
	Museum	A building or area used exclusively for preserving and exhibiting artistic, historical, or scientific objects.
	National Symbol/Monument	A site maintained for the public, often a statue or monyment, declared to be of national significance.
	Outdoor Theater/Amphitheater	An outdoor area consisting of a stage that is surrounded by rising slopes (either naturally occurring or artificially constructed) supporting tiered seating.
	Racetrack/Dragstrip	A constructed course with either natural or artificial surface prepared expressly for racing horses, dogs, automobiles, or race cars.
	Sports Arena/Stadium	An athletic field partially or completely surrounded by a structure designed to allow spectators to stand or sit and view the event.
	Zoo	A site where whild animals are kept for exhibition to the public that may also support breeding and/or study.

Assignments

Each student was given a specific assignment with a specific completion time—typically within 4 weeks. Assignments were organized so that the work was equalized as much as possible, based on USGS estimation of the number and relative difficulty of editing the structure types. Some students edited multiple structure types over several quads while some edited one structure type over all quads. Step-by-step guidelines for using the Potlatch 2 editor and for completing the editing and quality control processes were accessible through the project Web site (fig. 8). An online forum was set up for students to ask questions and post tips and tricks (fig. 9). Students tracked their progress and the number of hours spent on each feature code and USGS quad (fig. 10). Most students completed the editing assignments within a week or two of the due dates. Volunteers who completed their assignments received a certificate of participation from the USGS.

ADD A NEW FEATURE

You can add new points using the Aerial Imagery. This procedure is similar to editing a previously existing point but you must first drag the icon onto the map. All of the specifics outlined above should be followed when adding a new point.

Add a point for each unique structure *function* (e.g., police station, fire station), rather than adding a point for each unique structure *form* (e.g., building).

This medical facility is one building but contains multiple different entities. Therefore each function gets a point.

STEP 1: Identify feature not currently on map
- Use acceptable sources (as described above) to verify its existence and location.

STEP 2: Place point on map
- Drag the appropriate icon from the menu on the left to the feature's location on the aerial imagery.

STEP 3: Edit attributes
- While the icon is still selected, edit all of the attributes described in the editing existing points section above. Because the point is new, none of the attributes will be filled out. Please be sure to follow the guidelines for acceptable sources.

STEP 4: Change the Status to Edit

Figure 8. Sample page from user guidelines.

Figure 9. Online message board for students.

Figure 10. Example of student editing assignments.

Editing Structures Data

To edit existing points from GAZ, student volunteers had to determine whether structures were in the proper location, whether the attributes were correct and complete, and whether the

database contained all instances of a particular feature—for example, all the schools. Structures were to be located using the Denver-area USGS aerial imagery derived from NAIP with a resolution of 1 meter (USGS, 2008). Students were instructed that other background maps in the editor such as the OpenStreetMap image in the view screen, or the USGS quad reference maps, were only to be used for general confirmation. Locations, names, and addresses were to be verified by consulting an official source as discussed above. In some cases, this required a moderate amount of additional research, and students were sometimes asked to verify attributes by telephone. Names and addresses had to conform to the editing specifications, an example of which is given in appendix A. Students were also given instructions about how to deal with duplicate points.

To edit data, students clicked on a point in the Potlatch map interface and a pane with three tabs—Basic, Source, and QC Status—appeared to the left of the map (figs. 11 and 12). The fields in the Basic tab represent the set of attributes collected about a structure's name, address, and feature type as discussed in table 2. The Source tab allowed the volunteer to specify simple feature metadata about the source of data—such as where the point is positioned in relation to the structure. Centroid is the preferred position. In addition, volunteers were to specify how locational and attribute information were obtained, for example, by using aerial imagery (see table 1) . The QC Status tab indicated the stage in the quality control process for each point (see table 5).
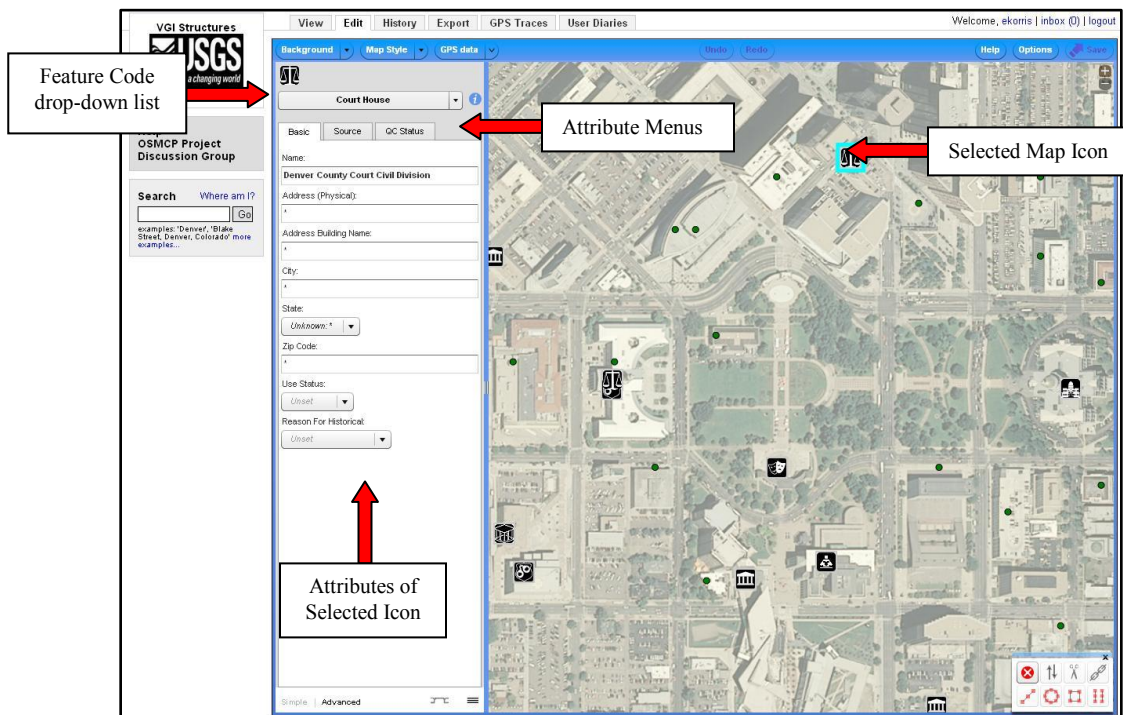
Figure 11. Potlatch editing interface showing icons and attributes.
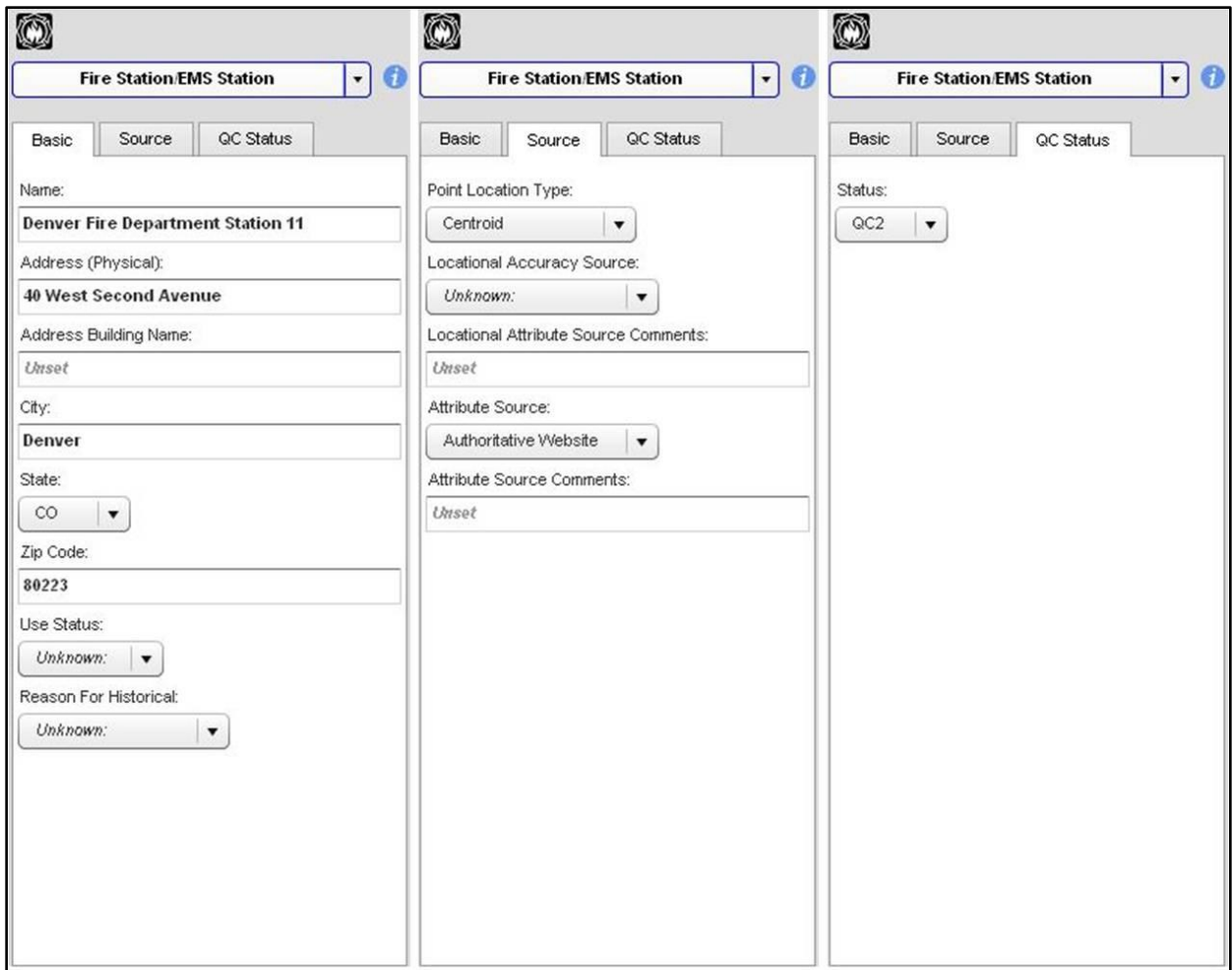
Figure 12. Editing tabs in Potlatch interface.

Quality Control

As the students finished editing they were given a quality control (QC) assignment. They were asked to check the data that other students had edited for one structure type within one quad. They reviewed the locational accuracy of the structure using the same techniques and resources used during the editing phase, and checked the spelling, formatting, and completeness of attributes according to the editorial guidelines (appendix A). Student editors also tracked the amount of time spent on editing and quality control. A second round of quality control was performed in house by USGS personnel following the same methods.

The edit and quality control phases resulted in four databases that were compared to analyze the differences from phase to phase. These are shown in table 5 along with the color coding that was used to indicate the project phase on the editing map.

Table 5.   Status of point, as indicated by color coding in interface.

Phase	Symbol	Description
GAZ		Initial data pre-loaded into the system from GNIS, not yet edited
Edit		Data that have been edited or added by student volunteers
QC1		Quality control, first stage. Students have checked the work of others
QC2		Quality control, second stage. USGS has checked data from QC1

Results—Data Quality

Methods

The data quality analysis compared the four datasets shown in table 5 in the sequence in which they were produced. That is, the results after the Edit phase were compared to the initial GNIS (GAZ) upload. The results of the student quality control, QC1, were compared to the Edit results, and in the final phase (QC2), USGS employees checked the quality of QC1. The software tracks each structures point contributed to the map, recording user identity, changes to location, attributes, and the time the change was made. Changes are stored in an OSM XML-standard planet file that the USGS downloaded nightly using Safe Software FME 12 Beta. The file from the previous day is not overwritten so a complete history of every point is available for a detailed analysis of the multiple changes to each structure as it was edited and quality checked.

For structures data contributed by contractors and partners (see fig. 6), USGS sampled a percentage (typically 10 percent) of the points for quality control. In Structures—VGI, a quality analysis of each structures point was undertaken. There were several reasons for this:

- No baseline of volunteered data against which the Structures—VGI data could be assessed existed, nor was the reliability of contributors known from previous experience. This project was expected to establish such a baseline for future volunteer projects.
- Given the small number of points, a statistical analysis could be skewed by outliers.
- A detailed analysis could demonstrate the reliability of VGI data and the VGI approach for *The National Map*.
- A thorough quality analysis of these data could provide feedback on how the editing and quality control processes in Structures—VGI might be improved for future phases of the project.

The USGS requirements for certifying structures obtained by contractors and partners for inclusion in the National Structures Dataset (USGS, 2012) formed the basis for the comparison of the four sets of points (table 5). These requirements can be classified according to the classic data quality measures used in the mapping community (Goodchild, 2007). These are positional accuracy, attribute accuracy, completeness, logical consistency, and lineage.

Positional Accuracy (Horizontal). Students were instructed to position structures points using the digital orthoimagery supplied by *The National Map* (table 1, fig. 4). The orthophotoquads

accessible through the Potlatch viewer are derived from the NAIP with a ground sample distance of 1 meter. NAIP imagery was collected under a contract stipulating that "all well-defined points tested shall fall within 6 meters of true ground at a 95% confidence level (NAIP, 2012)."

USGS digital orthophotographs, derived from NAIP imagery, must meet horizontal National Map Accuracy Standards (NMAS) at 1:24,000 and 1:12,000 scale, respectively. The NMAS specify that 90 percent of the well-defined points tested must fall within a radius of 40 feet (12.192m), commonly referred to as circular map accuracy (CMAS) (USGS, 1996; Zandbergen, 2008). Since field-based measurement by GPS or survey of actual ground truth is beyond the scope of this project, NMAS provides the standard by which horizontal accuracy is judged. The final quality check by USGS employees (QC2) visually compared each structure against NAIP imagery, thus the resulting dataset was taken as the reference dataset against which to measure the three test sets, GAZ, Edit, and QC1. Statistical sampling was not used, as each point was examined. Data were collected and managed in WGS 1984, a geodetic coordinate system (latitude and longitude). For the error calculation, the data were reprojected into the Colorado State Plane—North (FIPS 0501) which uses the 1983 North American Datum (meters). Positional accuracy is reported as mean absolute error (MAE) (eq. 1). Spatial data accuracy is commonly reported as the root mean square error (RMSE) but MAE is used here because RMSE is sensitive to large outliers. A confidence level is not included in this analysis since statistical sampling was not used.

$$MAE = \frac{\sum \sqrt{(x_{test} - x_{ref})^2 + (y_{test} - y_{ref})^2}}{n} \qquad \text{eq. 1}$$

Error is reported for structures that appeared in the original GNIS (GAZ) and also appeared in the reference dataset (QC2), and also for structures that were added in the Edit and QC1 rounds. The error value was calculated as the radial linear distance in x and y between the test point and the reference point expressed in meters. The errors were averaged and reported as MAE.

MAE was calculated for several other projections, but the difference in errors among all projections was less than a meter, well within the required positional error. By visual inspection of each point against the imagery, it was determined whether the point was in one of four possible locations, the center of the structure, the entrance or exit, a turn-off location, or an adjacent location.

Attribute Accuracy: For attribute accuracy, QC2 was the dataset against which others were checked using the same official sources first used by students in the Edit phase. In addition, each name and address in Edit and QC1 was checked for proper formatting according to the established standard (appendix A).

Completeness: There is no absolute baseline for completeness of structures data. New structures are continually being built and old structures destroyed or repurposed. Since the GNIS (GAZ) database originated in 1987, structures were added in each phase, based on official

sources such as county and city Web sites. Fidelity to these source materials was the best indication of completeness, but this value was not directly measured.

Logical Consistency and **Lineage** did not specifically form part of this analysis; however, the metadata on the points loaded into *The National Map* cite the volunteer program.

Table 6 provides an overview of the quality analysis. In this table, each row begins with the number of points that served as input to that phase and ends with the final number of structures at the end of the phase.

Table 6. Summary statistics on the student volunteer data for each phase

Phase	Structures in Initial Data-base	Existing Structures					New Structures		Structures in Final Database
		Number of Location Changes	Percent Location Change[1]	Number of Any Attribute Changes	Percent-Attribute Change[2]	Number Deleted	Number Added	Percent Added[3]	
Student Edit of GAZ and Collection of New Points (EDIT)	938	491	52[1]	633	67	19	187	20	1,106
Student Peer Review (QC1)	1106	205	19	700	63	26	75	7	1,155
USGS Review (QC2)	1155	108	9	1017	88	138	197	17	1,214

[1]Percent location change = number of changed structures / number of initial structures – deleted structures x 100.

[2] Percent attribute change = number of structures with any attribute change / number of initial structures – deleted structures x 100.

[3] Percent added = number of added / number of final structures x 100.

Analysis

Table 6 records changes to existing points, that is, structures points that were in the dataset at the beginning of each phase. We have recorded how many of these points were moved and calculated the percentage of points that were moved. The location changes were performed using NAIP imagery as a reference dataset. Each successive phase showed fewer location changes. Attribute changes to existing structures refers to any change in any attribute. These included changes to the mandatory elements (table 2) and also the non-mandatory elements (fig. 8). The figures for changes in attributes are somewhat inflated, as one point could have more

than one change. A certain number of points were deleted, from the database in each round, most likely because the structure in question no longer existed or had been converted to a function that was not part of the 30 structure types, for example, when a school became a condominium. A more detailed analysis of attribute changes is given below (see table 9).

Positional Accuracy:

In the first round of edits to the initial GNIS (GAZ) database of 938 points, 52 percent were moved (location change). In the student review (QC1), 19 percent of the 1,106 points were moved. The USGS review (QC2) resulted in 9 percent of the 1,155 points being moved.

The mean absolute error in point locations of each dataset, calculated as described above, is given below. In this analysis, 10 points out of the total of 1214 were found to be greater than 2 standard deviations from the mean. They constitute 0.8 percent of the points. These were removed from the dataset before analysis. These 10 points must be investigated further to determine the source of error. To give one example, a structure from the GAZ, the "Mountain View Church of God of Prophecy" was moved 14,842 m and renamed "Church of God Prophecy" in the Edit phase with no clear indication that this was the same congregation or an entirely new one. In the final USGS quality check, QC2, the name was changed back to the original, but the new location was retained.

In the initial database, only 49.57 percent of the points met the NMAS requirement of 1/5 inch at 1:24,000 map publication scale. After the first volunteer phase of edits, 84.22 percent met the NMAS requirement. After the volunteer peer review, the data exceeded the minimum requirements in the NMAS for 1:24,000 map publication scale (table 7).

Table 7.   Mean Absolute Error (MAE) in meters of all points in GNIS (GAZ), the volunteer edit, and peer review phases measured against points in the reference dataset (QC2) and the percent of points meeting NMAS accuracy requirements.

	MAE (meters)	Meets NMAS (error <12.129 meters) (percent)
GNIS (GAZ)	42.79	49.57
Volunteer Edit and Collect (EDIT)	20.10	84.22
Volunteer Peer Review (QC1)	13.67	92.23

Considering only structures added by the volunteers in the Edit phase (table 8), the MAE was slightly larger than the MAE of the entire database (table 7), but a higher percentage of

structures met NMAS. After volunteer peer review (QC1), data added by volunteers met minimum requirements in the NMAS for 1:24,000 map publication scale.

Table 8. Mean Absolute Error (MAE) in meters of only points that were added during volunteer edit and peer review phases as measured against points in the reference dataset (QC2) and the percent of points meeting NMAS accuracy requirements.

	MAE (meters)	Meets NMAS (error <12.129m) (percent)
Volunteer Edit and Collect (Edit)	21.54	89.19
Volunteer Peer Review (QC1)	11.84	91.14

Attribute Accuracy:

To analyze changes in attributes over the three datasets, attributes were grouped into three categories according to their importance in identifying the structure (table 9).

Critical: change to the feature code or the name.

Address: Change to Address, Address Building Name, City, State, ZIP Code.

Source: Change to the positioning of the point in relationship to the actual structure, that is, whether it is at the centroid or in another position (see explanation for table 2), as well as change to the source of the location information — derived from aerial imagery or another source.

Table 9. Accuracy of key attributes.

Phase	Number of Attributes Changed		
	Critical	Address	Source
Student Edit and Collect (Edit)	144	478	579
Student Peer Review (QC1)	35	100	211
USGS Review (QC2)	267	253	203

The large number of changes to attributes, particularly for name and address, in the USGS review phase (QC2) was due to the fact that these categories had multiple elements that might count as corrections, thus inflating the numbers, and also was due to the complex and precise editorial guidelines the students were expected to follow (appendix A). Many of the changes counted as errors in the name and address fields were of the following types:

- Not including the denomination of a church. For example, in QC1, "A New Beginning Church," and in QC2, "A New Beginning Pentecostal Church."
- Not following the standard in the ordering of elements. For example, in QC1, "Arvada Branch Jefferson County Public Library," and in QC2, "Jefferson County Public Library Arvada Branch."

Completeness:

Nine hundred thirty-eight structures were loaded into the database from GAZ at the start of the project. These were updated and enhanced by an additional 262 structures that were added in both student volunteer phases (Edit and QC1). The subsequent in-house review by the USGS (QC2) discovered an additional 197 structures that were appropriate for inclusion in the NSD, bringing the total of structures in the four quads to 1,214.

Lessons Learned

Working with Volunteers

- **Participation:** Participation was inconsistent, which generally aligns with experience of other VGI projects in which the majority of contributors sign up for an account and either never participate or contribute only a few points (Budhathoki and others, 2010). While the University of Denver students were required to participate in the project, the University of Colorado Denver students were not. Many students from the University of Colorado Denver who expressed interest in contributing did not follow through.

- **Timeliness:** It was difficult to encourage students to follow a specific timeline. Many did their editing at the last minute, and a significant number asked for extensions. The fact that there was a time limit likely increased the amount of data generated by the project, giving a potentially false picture of how much data can be expected in a given unit of time. This will need to be carefully considered in designing the next phases of the project. There are successful crowdsourced projects such as OpenStreetMap that do not impose a time limit; others do operate under time constraints, such as the emergency nature of the volunteer mapping of Haiti after the 2010 earthquake or the Audubon Society's Christmas Bird Count, which takes place over several weeks during the Christmas holidays each year.

- **Communication:** An online message board was set up so that students could ask questions and get answers from USGS staff and other students. It was expected that this would have a multiplier effect, improving the practical knowledge base of the project. The message board was not used as heavily as had been anticipated; most of the student questions were directed via email to the USGS point-of-contact rather than being posted to the message board. It may be that more active moderation and encouragement from USGS personnel are needed.

- **Usability:** During the in-house QC process, the USGS realized that the tasks may have been too difficult and complex. For example, the Potlatch editor had many functions inherited from OSM that may have caused problems. For example, volunteers had to save their work before it was committed to the database, and there was no prompt or reminder built into the system. The editor will be simplified in subsequent phases. Despite the availability of written guidelines on editing and QC, students had trouble deciding which structures qualified for historical designation, that is, structures that were in the initial GNIS (GAZ) database but which do not currently exist or have a changed use. Due to this difficulty, historical and inactive structures (61 points) were omitted from the data analysis. Students also had difficulty formatting names and addresses according to the exacting editorial specifications that were supplied (appendix A). Future projects will need to invest more time in simplification of tasks and instructions.

Technology

The original plan for Structures—VGI was to use Safe Software FME 12 Beta to read the XML planet file generated nightly from the database. The FME 12 Beta could work with the OSM file format, and it is a tool commonly used by the USGS for data development. It became apparent that different people worked at different rates. In order to encourage continued participation, the faster volunteers were allowed to start the QC1 phase while other volunteers

were still working on their initial edits. The end result was that the nightly planet files contained a mix of structures in EDIT and QC1 phases. FME was not able to completely separate the data from each phase (EDIT, QC1, QC2). Instead, a Python script was created using a Python module developed by an OSM community member to extract data from the OSMCP server. The Python script could look at each version of each structure in the version history stored on the server and select the appropriate version for each phase. The Python script was further extended to manage the complexities of the data development process. Using a more complicated tool, Python versus FME, that can distinguish among the updates results in more labor on the part of USGS to manage the software. In retrospect, FME might have been used successfully had the volunteer editing and quality control processes been more rigorously structured so that all volunteers moved at the same pace. However, it is possible that a more rigid structure could have discouraged the participation of volunteers eager to move on to quality control.

Process

Despite the numerous modifications student volunteers made to the data in the student peer review phase (QC1), volunteer QC alone was not entirely effective in improving data to current quality standards. A large number of changes were made during the internal USGS review process particularly in the area of identifying new points. This suggests that during QC, volunteers may have been focusing more on checking the existing points only, as opposed to also checking for new points. This information is important feedback that can be incorporated into training materials and training events during future phases of this project.

Additionally, the number of modifications that were made to existing points during internal USGS QC suggests that the process may have been too complicated. Changes to the data were required in order to make them conform to specifications; however, many of these changes were not major changes but were minor changes that were required to meet the data requirements of the NSD and GNIS programs. The edits made by the USGS mainly consisted of adding structures that had been added by the USGS and which the students had not found (completeness) and making editorial changes to the formatting of names and addresses (attribute accuracy). Given a larger cohort of volunteer mappers, completeness could be expected to improve over time, while formatting issues might be handled by automated software routines. Analysis of volunteer errors provides important feedback for future phases of the project. A review of the attributes should be conducted to determine which attributes are "essential" for future volunteers to collect and validate.

The Future

Structures—VGI established a process that generated VGI data of sufficient quality for incorporation into *The National Map*. The next phase of this project will explore how well this process scales to larger geographies and a greater number of more loosely coordinated volunteers. The USGS plans to collect data over the entire State of Colorado and to involve all types of volunteers from organized groups such as the Boy Scouts and from ordinary citizens. Data collection will be simplified, reducing the number of structures and mandatory attributes. The user interface and the editing guidelines will also be improved.

For More Information

The next phase of this project, collecting structures over the State of Colorado, is now active. More information can be found on *The National Map* Corps Web site (*http://nationalmap.gov/TheNationalMapCorps/*).

Acknowledgments

References

Budhathoki, N.R., Nedovic-Budic, Z., and Bruce, B., 2010, An interdisciplinary frame for understanding volunteered geographic information: Geomatica, v. 64, no., p. 11-26.

Goodchild, M., 2007, Beyond metadata: Towards user-centric description of data quality: Proceedings International Symposium on Spatial Data Quality, Enschede, Netherlands, p. 1-11.

Haklay, M., 2010, How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets: Environment and Planning B: Planning and Design, v. 37, no. 4, p. 682-703, doi:10.1068/b35097.

Orth, D.J., and Payne, R.L., 1987, The National Geographic Names Data Base: Phase II Instructions: U.S. Geological Survey Circular 1011, 133 p.

U.S. Board on Geographic Names, 1997, Principles, Policies and Procedures: Domestic Geographic Names, Reston, Va, United States Board on Geographic Names, 56 p.

U.S. Bureau of the Budget, 1947, United States National Map Accuracy Standards: Washington, D.C., accessed April 24, 2012, at *http://nationalmap.gov/standards/pdf/NMAS647.PDF*.

U.S. Geological Survey, 1996, National Mapping Program Technical Instructions, Part 2 Specifications, Standards for Digital Orthophotos: Reston, Va., accessed April 24, 2012, at *http://nationalmap.gov/standards/pdf/2DOQ1296.PDF*.

U.S. Geological Survey, 2008, National Agricultural Imagery Program (NAIP) Orthoimagery for Zone 13 (4 Band) [metadata for Arvada, Colo., Quadrangle]: accessed May 5, 2012, at *http://extract.cr.usgs.gov/distmeta/servlet/gov.usgs.edc.MetaBuilder?TYPE=HTML&DATASET=NAIP_Z13_G4*.

U.S. Department of Agriculture, 2012, National Agricultural Agriculture Imagery Program (NAIP): Information Sheet, accessed March 22, 2012, at *http://www.fsa.usda.gov/Internet/FSA_File/naip_2012_infosheet.pdf*

U.S. Geological Survey, 2012, Guidelines for Contributing Structures Data to *The National Map*: Reston, VA, U.S. Geological Survey.

U.S. Postal Service, 2010, Publication 28: Postal Addressing Standards: Washington, D.C.

Wolf, E.B., Matthews, G.D., McNinch, Kevin, and Poore, B.S., 2011, OpenStreetMap Collaborative Prototype , Phase One: U.S. Geological Survey Open-File Report 2011-1136, accessed June 26, 2012, at *http://pubs.usgs.gov/of/2011/1136/*.

Zandbergen, P. A., 2008, Positional accuracy of spatial data: Non-normal distributions and a critique of the National Standard for Spatial Data Accuracy: Transactions in GIS, v. 12, no. 1, p. 103-130.

Appendix A. Editorial Guidelines for populating the "Address" Attribute Fields in The National Structures Dataset

ATTRIBUTE NAME	EDITORIAL GUIDELINE FOR POPULATING THE ATTRIBUTE FIELD	EXAMPLES OF THE GUIDELILNE	
		INCORRECT	CORRECT
ADDRESS	NO ABBREVIATIONS (Spell out street types, prefixes, suffixes and directionals).	Ave	Avenue
		Blvd	Boulevard
		Cnty	County
		CR	County Road
		Dr	Drive
		E	East
		Hwy	Highway
		I	Interstate
		Int	Interstate
		NW	Northwest
		Pkwy	Parkway
		RT	Route
		Rte	Route
		SR	State Road
		St	Street
		SW	Southwest
		TSR	Township Road
		W	West
		Rd	Road
		Byp	Bypass
	NO APOSTROPHES IN STREET NAMES	Johnson's Way	Johnsons Way
		Sherman's Pike	Shermans Pike
	STREET NAME VARIATIONS: Use the name exactly as it is given by the local naming authority.[1]	--	· Third Street · 3rd Street · 3 Street
	SPELLING OF COUNTY, STATE, LOCAL HIGHWAY AND INTERSTATE ROUTES: Spell out consistent with the US Postal Service Address Standard[2].	Interstate Highway 680	Interstate 680
		I 55	Interstate 55
		US HWY 44	US Highway 44
		US 41 SW	US Highway 41 Southwest
		Hwy 64	County Highway 64
		KY ST HWY 1	State Highway 1
		Township RD 20	Township Road 40
	TITLE CASE CAPITALIZATION (Each word begins with a capital letter.)	US highway 40	US Highway 40
		County road 441	County Road 441

[1] Different jurisdictions follow different practices for numbered street names. For example, Pittsburgh spells out "First" through "Tenth" in numbered street names.

[2] US Postal Service standard. Appendix F - Address Standardization - County, State, Local Highways (See pages 79-80 of the document, .pdf file pages 82-83). http://nd911.homestead.com/USPSpub28.pdf (accessed October 2010).