# Pittsburgh Street Addressing Project

Chris Temple
Pittsburgh Public Schools

Bob Regan
Glenn Ponas
Ken Sochats
Visual Information Systems Center
University of Pittsburgh

## Abstract

Pittsburgh Public Schools, in collaboration with the Visual Information Systems Center of the University of Pittsburgh, has developed, using ESRI's ArcGIS, the most accurate current street range dataset to provide timely geographic based information to the Pittsburgh Public School's administrators, principals, teachers and parents regarding the composition of schools and the feeder patterns for those schools. Also, this data was provided to the public and university communities for the purposes of research, public policy decision making, and public safety.

This project was developed as an outgrowth of the PPS/VISC Data Atlas and Factbook. The Atlas provides, on CD and via a web site, maps of the district, each school, neighborhood, political subdivision or other relevant subdivision showing the locations of students within that area. Maps are also available showing the location of libraries, bus routes, community centers and other important assets.

The rivers, hills and other geographic features, combined with a rather fluent history of Pittsburgh provided a unique set of problems for student address geocoding. This paper provides the rationale for the Pittsburgh Street Addressing project, reviews the project's history and developments, and outlines the plans for future development.

## Background

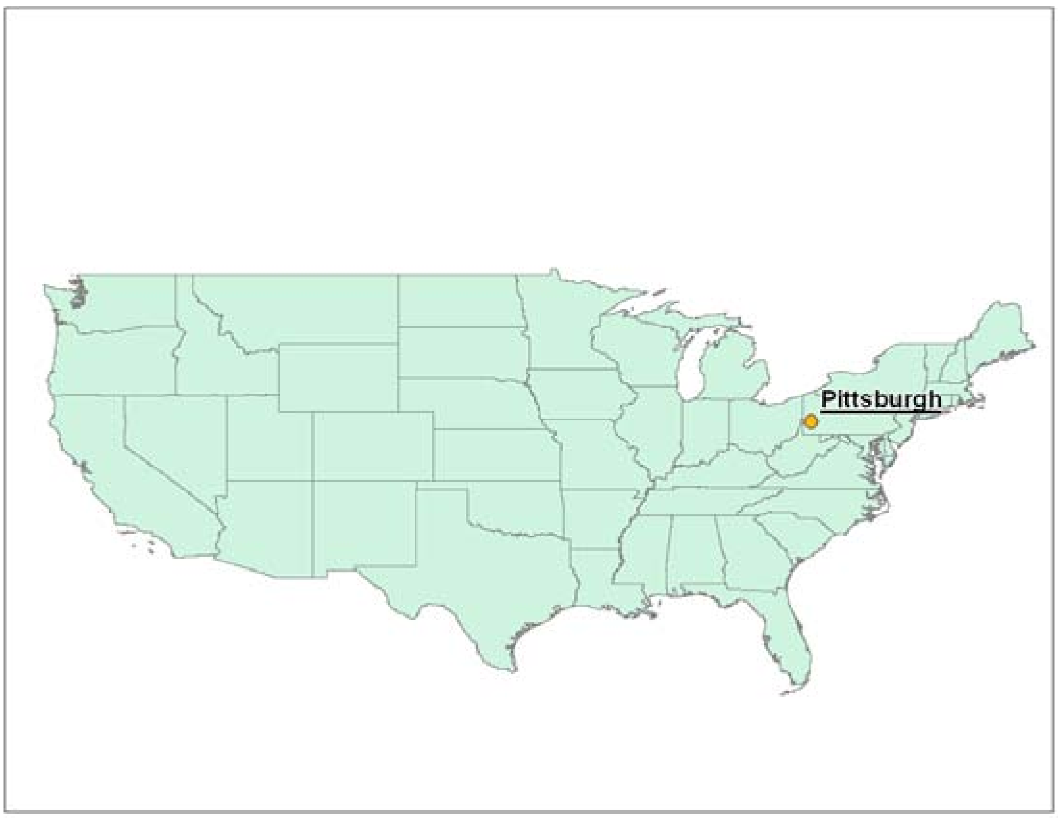
One method of locating a particular address within Geographic Information System (GIS) software, or any similar computer application, is termed address geocoding. In order to use this technique there must be digital street map data with address ranges for each street segment. Thus, each street segment will contain, in an ordered fashion, the lowest and highest even and odd addresses for that segment. Such information is utilized to locate residence and business addresses on maps and is vital to providing better municipal services, particularly emergency management services. This paper describes a unique project undertaken by the City of Pittsburgh, Pittsburgh Public Schools and the Visual Information Systems Center at the University of Pittsburgh to develop a methodology for defining accurate street range data.

## History and Rationale

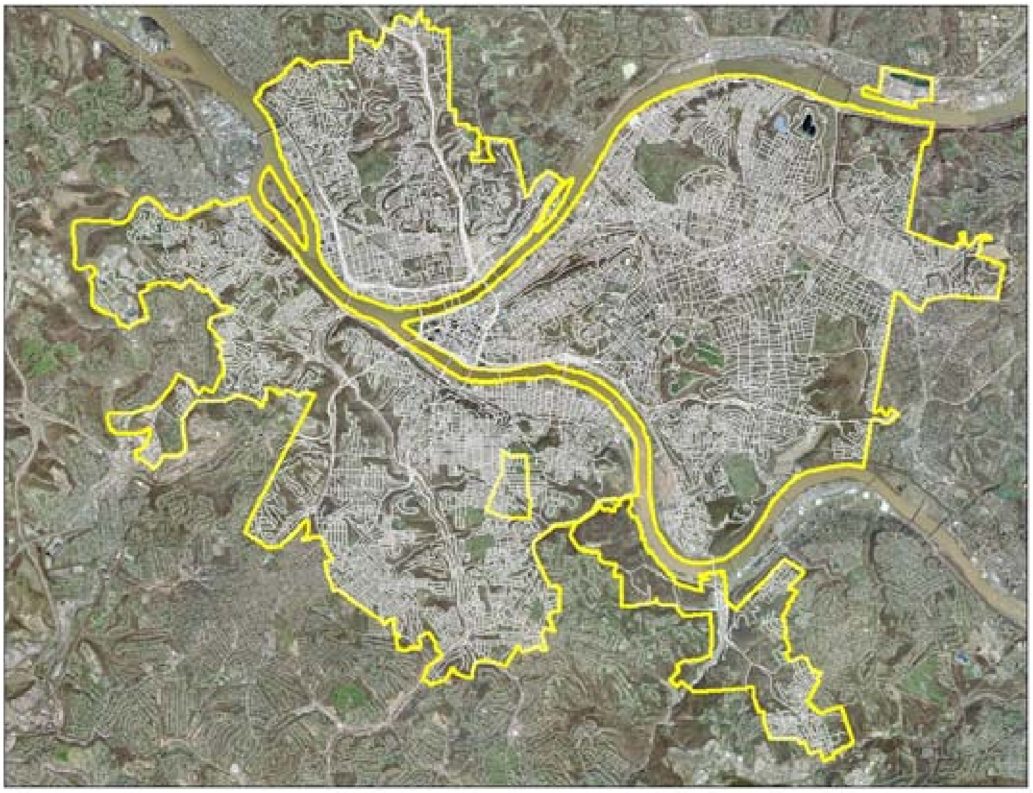
This project grew out of an effort by the Pittsburgh Public School District to analyze student performance and enrollment information called VIPER. In Pittsburgh Public

School’s case, the geocoding of student addresses was vitally important to any geospatial analysis of the student body.  The effort associated with this geo-coding was compounded by the unique natural and man-made characteristics of the city of Pittsburgh.

Pittsburgh is a medium sized city in the Mid-Atlantic region in Allegheny County in the Commonwealth of Pennsylvania. Located at the confluence of three rivers, The Allegheny, The Monongahela, and The Ohio, Pittsburgh has a current population of 335,000 people living in 163,000 households. Its 94 neighborhoods cover 55.5 square miles.



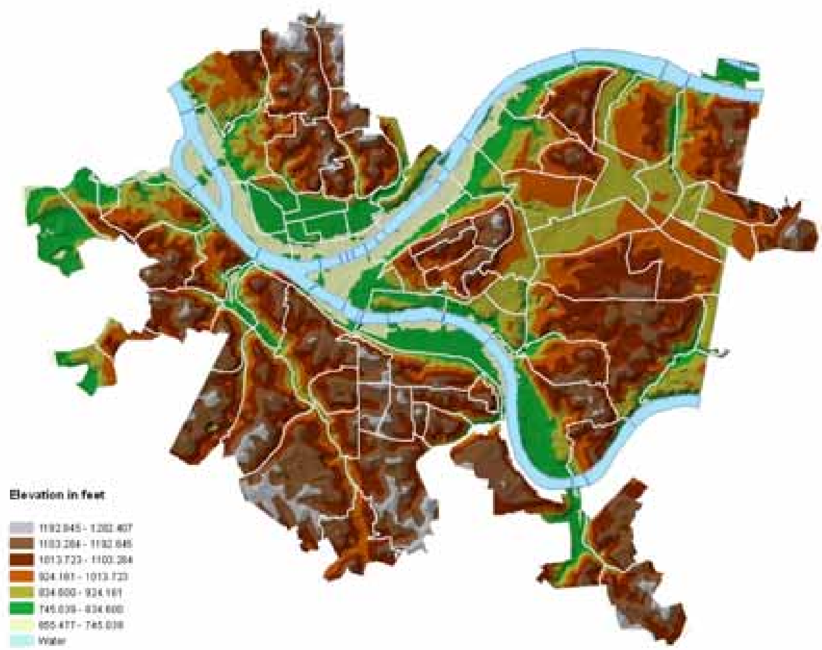

One of the most striking features about Pittsburgh is its irregular boundaries and landscape. Built on the sides of steep Appalachian river valleys, The City of Pittsburgh has enlarged itself over the past two centuries by annexing and consolidating adjacent communities, the largest of which were Birmingham, on the south side of the Monongahela River, and Allegheny City, located on the north side of the Allegheny River. Mt. Oliver, an independent borough completely surrounded by the city of Pittsburgh, is located in the south-central part of the city. Its borders are clearly visible in the map below. Since it is part of the Pittsburgh Public Schools District, and shares many services with its surrounding neighbor, Mt. Oliver streets were included in the street data editing project.



Typically, throughout the country, where streets are arranged in regular patterns, address ranges can be readily developed. Unfortunately, this is not the case in the Pittsburgh area wherein topography and historical development has resulted in an erratic and irregular distribution of streets and associated addresses. Unlike many newer cities, such as Phoenix, or rebuilt cities, such as Atlanta, Pittsburgh has grown and developed without any urban planning or visionary design to its street layout. During the steel boom of the late Eighteenth and early Twentieth Centuries, Pittsburgh's streets were in most cases created without official input or the approval of municipal authorities Indeed, Ernie Pyle's April 16, 1937-syndicated column, which focused on Pittsburgh, best summarized the situation. He wrote, in part:

> *"Pittsburgh is undoubtedly the cockeyedest city in the United States. Physically, it is absolutely irrational. It must have been laid out by a mountain goat. …"*

The map below gives an idea of the wide ranging geographical complexity of the city of Pittsburgh.




Over its relatively small area, Pittsburgh has dramatic changes in elevation, in many places jumping from about 700 feet above sea level to over 1400 feet above sea level in the course of less than a tenth of a mile. This rugged landscape is filled with dramatic cliffs and hillsides that plunge down to the riverside with very little flat land for development. Streets in Pittsburgh reflect this landscape by having sharp bends, steep slopes, and irregular paths.

When Pittsburgh Public Schools began the process of utilizing GIS technologies for analysis and management, the need for an accurate street data was of primary importance. A search of all available commercially and publicly available digital street data resulted in the conclusion that none is ideally suitable for accurate spatial location of addresses. While all will result in providing a location within a block or two of the actual location, none have the precision required for accurate location vital to making decisions that face the Pittsburgh Public School system.

## Geocoding

In the development of a dynamic GIS for Pittsburgh Public Schools, Christopher Temple, Senior GIS Analyst, faced a major challenge in geocoding the entire student data record set, encompassing over 30,000 records. Geocoding is defined by the ESRI technical document, “ArcGIS 9.0: Geooding in ArcGIS”, as:

> Geocoding is the process of assigning a location, usually in the form of coordinate values, to an address by comparing the descriptive location elements in the address to those present in the reference material. Addresses come in many forms,

> ranging from the common address format of house number followed by the street name and succeeding information to other location descriptions such as postal zone or census tract. In essence, an address includes any type of information that distinguishes a place. (ESRI Pg.1)

On the first pass of the student address geocoding process, less than 60% of the records matched to a spatial location with an acceptable degree of confidence. The file was checked for entry errors and, although a few were found, the main culprit in the low match rate was found to be the quality of the available street range data.

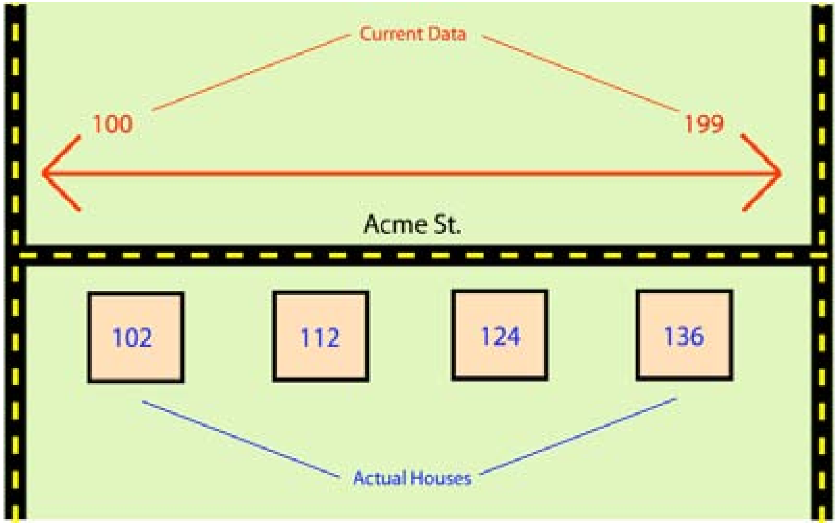
## Pittsburgh Street Address Problems

Mr. Temple discovered that no one street range dataset was good enough for the district's purposes. Data available from the City of Pittsburgh's Planning Division was spatially correct, but inaccurate in terms of numeric attributes. In the example below, the geocoded postal address points have a dramatically obvious tendency to "clump" at the ends of the street, while the building footprint file shows addresses all along the range. This is due to the fact that the street range numbers are much wider in scope than the extent of actual building addresses found along the segments.

Tiger based street data from GDT was much better numerically, but too spatially imprecise for use at the neighborhood and block level. The problems with the spatial orientation of the Tiger based file are visible when superimposed over aerial photography. The street centerlines simply do not lay on top of the streets. From a state or national level, this level of accuracy is acceptable. However, for school district use this level of error can lead to major problems. With 86 schools spread out over only 55 square miles, attendance patterns are measured in sub-block increments. Also, the overall vision of using GIS technology is to integrate information concerning all aspects of a student's life and environment into a concise picture. If the streets and, subsequently, the student points, do not accurately interact with other feature datasets such as parks, playgrounds, police stations, hospitals, etc., the value of GIS as a tool for communication and analysis is diminished.

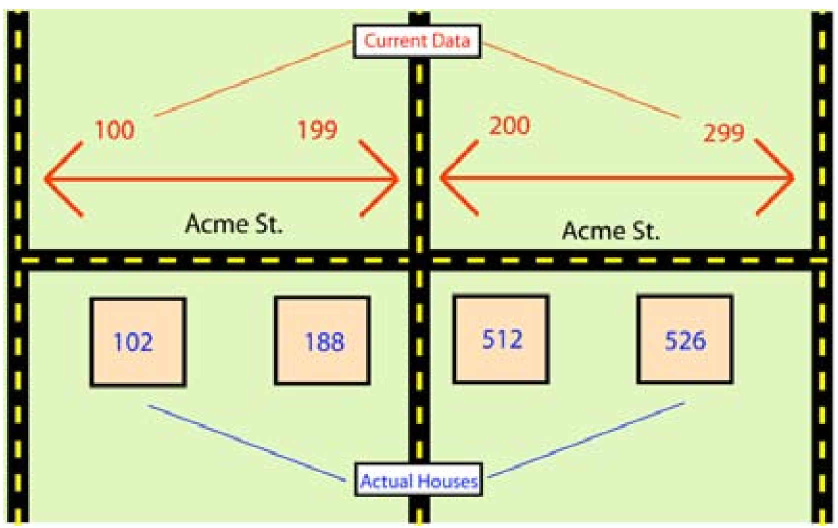
The major problems noted with the existing data were as follows:

a) Short Blocks—The typical address information in all digital street map data throughout the United States has contiguous hundreds based ranges for each block (or street segment), e.g., the 200 block would go from 200 to 299. Although all available digital street map data for the City of Pittsburgh adhere to this convention, it is not borne out in reality. There are 20,000 street segments in Pittsburgh and based on our analysis we would estimate that 90% are short blocks (and multiple short blocks). That is to say that in the 200 block example, the range in Pittsburgh is more typically 200 to 240 and often many segments have several ranges such as 200 to 240; 300 to 320, mistakenly reported in the street data as 200 to 400.



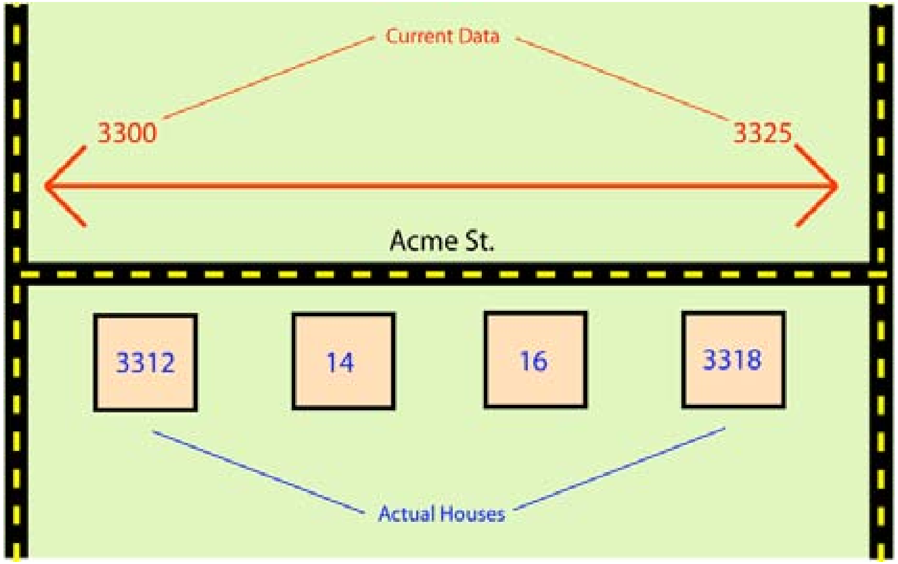
*Short blocks*

b) Missing Intervals—Throughout the United States the street range data is consistently monotonically increasing, i.e., the 100 block, the 200 block, the 300 block, etc. This is not true in about 50 % of the contiguous street segments in Pittsburgh where there are often times missing gaps, dramatic jumps in ranges, and even decreasing ranges.



*Missing Intervals*

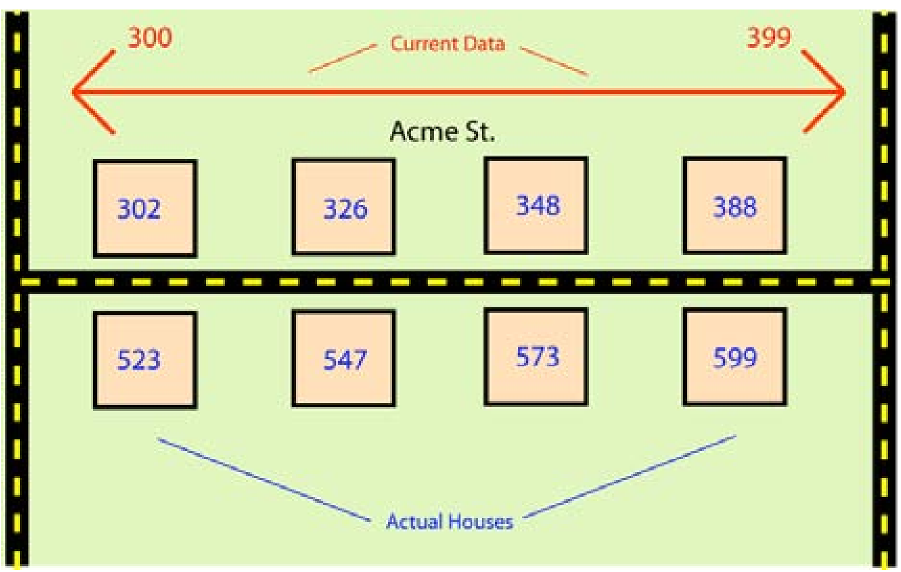
c) Mixed Addresses—Approximately 20 percent of the street segments in Pittsburgh contain mixed addresses. The predominant problem is 2-digit addresses mixed within a range of 3- and 4-digit address ranges. In many areas the older homes in a street segment will have 2 digit addresses while the newer homes have 3- and/or 4- digit addresses. The available digital street map data either do not or incorrectly reflect this.



*Mixed Addresses*

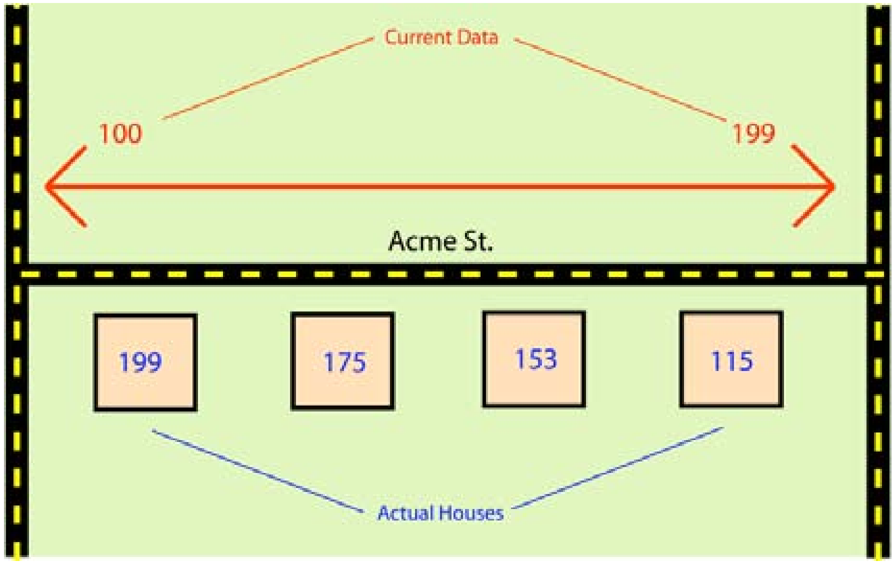
Another problem is that while in most areas even addresses are on one side of the street and odd addresses on the other, this is not true in approximately 5% of the street segments in the City of Pittsburgh.

d) Mixed Ranges—Approximately 1 % of the street segments in the City contain mixed ranges. An example of this is a segment with the 400 address range on one side of the street and the 500 address range on the other.



*Mixed Ranges*

e) Reversed Sense—Approximately 5 % of the available digital street map data street segments have a reversed sense wherein the addresses are increasing (or decreasing) in the wrong direction.



*Reversed Sense*

f.) <u>Steps and inclines are streets</u>—In over 300 instances, streets listed as navigable thruways on all available maps for both public and private use are actually what are known in Pittsburgh as "City Steps". City steps are publicly owned and maintained staircases providing access to physically steep sections of the city. In many cases homes have their main front address located on these steps. Defining which streets are actually steps (and not accessible by vehicle) is crucial in any automatic routing or transportation system.


*Photo Courtesy of Dr. Robert Regan*

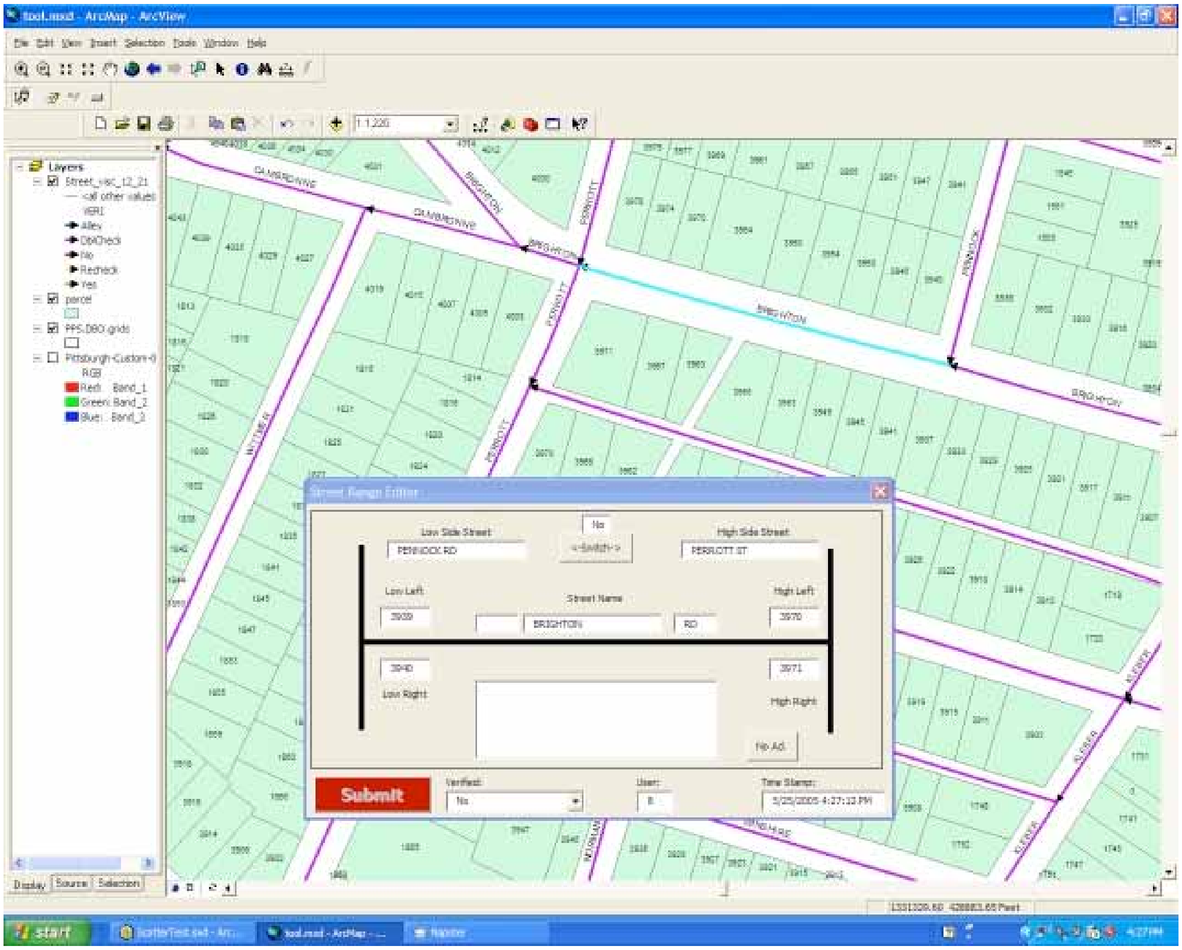
These problem areas are some of the reasons that it was not possible to accurately spatially locate a given address with the available digital street map data. In order to correct this situation, Visual Information Systems Center (VISC) personnel conducted an "on the ground" physical survey of the address ranges in every street segment in the City of Pittsburgh, The collected data was used to edit and revise the best available digital street map data file to provide one of unparalleled accuracy and precision.

## The Solution

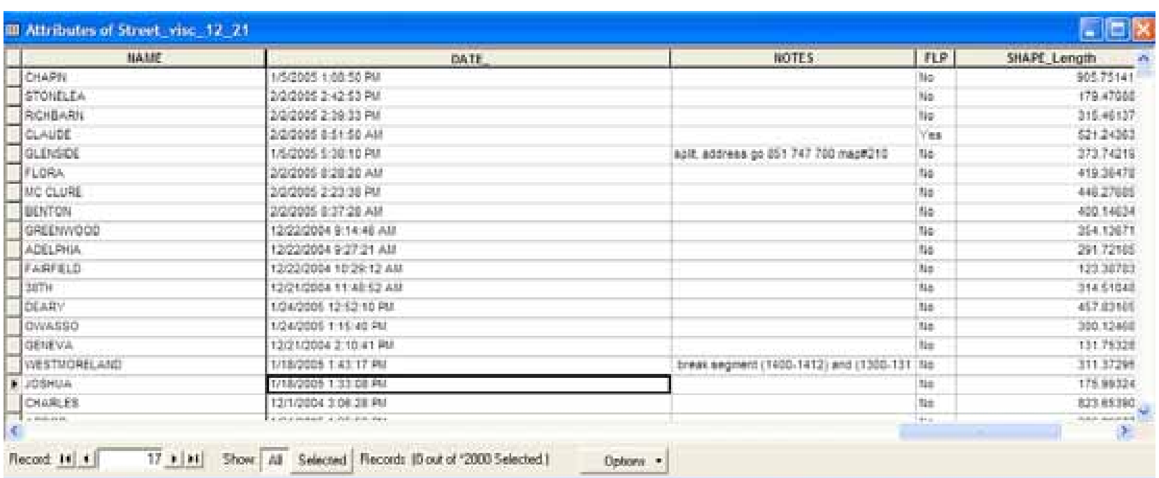
The solution to this problem came from collaboration between the Visual Information Systems Center's Dr. Robert Regan and Pittsburgh Public School's Christopher Temple. Using the Planning Division's street shapefile as a starting point, a total revision of the street range data for the city of Pittsburgh was performed in eight months and at no cost above the existing software licensing and the time cost of the two principal participants. The process began with Dr. Regan riding the streets on bicycle and reading the street ranges into a Dictaphone. Next, he transcribed his audio notations onto large format printouts, marking the high and low house numbers for each street segment as well as any necessary changes to street names and topology. Mr. Temple then took the paper maps, designed a custom ArcObjects interface built on top of ArcMap 9.0 and ArcSDE 9.0, and organized as group of Pittsburgh Public Schools interns to enter the data into a digital form using the Dr. Regan's paper maps as a guide.

The framework of this custom interface was designed and built around the goal of allowing absolutely inexperienced users to utilize fairly advanced ArcGIS functionality. Since the cast of intern assistants was constantly changing, the ability to sit new users down and get them up to speed immediately was crucial. The interface accomplished this goal by limiting the options for data entry errors, automatically tracking all changes by user and time, and by creating daily backups of the dataset.

A view of the custom interface is visible below. The pop-up window in the foreground displays editable fields that correspond to the database attributes of the selected street segment. After the user makes the necessary changes to the feature's attributes, selects their level of confidence in the accuracy of the information from a pull-down menu, and types in any necessary notations in the large central field, the "submit" button writes the newly entered information back into feature's database fields. The color symbology of the street layer is keyed to the editing status of the field so that multiple users do not waste efforts in redundant work and all members of the project team can easily see where work on the file is necessary.




A view of the data structure behind the street layer shows that information concerning the street attributes and editing status were collected and recorded in separate fields for quick analysis and display.



Attributes of Street_visc_12_21

| NAME | DATE_ | NOTES | FLP | SHAPE_Length |
|---|---|---|---|---|
| CHAPIN | 1/5/2005 1:08:50 PM | | No | 905.75141 |
| STONELEA | 2/2/2005 2:42:53 PM | | No | 179.47088 |
| RICHBARN | 2/2/2005 2:39:33 PM | | No | 315.46137 |
| CLAUDE | 2/2/2005 8:51:50 AM | | Yes | 521.24363 |
| GLENSIDE | 1/5/2005 5:38:10 PM | split, address go 851 747 700 map#210 | No | 373.74215 |
| FLORA | 2/2/2005 8:28:20 AM | | No | 419.36478 |
| MC CLURE | 2/2/2005 2:23:38 PM | | No | 446.27685 |
| BENTON | 2/2/2005 8:37:28 AM | | No | 400.14634 |
| GREENWOOD | 12/22/2004 9:14:46 AM | | No | 354.13671 |
| ADELPHIA | 12/22/2004 9:27:21 AM | | No | 291.72105 |
| FAIRFIELD | 12/22/2004 10:29:12 AM | | No | 123.38703 |
| 38TH | 12/21/2004 11:48:52 AM | | No | 314.51048 |
| DEARY | 1/24/2005 12:52:10 PM | | No | 457.83105 |
| OWASSO | 1/24/2005 1:15:40 PM | | No | 300.12468 |
| GENEVA | 12/21/2004 2:10:41 PM | | No | 131.75328 |
| WESTMORELAND | 1/18/2005 1:43:17 PM | break segment (1400-1412) and (1300-131 | No | 311.37298 |
| JOSHUA | 1/18/2005 1:33:08 PM | | No | 175.99324 |
| CHARLES | 12/1/2004 3:06:38 PM | | No | 823.65390 |

Record: 17 Show: All Selected Records (0 out of *2000 Selected) Options

Over the course of three months, two or three interns at a time worked on the same versioned street dataset, altering the range data as well as adding notations concerning spatial corrections that needed to be made. These interns, who were Pittsburgh Public Schools seniors, gained valuable experience in GIS technologies and did an excellent job of diligently poring over the paper maps and making the necessary corrections and notations.

The error correction process involved three phases of review. At the first phase, simple errors in numbering were corrected and separated from those segments which required more advanced spatial edits. At the second phase, advanced edits were performed by Mr. Temple using the segment by segment recommendations of the interns. At the third phase, the corrected dataset was used to create a geocoding service. Dr. Regan and Mr. Temple then ran a complete database file of known City of Pittsburgh postal addresses provided by the United States Postal Service against the geocoding service to uncover address ranges that did not receive matches. Checking these “no match” ranges with the City’s shapefile of building footprints gave a good indication of which streets needed re-editing.

## Conclusions

At the conclusion of the project, Visual Information Systems Center and Pittsburgh Public Schools were in possession of the most accurate street range dataset ever created for the City of Pittsburgh. At an informal meeting covered by the Pittsburgh Post-Gazette newspaper, Dr. Regan and Mr. Temple presented the City of Pittsburgh municipal authorities with the completed file free of cost.

The map below shows the approximate area displayed in the section on Pittsburgh Street addressing problems. The map in that section shows the addresses geocoded in clumps on the streets. As the map below with the corrected street address scheme demonstrates, the geocoded postal addresses correspond much more closely with the actual building locations.

Over 20,000 distinct street segments were reviewed and edited, with over 200 new segments added to the dataset. Numeric, spatial, or naming attributes were changed in virtually all of the 20,000 plus records associated with the layer. The accuracy of the edits was tested against hi-resolution aerial photographs and tax assessment parcel layers to insure that the data was ready for dissemination.

For Pittsburgh Public Schools the value in having high quality street data and accurate student address data points is twofold. The first is redistricting. In the last five years Pittsburgh Public Schools has decreased in enrollment from 36,000 students in 1999-2000 to around 31,000 in 2004-2005. Overall, the total population of the city has decreased by 50% in the last 50 years. This ongoing situation has left Pittsburgh Public Schools with about 26% of its educational capacity unfilled. The cost of maintaining half empty buildings is too great for the district and the taxpayers to bear. Subsequently, school closings are an inevitable part of prudent long term district management. To accomplish the goals of closing schools with a minimum of disruption to the fabric of the community and an eye towards long term planning, GIS was utilized in the information dissemination and decision making processes concerning which schools to close and the subsequent placement of the affected students for the 2004-2005 school year. The second area where accurate street range data has value to Pittsburgh Public Schools is in performance and demographic analysis. GIS and spatially located student data plays a crucial role in the ongoing research project known as VIPER (Visualizing Information for Public Educational Research). Project VIPER is a collaboration of the Pittsburgh Public Schools Office of Information and Technology (OIT) and The University of Pittsburgh’s Visual Information Systems Center (VISC). The aim of the project is to develop, test, and validate a new generation of software for visualizing student data. VIPER allows end users to create dynamic visualizations of data including multi-dimensional, spatial analyses and GIS-based representations. The visualizations are viewable at varying degrees of granularity and will paint robust pictures of student achievement. Project VIPER enables practitioners and researchers to map the relationships between the myriad variables affecting student learning.

For the University of Pittsburgh, the value in having high quality street range data is in how it can be used to support research projects in areas ranging from urban planning to psychology to homeland security. The ability to precisely locate addresses spatially gives researchers confidence in making conclusions about local trends in their respective areas of study.

High quality geo-referenced addresses are of vital importance in a wide variety of applications. One often thinks of GIS as being focused in the areas of public policy and land management. However, the private sector also can benefit from the availability of this data. The marketing and transportation of products can be accomplished with greater efficiency and at a lower cost as local businesses begin to realize the value of this new information.

This data will be of invaluable use to emergency responders and urban planners. Before this data was available, fire, police and EMS personnel in the City of Pittsburgh often had to rely on the geographic knowledge of veteran employees to quickly and accurately locate incidents. With the City's ongoing budgetary difficulties, these highly paid individuals are being encouraged to retire at an alarming rate with without being replaced. Armed with this new information, the City's first responders can better ensure the safety of Pittsburgh's citizens now and in the future.

## References

ESRI. <u>ArcGIS 9.0: Geocoding in ArcGIS.</u> Redlands: ESRI Press, 2004

Regan, Robert. <u>The Steps of Pittsburgh</u>. Pittsburgh: Local History Company, 2004