City of Philadelphia Water Department: From Paper to the Geometric Network

Track: Water, Wastewater, and Storm Water

Author(s): Roxanne Gregorio, Ed Roworth, Larry Szarek, Rami Raad

Abstract

The City of Philadelphia Water Department (PWD) is in the process of completing one of the largest data conversion projects in the industry. The project included development of three citywide GIS coverages for over 6,600 miles of water, sewer, and high pressure fire infrastructure. To support the long term maintenance of this data, PWD and CH2M HILL recently migrated these coverages to a multi-user enterprise geodatabase. Geometric Network object models were created for each of the five systems to maintain the highest possible data integrity and capture engineering principles within the database. These models, in combination with SQL server and ESRI ArcSDE, provide PWD efficient data access, multi-user editing, versioning, conflict resolution, and integration with legacy databases. This paper provides an overview of the design methods used during the migration process, lessons learned, and the present and future benefits associated with this GIS.

Introduction

The mission of the Philadelphia Water Department (PWD) and Water Revenue Bureau is to serve the greater Philadelphia region by providing integrated water, wastewater, and stormwater services. The utility's primary mission is to plan for, operate, and maintain both the infrastructure and the organization necessary to supply high-quality drinking water; to provide an adequate and reliable water supply for all household, commercial, and community needs; and to sustain and enhance the region's watersheds and quality of life by managing wastewater and stormwater effectively.

One important way in which PWD is fulfilling its mission is through the Geographic Information System (GIS) Data Conversion project for water and sewer infrastructure. GIS technology is an increasingly valuable tool for improving overall facility maintenance and strategic utility decision-making. The main objective of this project is supporting the Philadelphia Water Department's needs to fully establish an electronic information management system. From 2002 until 2005, The Data Conversion Team compiled a broad spectrum of engineering data for over 6,620 miles of water, wastewater, stormwater, and high pressure fire infrastructure into citywide ESRI ArcInfo coverages.

For a detailed description of the phases and tasks involved in the PWD Data Conversion process that resulted into these GIS citywide coverages, please see the companion to this paper "GIS DATA CONVERSION PROCESS FOR WATER AND SEWER INFRASTRUCTURES WITH APPLICATION FOR LINKING LEGACY DATA " Gregorio, Raad, Peluso (ESRI1551) ESRI UC2005.

About midway through the Data Conversion Project, the PWD GIS Data Conversion team began discussing whether the final delivery of seamless ARC INFO coverages within the City boundary was really the best outcome of this three year-long Data Conversion Project. At this stage, ESRI geodatabases had matured to a point where they were no longer "bleeding-edge" technology. Also, The PWD GIS Manager and the Senior PWD GIS Specialist had recently attended an ESRI Class on geodatabases which touched on the geometric network. Based on the classroom discussion and some further investigation, it was decided to adjust the work plan to allow a final delivery of a geometric network to PWD in lieu of ESRI ARC INFO coverages. The reasons for leaving the "tried-and-true" coverage world for the world of the geometric network consist of the following:

- Better organization of feature classes
- Domain tables and subtypes for editing sessions
- Extensive rules and ability to validate features based on rules
- Multi-user editing environment with versions; post, and reconcile functions.

<u>Better Organization of Features</u>

The coverage deliveries for this project were divided into three main categories: Water, Sewer, and High Pressure Fire Service. For each of these categories were delivered arcs, nodes and points in coverage format. This format, though well-established and topologically correct, is in itself "disorganized" because features are lumped into groups of arcs, nodes and points. In reality, PWD manages a Raw Water Network (non potable), a Water Network (potable), a Waste Network (sanitary and combined systems), a Storm Network, and a HPFS Network (High pressure fire service). [As of 2005, the HPFS network has been decommissioned, but the GIS unit will continue to serve this data to its users.] The geodatabase with feature classes and networks is a much more organized data model which allows specific features to be a subtype of a feature class which participates in the network. For example, water valves and tees are subtypes of the water fittings feature class, and are not all lumped together into water nodes as in the water coverage, but participate as feature classes with subtypes in the Water Network. Additionally, manholes, inlets and wyes are subtypes of a feature classes that reside in the Waste Network and so on. All the captured data, by processing it into a geometric network are able to be organized into their own feature classes and subtypes based on their function. Thus the groupings are intelligent.

Upon entering into the Geometric network, the terminology of arc, node and point vanishes and is replaced by Edges and Junctions. All Edges must have a coincident junction at their end point. Also, Edges need not be broken by various Junctions along their length. Why should the addition of a tee on a water main line cause the pipe to be broken and the unique ID of that pipe to be changed? The breaking of arcs caused by adding nodes in the coverage format causes problems when other data is tied to the unique IDs of the arc. Breaking the arcs constantly requires new unique IDs to be generated and pre-existing links to other data becomes orphaned and must be re-linked.

When a small GIS group like PWD is managing 1.35 million features, the geometric network organizational structure becomes very appealing for all the reasons just described.

Domain Tables and Subtypes for Editing Sessions

By properly employing domain tables and subtypes in a geometric network, editing sessions become streamlined, efficient and almost foolproof. For example Water mainlines may have an attribute called PipeShape which connects to a domain table of the acceptable shapes for Water mainlines. In the editing session, the technician should only pick from the drop down list for that attribute, thereby limiting the possibility of user error by free-hand typing. PWD employs many domain tables for its 5 Networks.

Subtypes were heavily utilized for many of the feature classes within each network. For example, under the Water Main Lines, the subtypes which exist are Mainline, Supply Connection, Cross Connection, Round Connection, Shaft, and Vertical Offset. Default subtypes can be chosen which in this case is "Mainline." Once again, this type of organization allows technicians to edit in an efficient manner with built in devices for reducing errors.

Extensive Rules and Ability to Validate Features Based on Rules

One of the major strengths of the geometric network is the ability to assign an extensive rule set to the feature subtypes within the network. Aside from the intelligent groupings described earlier, extending this intelligence by applying rules to the connection behavior of the features increases the intelligence of the database. For example, a water fitting Tee must connect to three pipes, a water fitting Cross must connect to four pipes. Thus, when a technician adds new features to the network, upon validation, violation of any connectivity rules for the Edges and Junctions becomes apparent and changes can be made by the editor. An additional feature of the rule set is the ability to set default junction types for edges. For example, a hydrant line will automatically connect to a main line through a Tee fitting and will end in a Fire Hydrant. This happens as soon as the GIS technician places the hydrant line. They do not even have to add the tee or the hydrant as their placement is automatic.

Multi-User Editing Environment with Versions; Post and Reconcile Functions

A lot has been mentioned about the ease and efficiency in editing sessions when utilizing a geometric network. What has not been highlighted is the fact that utilizing a geometric network within an enterprise geodatabase allows multi-user editing sessions on the dataset. All PWD networks reside in

a SQL Server 2000 Database using ArcSDE 9.0 as the geospatial translation engine. If PWD had stayed the course with a final delivery of coverages, these flat file formats would not have allowed multi-user editing. PWD GIS unit employs three full time GIS technicians, a senior level GIS specialist, a programmer and a manager and is expanding. All staff members can use or edit the data; the data can be served in a variety of web applications simultaneously, perused by end users in other ESRI software, and processed by other applications which connect to it.

For these reasons described above, it would almost have been considered an error for PWD to continue down the path of final ESRI coverage deliverables. The effort it would have taken to build applications around the coverages to mimic the rule validation, safety nets, and multi user editing environment already present intrinsically within a geometric network would have constituted a whole other project within itself.   The PWD Data Conversion team was able to make the needed adaptations to the work plan, PWD was able to finance the additional hardware for a more than adequate Server, and ESRI was able to deliver the technology to make the Five PWD Geometric Networks a reality.

The following sections provide an overview of the design methods used during the migration process to the Geometric network from the coverages, lessons learned, and the present and future benefits associated with this GIS Data Conversion Project.

Geodatabase Design and Development

PWD, ESRI and CH2M HILL held a series of design workshops to develop the conceptual, logical, and physical geometric network models for five PWD systems. Models for Waste, Storm, Raw Water, Water, and High Pressure Fire systems were designed using Unified Modeling Language (UML) case tools and MS Visio. ESRI provided a great starting point for these and similar systems in their example Visio UML models. Due to the complexity of the City's systems, and the detail of the data dictionary for the coverages that were to be migrated, PWD decided it was best to create their own custom models not based on the ESRI examples.

Using the coverage data dictionary as a starting point, each coverage contained an attribute for the type of feature represented. These point, node, and arc features with the equivalent of subtypes where then individually examined and grouped into similar feature classes. Feature classes fell into four main categories: those that were to take part in the geometric network such as Junctions and Edges, and those that would represent simple point or line feature classes—not taking part in the network. The non-networked features were decided upon based on the concept of using structures through which water or waste flowed as being part of the Geometric Network, i.e. valves. Items such as casings for water lines were placed into the non-networked feature category as there is no flow through the casing which protects the water line. Common attributes within the coverage's were than targeted for inclusion within four abstract classes one for each of these types of feature classes.

The conceptual phase of the design process was aided by creating an Excel spreadsheet or migration matrix of the proposed abstract classes, network junctions and edges, point and line feature classes for each system. Each worksheet within the spreadsheet represented one of the above objects with existing coverage attributes mapped to the geodatabase field names. For each feature class, the existing coverage(s) and subtype(s) to be included in the data migration was identified. SQL queries for loading the coverage data were included within the worksheets because some of the feature classes had more than one source of coverage data to include because of the grouping of subtypes. Once the initial migration matrix was prepared and all coverage subtypes were accounted for, the process of designing the logical UML model began. The result of the conceptual phase of the design were five spreadsheets listing all the proposed abstract classes, network junctions and edges, point and line feature classes, valid values (domains), and the source coverage for each.

The logical design of the geodatabase proceeded by using the migration matrices as a guide for all the types of objects that needed to be created in the UML. MS Visio is a powerful tool for the design of the geodatabase primarily because of class inheritance (the parent child relationships) maintained within the UML. The first step in the UML design process was to create "parent" geometric network objects. Then edge and junction classes that were children of the network were defined. At this level, also included were the abstract classes for edge and junction features. Now, the framework for adding specific edge and junction features and the subtypes inheriting from the abstract class and the

network existed.   A similar process was followed to add the non-network point and line feature classes with the exception of not having these feature classes take part in the parent network inheritance.

Domain tables are important during the data validation stage and for maintaining valid values for the feature classes. Valid values for our coverage attributes were easily used to populate range and coded value domains within Visio. Each network feature class contains domain tables that are specific to certain attribute fields. For feature class subtypes that had multiple fields with domain tables, it was learned that a default value for the field needed to be defined or else the domain tables became unlinked during conversion to a geodatabase (the physical design process). UML design can be very tricky when using multiple subtypes within a feature class with subtypes having their own attribute fields and domain tables.

Once the logical models were fairly complete, the physical design process consisted of using ESRI's XMI export and validation macro's within MS Visio and then using Arc Catalog to apply the XML document to a blank personal geodatabase (pgdb).   This was an iterative process where errors in the UML were fixed when found and the error checking process restarted. XML was exported from Visio and applied to a pgdb. The schema of the pgdb was than thoroughly checked by producing a report using ESRI's Geodatabase Reporter tool. Not all errors were found by using the validation macros within Visio such as the one described for subtypes and domain tables.   So it was necessary to check the physical database schema before being confident it reflected the logical UML design.

Geodatabase Preparation and Data Loading

The first step in the coverage to feature class conversion process was to create a set of intermediate feature classes that could easily be queried and manipulated to load subsets of features into the target geodatabase feature classes and subtypes (**Figure 1**). Preparation of the coverages included joining attribute information stored in related MS Access tables, renaming coverage field names to the final geodatabase field names, and changing field type definitions to prepare them for linkage to range and coded value domains. For example, in the geodatabase, subtypes are stored as integer values while in the coverages, subtypes are stored as strings. This intermediate step was the key to being able to proceed to the data loading stage for the 1.35 million features.
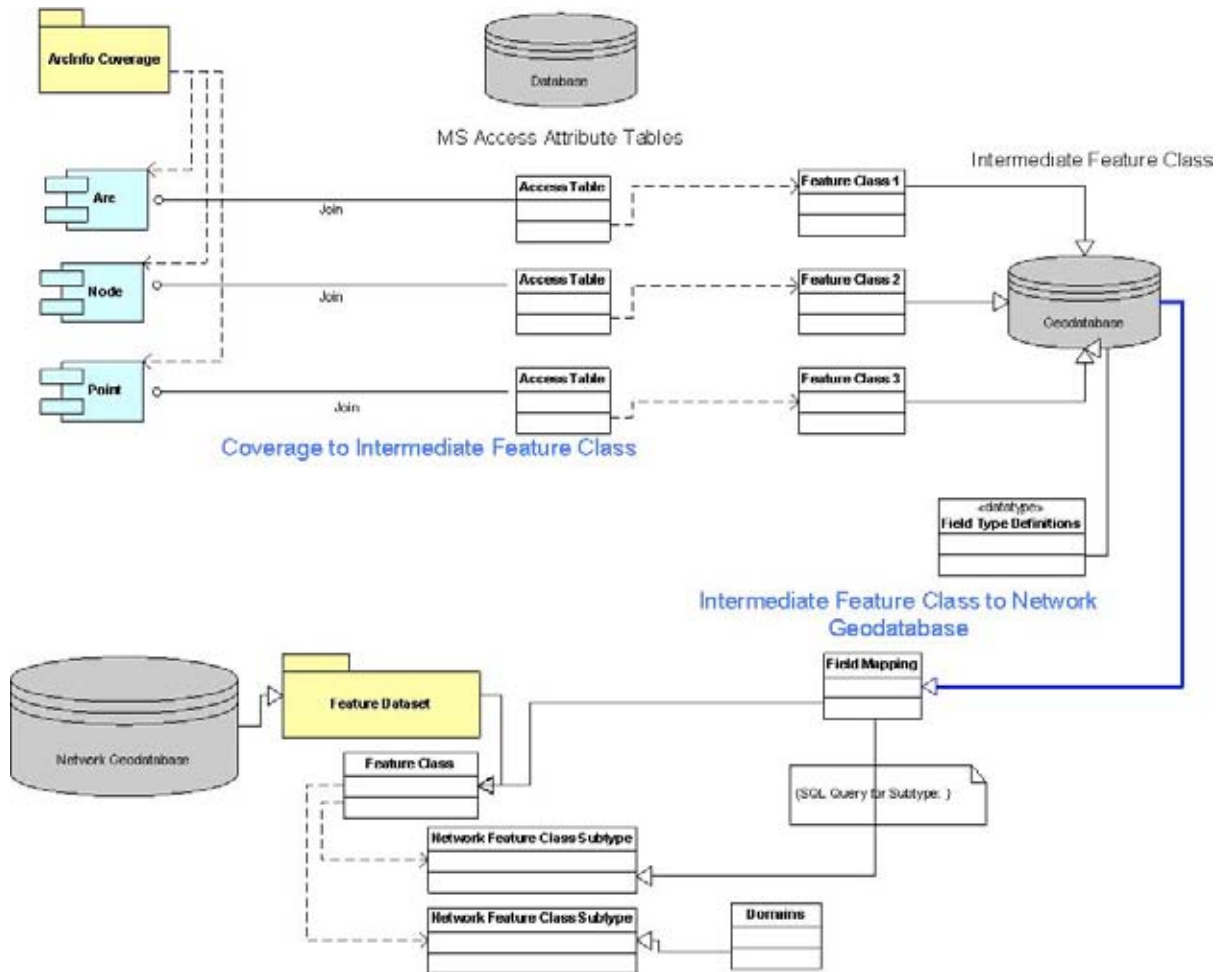
Figure 1. PWD's coverage to Geodatabase feature class conversion process.

Before data loading of the intermediate feature classes could commence, the various "out of the box tools" for loading data within ArcCatalog were explored. The Simple Data loader within ArcCatalog was used to test the data loading process on a personal geodatabase built from the UML models. This tool proved to be very time consuming because of the wizard like step by step dialogues that needed to be filled out and checked repeatedly. Simply needed was a tool that would allow the entry of a SQL query to filter the intermediate data and append the result to the target feature class. ESRI's Arctoolbox ModelBuilder allowed this flexibility.

ModelBuilder models are created in a visual editing window where functions or components of the model are dragged into the editing window by selecting the appropriate tools from ArcToolBox. This flexibility allowed the Data Conversion Team to piece together and set the various tool parameters and variables. In **Figure 2**, it is shown how the simple data loader makes use of the Make Feature Layer and Append tool. The Make Feature Layer tool allowed the entry of a SQL expression to filter the data and stored this temporary subset in memory. Then this input was used with the Append tool which allowed the option of selecting an output target feature class. The key to using this simple Model was the fact that the Append command **(Figure 3)** checks the input schema of the intermediate feature class and if the field names match the Append command appends the features with attribute values. The secret to this command is that the schemas do not have to be a 1 to 1 match in total number of fields; just field names and values that match the output are appended. Luckily, most of the geodatabase feature classes had their own set of fields and many of the geodatabase feature classes had more or less fields than the original coverages' fields. Once this simple data loader was

created, the five UML schema's to the as yet to be created SQL server / SDE based geodatabase were able to be applied.

PWD Dataloading Template - Lines
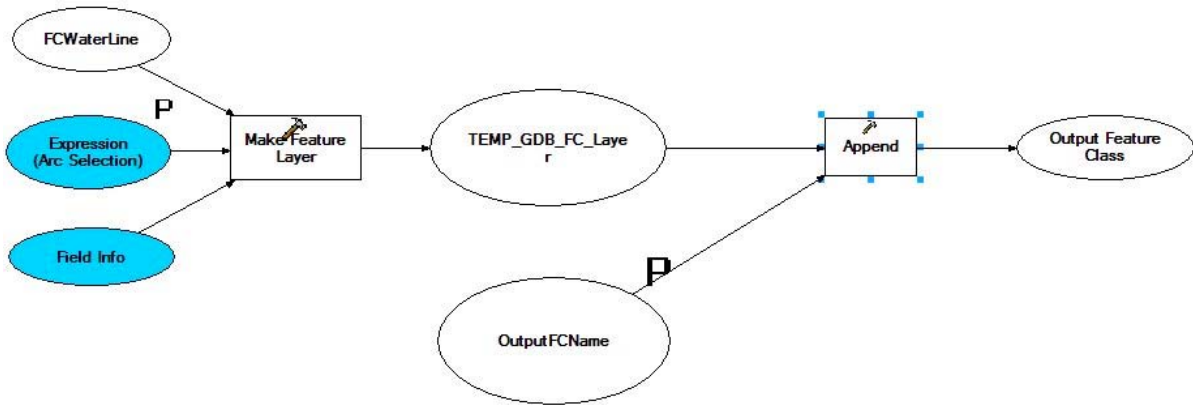
Load Water Lines

Figure 2. ESRI's ModelBuilder was used to create a customized Simple Data Loader tool.
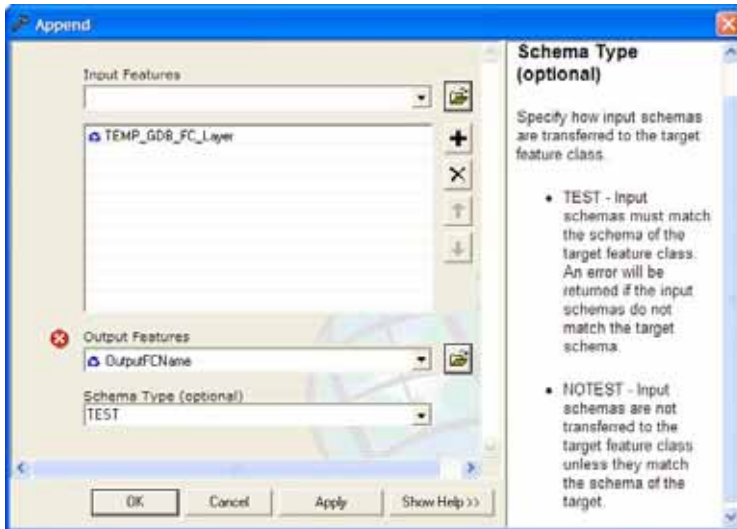
Figure 3. The Append command within ArcToolbox contains the option to Test input schema's to match target output schemas.

Once a method for data loading was created, the UML models were applied to the SQL server database. This step worked as expected after thorough testing on a pgdb. However it was found that some domain tables were named identically to other domain tables in different UML models and, therefore, domain tables where overwritten if they already existed in the geodatabase. This is an important caveat to designers of UML models. This error was fixed by merging some domain table values together and renaming others.

Then all of the coverage data from the intermediate feature classes were loaded. Each SQL query written while preparing the migration matrix was performed on the intermediate feature classes. The number of features returned from the queries was totaled and the total was carefully tallied to make sure all of the data was migrated. In this way an expected feature count was maintained which was later verified while loading data into the SQL server geodatabase. This is an important check in finding missing data or rather data that was not migrated. Each system took on average 2 days to load and check with a two person team verifying each step.

Database Design

PWD uses Microsoft SQL Server and ESRI's Arc SDE 9.0 in a multiple geodatabase configuration where SDE system tables are stored within each geodatabase. GIS data for the five networks discussed in this paper are contained within a single geodatabase. Other city-wide planimetric features and imagery are stored in their own geodatabase. This compartmentalizing of data allows fine control of user and group permissions needed to police read and write access to the features. Database design for an enterprise system can be the topic of another paper so it will not be covered here. However, the most important aspect of the database design was to have solid plan for maintaining data integrity once it was loaded. Versioning of the feature classes was immediately implemented upon completion of the data loading. In addition a database backup procedure was put in place that started the first day of data-loading. PWD has implemented a surrogate default versioning scheme as documented on page 4 of ESRI's "Versioning Workflows" technical paper (www.esri.com Jan 2004) (**Figure 4.).**

Figure 4. SDE versioning scheme.
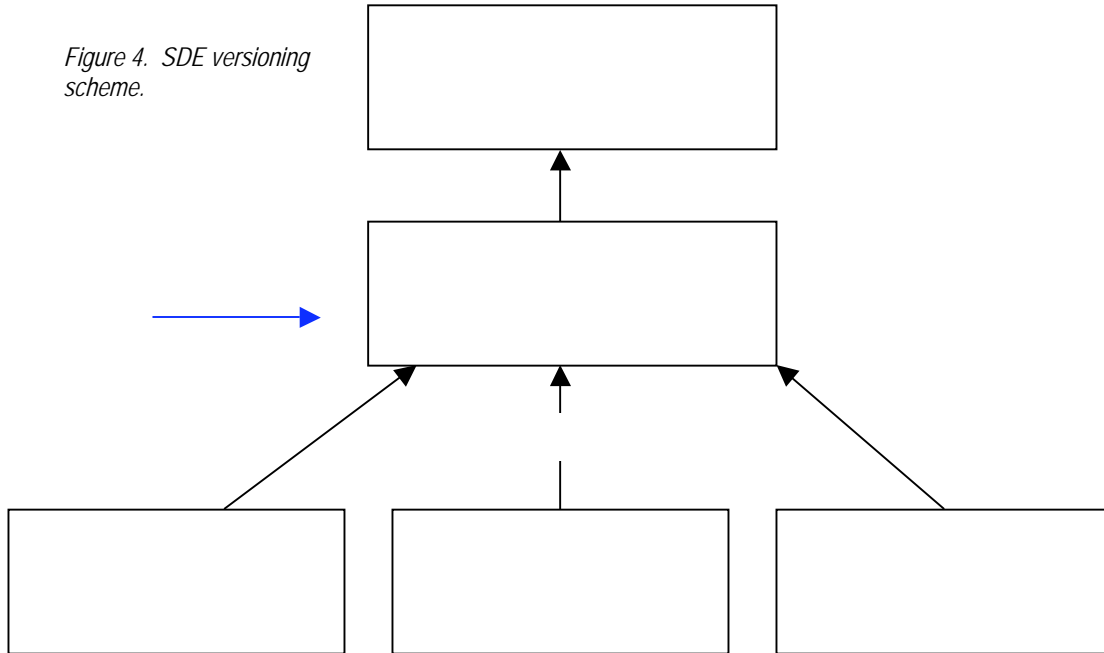
The benefit of versioning becomes apparent when PWD technicians are maintaining a section of a network. They can create their own version of the data and validate it. When they are satisfied with their edits they can reconcile their version to the editing default version, a sort of safety net. This editing default version can be reconciled and posted to the default dataset by the lead GIS specialist. The default version is considered the active dataset which end users may access via ESRI's ArcMap Desktop and various web applications. This concept of versioning allows these various safety nets to be set up so that validation and reconciliation of data can take place before making that final leap to the default version of the dataset.

Geometric Networks

Using the Geometric Network wizard within ArcCatalog requires some advance knowledge of how database precision and snapping tolerance interact when loading data into a geodatabase. Upon loading the data into the geodatabase it was discovered that some of the data had shifted by an extremely small amount in an inconsistent manner. In most cases this shift was less than .00006 feet. However this shift was enough to break geometric connectivity if no snapping was used upon building the geometric network. This was odd behavior since the data had originally resided in coverage format with topology. It was important then to create the geometric network without snapping, since once the data was snapped and moved, it could not be unsnapped.

Once the Geometric Network was created, network junctions were automatically placed where there was a break in connectivity. After a period of trial and error on a pgdb, a snapping tolerance of .0007 feet was chosen. This trial and error of the snapping tolerance consisted of reducing the number of network junctions automatically created while not creating invalid connectivity. In addition, only the network junctions were allowed to move during snapping. The final snapping tolerance was then applied against the SQL server geodatabase. The lesson to share here is to not snap data when defining a geometric network for the first time, but instead to make a pgdb copy of it and than adjust snap tolerances as needed.

Network Validation

Before editing of network junctions and fixing of connectivity issues could begin, an understanding of validation needed to be developed. There are several short-comings to the existing Arcmap editing

"Validate Features" tool. These short-comings are related to the general approach ESRI uses to validate features and the intended use of this "out of the box" tool.

The general approach ESRI uses for validating features is that the validation process should not result in valid features being flagged as invalid (false negatives); it is, however, allowable to have features that are invalid being reported as valid (false positives). If the geodatabase did not enforce any validation, every feature would effectively be valid. When performing validation on a particular feature, the validation process occurs in three steps (there are additional steps but these are the ones that PWD employed):

1. Validate the subtype (if applicable)

2. Validate the attribute values (domains).

3. Validate the network connectivity rules (if a network feature).

This strategy means the least extensive validation is performed first. The validation process stops once a feature is found to be invalid. So, for example, if a feature fails the validation test for check number 1, then checks 2 and 3 are never executed. When checking connectivity (rule number 3) all associated rules must be valid. With network connectivity rules, a user does not have the ability to specify only one connectivity rule, the user has to specify them all at once. Thus, if a type of connectivity exists that does not have an associated connectivity rule, the network feature is flagged as invalid.

The validation tool and approach that ESRI includes out of the box is not intended to validate features in a data loading scenario. For example, the first occurrence of an invalid attribute (2) or rule violation (3) will stop the validation process. The feature with the violation is automatically selected and one of two things may occur. First, if this is the only violation in the set of tested features, the message returned may indicate an invalid attribute value (checks against domain tables) or it may return a description of the network connectivity rule that's been violated. This message is returned in a popup dialogue box. Second, if more than one rule is violated in the set of tested features, then all records with a violation are selected, but no feedback for the type of violation is provided. This limitation necessitates that each invalid feature be selected again and tested individually to find the specific error message.

The validation limitations had to be addressed before the correction of the network junctions could begin. Available VBA macros were a great aid in attribute validation that helped identify invalid or missing domain values. A list of these types of attribute or domain errors was created and fixed within SQL server by performing update queries. Once these domain errors were corrected, editing and correcting of the network junctions proceeded.

Network Connectivity

Each geometric network has its own set of connectivity rules. The connectivity rules for the five networks were created by a team of individuals who have knowledge in either GIS or the engineering practices of the department. The Connectivity Rule Team consisted of a GIS analyst familiar with the GIS network, a water wastewater engineer, and a group of field personnel. The combined knowledge of the field personnel, more specifically their experience with the engineering practices of these digital GIS features, was captured in the resultant connectivity rules.

When creating our set of connectivity rules, two data type scenarios had to be addressed. The first, Edge-Edge rules, determines what type of pipes may connect to each other and through what type of junction they can connect. For example, a hydrant line (edge) can connect to a main line (edge) thru a tee (junction) or a valve (junction). When this rule is implemented in the network, any other junction connecting these two types of pipes will produce a validation error. The second type of rule that was input was an Edge-Junction rule. This second rule-type is defined by two types of connectivity: firstly, what type and how many junctions can be connected to an edge, and secondly how many edges can connect to a given junction. For example, an Edge-Junction rule that was created stated that a valve can connect to a mainline pipe, and that valve can have up to three mainline pipes connected to it, and that mainline pipe can have up to two valves connected to it. In other words, it is

a valid situation if one has three mainline pipes connecting at a valve or if one has a valve on each end of a mainline pipe.

To aid in creating the rules, it was found helpful to make two spreadsheet workbooks for each network. This is referred to as a "Connectivity Matrix". The first workbook contained the structure of the Edge-Edge rules and the second contained the same for the Edge-Junction Rules. These workbooks were designed so they contained every possible combination of subtype for Edge-Edge (**Figure 5**) and Edge-Junction feature types (**Figure 6).**

Figure 5. Portion of an Edge-Edge Rules Connectivity Matrix.

Subtype1	Subtype2	TEEFLD, TEEMAN Tee	ATTRIBU, FLARE,STEP,CASEN AttributeChange	BULK Bulkhead	COLLAR Collar
GravityMain	GravityMain	X	X		X
	GravityDryWeatherOutlet	X	X		X
	GravityInterceptor	X	X		X
	GravityRelief	X	X		X
	GravitySiphonInterceptor	X	X		X
	wwAquaduct				
	wwInletPipes				
	wwVentPipes				
	ForceMain	X			
	ForceInterceptor	X			
	wwOffsetAccesss				
	wwVirtualLink				

Figure 6. Portion of an Edge-Junction Rules Connectivity Matrix.

Edge	Subtype	TEEFLD, TEEMAN Tee	ATTRIBU, FLARE,STEP,CASEN AttributeChange	BULK Bulkhead	COLLAR Collar
stGravityMain	GravityMain	03,02	02,02	02,02	02,02
	GravityStormwaterOutfall	03,02	02,02	02,02	02,02
stInletPipe	stInletPipes		02,02	02,02	02,02
stVentPipe	stVentPipes		02,02	02,02	02,02

Once these matrices were created, it was more easily apparent to see that every possible subtype to subtype connectivity scenario was addressed. These matrices were also useful to the GIS technician when it came time to enter the rules into the Geometric Network Properties in ArcCatalog. This allowed the technician to stay organized during a very tedious data entry process.

Once a complete set of connectivity rules was entered into the geometric network, it was time to validate the network. When dealing with this extremely large network, it was helpful to break the network into smaller sections, and then validate each piece separately. Also, the Arcmap "Validate Command" did not respond well to validating the entire network at once. The process found to work best was to use a polygon index layer to break the network up into manageable selection sets. An Arctoolbox ModelBuilder tool was used to perform the spatial selections based on a polygon index value; the selections were filtered to remove abandoned features based on an attribute. Once a final selection set was created, the Arcmap "Validate Command" was employed. To address the shortcomings of the Arcmap "Validate Command" and the error reporting problem which was described earlier, a custom validation tool using ArcObjects was created that reported validation errors in batch to a csv file. The csv file for each validation run contained the feature class name, the Object ID, the actual error description as would be reported from the Arcmap "Validate Command", and some other useful attribute information for each invalid feature.

Once all the validation spreadsheets were appended together into one large spreadsheet of validation errors for a Network, each error was investigated. The resolution of these errors usually fell into one of three categories:

Firstly, an edit was needed. In this case a feature could have been attributed wrong (subtype), or in some instances the geometry needed adjusting. For example, numerous errors were generated due to a subtype being incorrectly attributed to a feature. In this case the error was resolved by simply changing the subtype to the correct value.

Secondly, a new rule was created as it was overlooked initially. An example of this type of error is when hundreds of occurrences of the same validation error occurred. For example, a specific type of pipe was invalid when connected to another type of pipe. After reviewing the situation with the Connectivity Rules team, it was determined that this was actually a valid connection and a new rule was required. After entering the rule into the network, these errors were resolved and the Team could continue working through the rest of the issues.

Lastly, a change was required in a domain table. This case occurred when a value entered into a field did not appear in the domain table. Sometimes the initially invalid value was added to the domain table, or the initially invalid value was changed to a value that was already present and valid in the domain table.

When all of the errors had been addressed and validation was run without any errors being returned (or only known errors were returned that reflect engineering anomalies to be changed at a later time by field personnel), the Team was confident that a very accurate set of data now existed. Validating the network in this manner is as close as the Team could come to a 100% QA/QC process, since every single feature in the five network s (over a million features) has been checked against a strict set of connectivity and attribute rules.

Associated Applications

Largely in part to its attribute and connectivity accuracy and the painstaking methodology by which it was assembled, it is easy to visualize the PWD GIS dataset becoming the data hub of the PWD organization.   As the data represents the various PWD facilities above and below the street level, it can easily extend its reach to be incorporated with the multitude of planning applications and inventory management systems which currently exist within PWD.

       Planning Applications

- Leak Detection System
- CAPIT (Capital Management System)
- SAP-DSS (Sewer Assessment Project-Decision Support System)
- Defective Lateral System
- Hydraulic Water Model

       Inventory Management

- Legacy Work Order Management Systems

These Planning and Inventory Management Applications report on various aspects of the data which can now be represented in one of the 5 networks. Until now, these applications have used the Citywide Streets centerline data to provide a graphical link in each application if any graphical link exists at all. However, a street centerline is no replacement for the actual PWD feature that is being reported on in an application. While personnel need to locate their projects via a street centerline, i.e., along street, from street, and to street location, they are not reporting on the streets' characteristics themselves but on the actual water and sewer features involved in their work. It follows that with the existence of the PWD GIS dataset, their applications will need to be shifted to connect to the actual water and sewer assets via a unique GIS FacilityID. This constitutes a large paradigm shift within the PWD organization. For over a hundred years the actual PWD facilities have almost been masked by their along, from, and to street locations. Now this positional description can be joined with actual

PWD asset attribute and geospatial information to provide a more robust application environment for these planning and inventory management applications. While the Legacy Work Order Management Systems, the Hydraulic Model and SAP are currently linked to the new GIS data, the remaining applications will require some re-work to get them to link directly into the PWD GIS datasets. Also, the Legacy Work order management system (WMS) will be migrated into a new WMS which will link more efficiently to the GIS datasets.

Maintenance of the PWD GIS data is of utmost importance to the PWD Data Conversion team and Geometric Network Team. Maintenance was really one of the main reasons for transitioning to a geometric network in the first place. An investment as significant as this Data Conversion Project simply cannot be allowed to go out of synch with "what is happening on the street."   A great deal of time and effort has been invested in creating maintenance procedures for both large edits for new contracts (Capital Edits) and minor plumbing edits (Non-Capital Edits). These procedures involve various tools created within the Arcmap Environment as well as some asp.net web application tools for use by the PWD GIS Staff in finding their work.   These are descriptions of the various tools which were created to manage such a large system of data:

Dimension Line Tool -- This tool allows GIS Technicians to create construction lines and construction circles to precisely place new features adjacent to Citywide curblines. This tool allows persistent and precisely drawn (length, angle, perpendicular, parallel) circles or lines to be drawn to accurately place system features in a "snapping" environment.

Abstract Class Attribution Form -- The abstract class attribute form allows the GIS technician to enter in attribute values once and apply these values across feature classes instead of having to enter in the values multiple times by feature class.

Unique Facility Identifier -- This SQL server code allows the GIS systems administrator to populate and persist a custom global unique identifier without having to manually populate this ID by feature class.   The unique facility ID is used to individually identify each asset within the GIS network and link to the maintenance management system.

With these tools and a dedicated in-house GIS staff, it is assured that the GIS dataset will not become obsolete due to lack of maintenance.

Conclusion

The PWD GIS Dataset is a highly sophisticated dataset in that it was drawn piece by piece from the as built drawings that describe, in a block by block fashion, the entire water, storm, hpfs, and waste water system for Philadelphia. For the first time in the history of PWD there is a seamless dataset available that represents the entire city's water and wastewater system. First phase usage of the dataset revolves around maintenance and basic reporting on the features and their attributes. Maintenance all occurs within the ESRI environment which has been discussed in the previous section. Basic reporting functionality on the dataset will be accomplished through a thin client web application. This will allow a large number of managerial and field personnel to review pre-set reports on the data and to visually inspect the data by navigating via the streets of Philadelphia.   This will satisfy the immediate needs of the current GIS user population. Then the needs of the current applications which link or will link to PWD GIS data must be met. The Planning and Inventory Management Applications discussed in the previous section must be easily integrated with the new GIS dataset. SAP-DSS and the Hydraulic Water Model currently have a process in place for both using the GIS data in real time and on reporting back any anomalies that are found during their usage.   The other applications mentioned will require some intervention in integrating them with the current GIS dataset. However, the most important integration of the GIS dataset with an application must be with the new Work Order Management System. Linking the GIS Spatial and attribute data with the work order histories of the data will be an extensive set of data upon which many types of reporting can happen to assist in proactive maintenance of the Philadelphia Water and Wastewater System.

Locations where the GIS dataset has less precise horizontal accuracy based on the lack of planimetric information is being compiled and projects are being devised to GPS those locations and use the coordinates to adjust the GIS data in those areas.

ESRI's Spatial Analyst Extension along with TIN data is being used to fill in the elevation information for manholes and other Waste and Storm Network features where elevation information was not found on the source documents. This one repository of asset elevation information will prove very useful to planners and analysts of the system.

As time goes on, more and more complex spatial and attribute queries will be devised to gain insight into the system, its behavior, and possible future paths for changes to it.   And continuous checking of the data against real world assets by field personnel will result in an increasingly accurate dataset. It is envisioned that monthly meetings of the GIS Steering Committee will continue to assist in prioritizing all projects that involve usage of or linkage to this new and sophisticated GIS dataset. The GIS dataset should become a central hub in the wheel of organization of the PWD in order to maintain its main mission of providing clear, clean, quality drinking water to the residents of Philadelphia.

References: Information contained in this document is from user experience and information contained in ESRI's documentation library. Some of the references listed below are potentially used but not cited in a specific instance:

Building Geodatabases II (For ArcEditor 8.3 and ArcInfo 8.3) Course Lectures. ESRI Educational Services, Patricia McGray. Course Version 1.2. Revised May 2003.

"The City of Philadelphia Water Department GIS Data Conversion Services Data Conversion WorkPlan." CH2MHILL Data Conversion Team: CH2MHILL, AGRA Baymont, Inc., Schoor DePalma, Inc., CSA North America, Inc., LAM Associates, Inc. March 27, 2002.

Designing Geodatabases: Case studies in GIS Data Modeling. David Arctur; Mike Zeiler, 2004 .

The Drinking Water Dictionary.   Water Amer, American Water Works Association, Lee Jr. Bradley, James M. Symons

Getting to Know ArcGIS Desktop.   Tim Ormsby, Eileen Napoleon, Robert Burke, Carolyn Groessl, and Laura Feaster, 2001.

Modeling Our World: The ESRI Guide to Geodatabase Design.   Michael Zeiler, 1999.

Author Information:

Roxanne Gregorio – GIS Manager, Philadelphia Water Department. Aramark Tower 1101 Market Street, Philadelphia, PA 19107 p. 215 685 6333 f. 215 685 6207 roxanne.gregorio@phila.gov

Rami Raad – Project Engineer, CH2M HILL 1700 Market St. Suite 1600, Philadelphia, PA. 19103-3916 p. 215 563 4220 f. 215 563 3828 rraad@ch2m.com

Edwin Roworth – GIS Analyst, CH2M HILL 1700 Market St. Suite 1600, Philadelphia, PA. 19103-3916 p. 215 563 4220 f. 215 563 3828 eroworth@ch2m.com

Larry Szarek- GIS Specialist III, Philadelphia Water Department. Aramark Tower 1101 Market Street, Philadelphia, PA 19107 p. 215 685 685 4924 f. 215 685 6207 larry.szarek@phila.gov