An Esri® White Paper • July 2011

GIS Data Quality Best Practices for Water, Wastewater, and Stormwater Utilities

Esri, 380 New York St., Redlands, CA 92373-8100 USA
TEL 909-793-2853 • FAX 909-793-5953 • E-MAIL info@esri.com • WEB esri.com

GIS Data Quality Best Practices for Water, Wastewater, and Stormwater Utilities

An Esri White Paper

Contents **Page**

GIS Data Quality Best Practices for Water, Wastewater, and Stormwater Utilities

The Challenge

Water, wastewater, and stormwater utilities recognize a geographic information system (GIS) as an authoritative repository of asset information that should be shared throughout the entire utility and integrated with other enterprise utility business systems. To maximize the benefit of utility-wide GIS, the data being shared must reliably meet the needs of the utility. Asset data in the GIS must be accurate positionally (in the right place), descriptively (describes the asset appropriately), and temporally (up-to-date). As a result, water, wastewater, and stormwater utilities are now heavily focusing on quality assurance (QA) and quality control (QC) to ensure that their GIS data truly meets their needs.

Why Data Quality Is Important for Water Utilities

Data Access

The evolution in GIS technology has made it easier to ubiquitously deploy geospatial data throughout all departments within a utility. Utility-wide GIS access means that more people have access to the most current data/information. Poor quality data impacts decision making and is often expensive to fix.

Data Integration

Utility-wide GIS also means the ability to integrate with other enterprise business systems to spatially enable them. It is typical for work order management, customer information, and laboratory information management systems (LIMS), as well as financial, billing, hydraulic modeling, and other water utility systems, to be integrated with GIS. This type of integration has raised the bar for water utilities to ensure data accuracy and currency.

Data Usage

The way water utilities are using GIS is also driving an increased focus on data quality. Several years ago, GIS was primarily used to produce paper maps. As long as the features looked accurate at the scale of the printed map, everyone was satisfied with the quality of the data. Now, most water utility staff are exposed to GIS via web-based and mobile applications. This allows interaction with GIS data at many different map scales and the ability to drill down for detailed information.

Data Content

Lastly, data quality is paramount when using GIS to perform spatial analysis and for reporting purposes. For example, water utilities use GIS to produce statements such as impervious surface calculation for stormwater billing and asset reports to regulatory bodies and for real-time or periodic generation of key performance indicators (KPIs) for utility management. The increased usage of GIS for reporting—especially when related to regulatory approvals, customer billing, and KPIs—drives the need for quality data.

QA/QC Overview

While it may be easy to engage staff members in cleaning up bad data, it is not so easy to find out where edits/corrections need to happen or to define workflows that allow them to swiftly make required corrections to the data. This is true for ongoing data updates made through redline submissions. To effectively and efficiently identify problems in data and make the most of data correction efforts, QA/QC procedures and tools must be introduced into workflows to prevent errors from occurring in the first place.

ArcGIS® offers a wide range of tools, extensions, and database modeling capabilities that can be used to improve data quality. This paper will explore some of the options that can be considered best practices for water utilities.

Quality Assurance Techniques

QA is the process of establishing a set of guidelines or best practices to ensure the production of quality GIS data. Performing QA has many benefits, including less data reworking, because quality requirements have previously been identified and are being measured and monitored continuously.

QA is often thought of as performing tests on data to ensure its accuracy. In actuality, this is QC. While QC is a key part of QA, quality assurance takes a step back and looks at the overall structure and workflows to implement measures that prevent the introduction of errors. The following are some useful features of ArcGIS that can provide key advantages when implementing QA:

Use a Geodatabase

Shapefiles and coverages are used mostly for data export and interchange and do not provide the appropriate GIS data repository for the needs of a water, wastewater, or stormwater utility. The geodatabase, on the other hand, is a data storage structure for ArcGIS that enables water utilities to encapsulate both data and business rules in one location. Some of these rules prevent you from making editing mistakes, such as not allowing invalid attribute values to be entered, while other rules ensure that network connectivity and relationships between connected features are maintained. The geodatabase is the receptacle that provides the ability to implement these rules; for the geodatabase to be effective, you need to properly design and maintain your data model.

Have a Properly Configured Data Model

A data model or schema is how you represent real-world features in your geodatabase. The data model is used to configure your data's structure to suit your needs. A properly configured data model, especially one that utilizes domain values wherever appropriate, minimizes the possibility of misrepresenting features in your GIS. That is, a well-designed data model that is implemented in a geodatabase prevents the user from creating invalid data. While it is impossible to design a data model that prevents all editing errors, it is possible to minimize errors by using data modeling within your organization. The Esri water, wastewater, or stormwater data models offer a good way to start, as much of the initial data modeling work has already been completed, thereby saving time and money and allowing you to utilize industry best practices.

Note: The water, wastewater, and stormwater data models are part of the local government model.

Use Geometric Networks

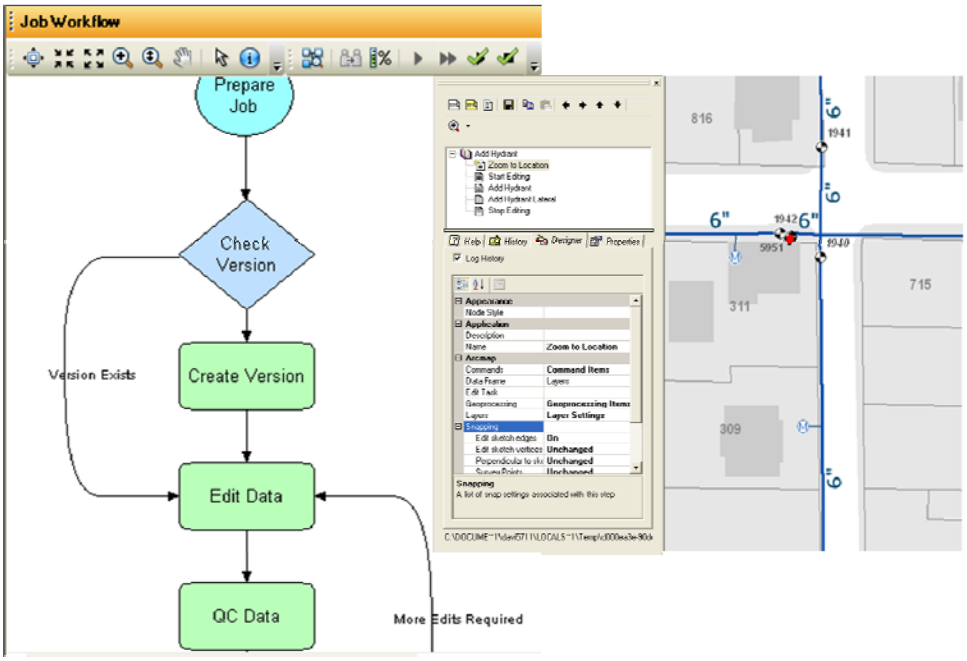
Another key feature of the geodatabase is the geometric network, which is a set of connected edges (lines) and junctions (points), along with connectivity rules that is used to represent and model the behavior of a common network infrastructure in the real world. The geometric network provides extra behavior that allows you to define how edges and junctions connect and how water, wastewater, and stormwater flow through the network. In a water network, the flow of water is from a reservoir to a customer via various appurtenances such as valves, fittings, and pump stations. Connectivity rules help define how the features in the network connect to each other. For example, a fire hydrant can only be connected to a service lateral of type hydrant; a fitting of type tee should be connected to exactly three pipes.

While a well-defined data model allows you to store information about real-world features and assets, the geometric network allows you to encapsulate how your distribution or collection systems actually operate.

More information on geometric networks is available at blogs.esri.com/Dev/blogs/waterutilities/archive/2009/11/17/Geometric-networks-for-water-utilities.aspx.

Formalize Your Workflows

The next step toward producing quality data is to standardize and formalize your organizational workflows. This is especially important when multiple editors are accessing your data. Formalization of workflows could be as simple as having clearly written and documented standard operating procedures. However, in reality, most water utilities are looking to design and distribute workflows. This can be achieved by developing custom tools suited to your organizational needs (this option would require that you update and maintain the source code for every release). Alternatively, you can utilize the out-of-the-box tools within the ArcGIS Workflow Manager extension to design, execute, track, and distribute your workflows.



Design and execute workflows.

With Workflow Manager, utility managers and supervisors can query data and generate reports to summarize what is actively being worked on, who has been assigned to complete the task, what work is behind schedule, and so on. Task Assistant Manager, a component within Workflow Manager, provides microlevel workflows that allow you to streamline tasks within ArcMap™. This is a great way to guide users through defined processes within ArcMap, execute ArcMap commands or geoprocessing tools, and set layer properties such as visibility and snapping. Using these workflows can prevent errors and help execute common processes consistently throughout the utility.

Take Advantage of Versioning

With multiple editors, using a versioned geodatabase will enable you to implement an editing workflow that can allow you to identify and resolve potential data issues, such as conflicting edits, and perform quality control on recently edited data before all the GIS users and integrated systems see the edits. Versioning is highly beneficial to safeguard your data.

A typical use case for versioning is when designing proposed projects or plans for expansion. You may want to use the GIS to model the potential results of placement of pipes, for example, for a proposed construction project. However, you will want to isolate these proposals from the production database until the project is finally built and accepted. By isolating these changes to a different version, you can prevent others from performing analysis on the wrong data. Managing versions in a large utility to ensure that data is published at the correct times can be challenging. ArcGIS Workflow Manager allows you to automate version management to avoid having to manage them manually. Instead of using ArcCatalog™ to administer job versions, you can simply allow the Workflow Manager workflow to create, reconcile and post, and delete the versions. With Workflow Manager, you can take advantage of versioning without the typical overhead associated with version management.

Use Geoprocessing Models and Scripts

Within a utility, if there are common data editing or manipulation functions that occur on a regular basis, it may be worth the effort of automating these tasks with geoprocessing. You can easily turn data manipulation and editing tasks into repeatable tasks and combine functions by using geoprocessing models or scripts, which can be scheduled to execute in a batch mode. You can preset values to reduce the likelihood of incorrectly entering information. Geoprocessing models and scripts ensure that functions are executed successfully and in the same order.

Configure the Infrastructure Network Editing Template

The Infrastructure Network Editing template is an ArcGIS 10 editing map and toolbar for managing water, wastewater, and stormwater utility data. The Infrastructure Editing toolbar contains a series of custom editing and reporting tools that enrich the editing experience for utilities working with infrastructure data. For example, there are tools for the following tasks:

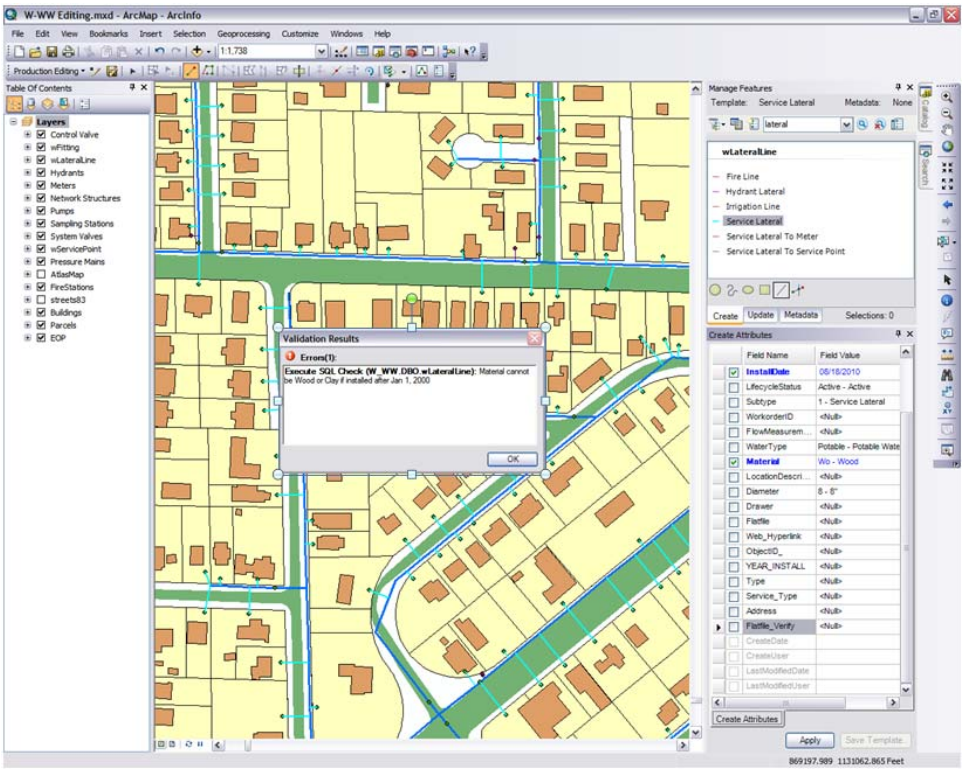
- Automatically connect service connections to laterals and their mains.
- Report tracing results along the utility network.
- Graph the profile of a main.

The template also includes an editor extension called Attribute Assistant. This extension uses a series of predefined methods to automatically populate attributes when you update and/or add new features to the geodatabase.

J10034

Utilize On-the-Fly Validation

A geodatabase is used to model predefined attribute rules. For example, setting up domains can limit the values only in an individual field. However, in reality, attribution rules are often more complex, and in many cases, the value in one field may depend on the value in another. Esri® Production Mapping allows you to create extended validation rules so that when you are creating new features or updating existing features, it is possible to validate attributes according to these rules. If an incorrect attribute value is entered, an error message is returned. You must fix the error before the change is committed to the feature. You can also choose to return a warning depending on the severity of the error. In such cases, the change will be committed, but it is recommended that you correct the issue. This type of attribute validation on the fly allows you to ensure that there are no invalid combinations of values assigned to a template or a selected feature before committing the values to the database.



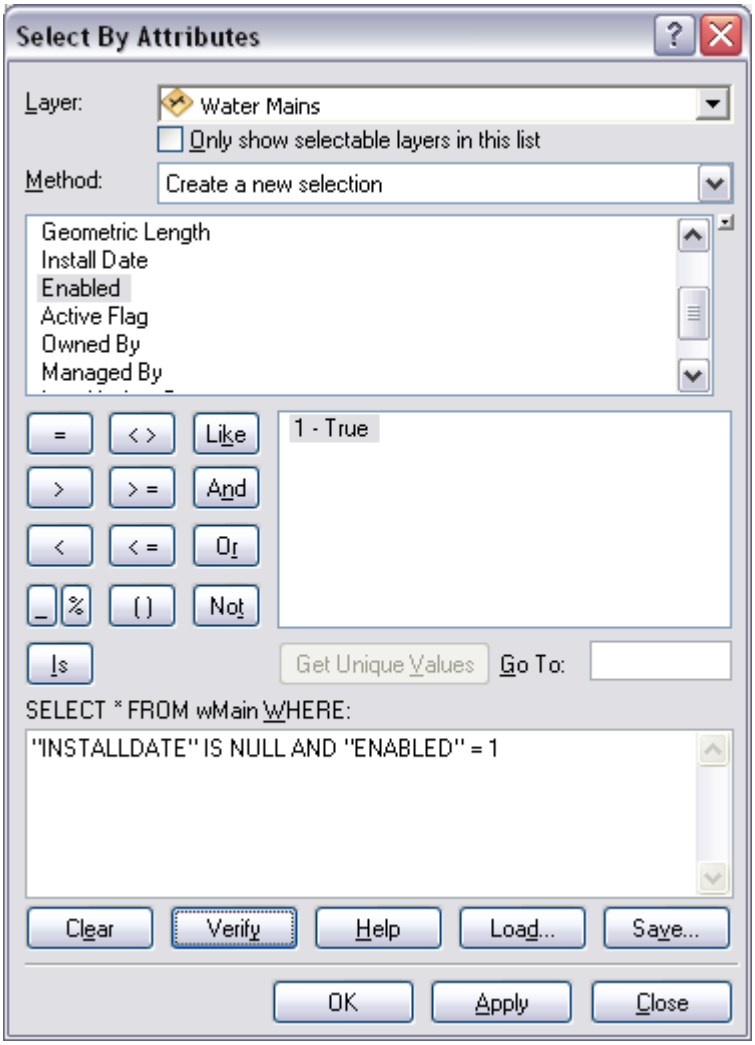
Increase the efficiency of GIS data creation using on-the-fly feature validation.

Quality Control Tools

Quality control is focused on identifying and resolving existing errors. The key to quality control is to regularly check your data for errors. Data checks can range from running queries (spatial, attribute, or both) to generating and reviewing random samples to perform visual exploration of data. Data checks can (and should) be formalized workflows that are automated and reusable where possible.

Finding Issues with Queries

The ArcMap Select By Attribute and Select By Location query dialog boxes allow you to define parameters and select features that meet those parameters. You can build queries to find attribution errors, for example, pipes installed after January 2000 with a material type of wood or clay. Spatial queries can be used to find features that may not be in the correct geographic location. When performing spatial queries, you can compare features that are within your water network as well as other features. For example, you can compare your service connections to parcel basemap data to identify connections that are not within a parcel.

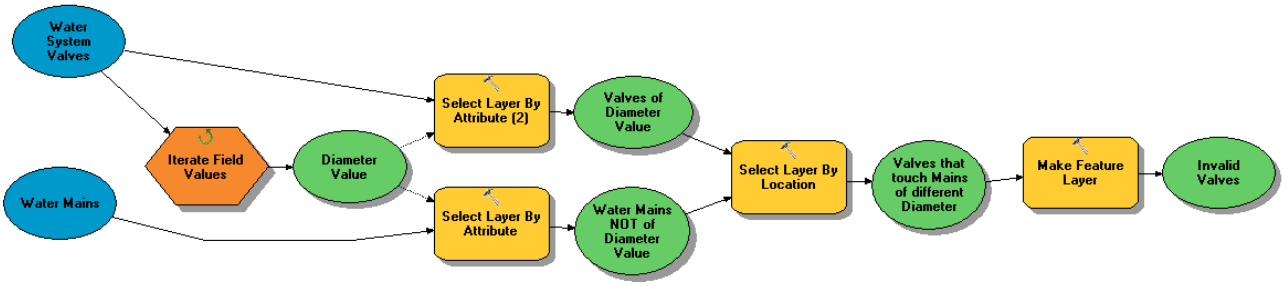


This sample Select by Attributes query has been configured to find water main features that are enabled and have a null value for the installation date.

J10034

Geoprocessing Models or Scripts

With queries using the Select By Attribute or Select By Location tools, the parameters must be filled out every time you want to run the query. This may introduce the possibility of user errors while entering the parameter values. With geoprocessing models or scripts, you can store as many preconfigured values as are needed, which provides consistency in running the queries. Geoprocessing also offers the ability to perform (and automate) complex queries and additional behaviors such as combining multiple queries and iterating.



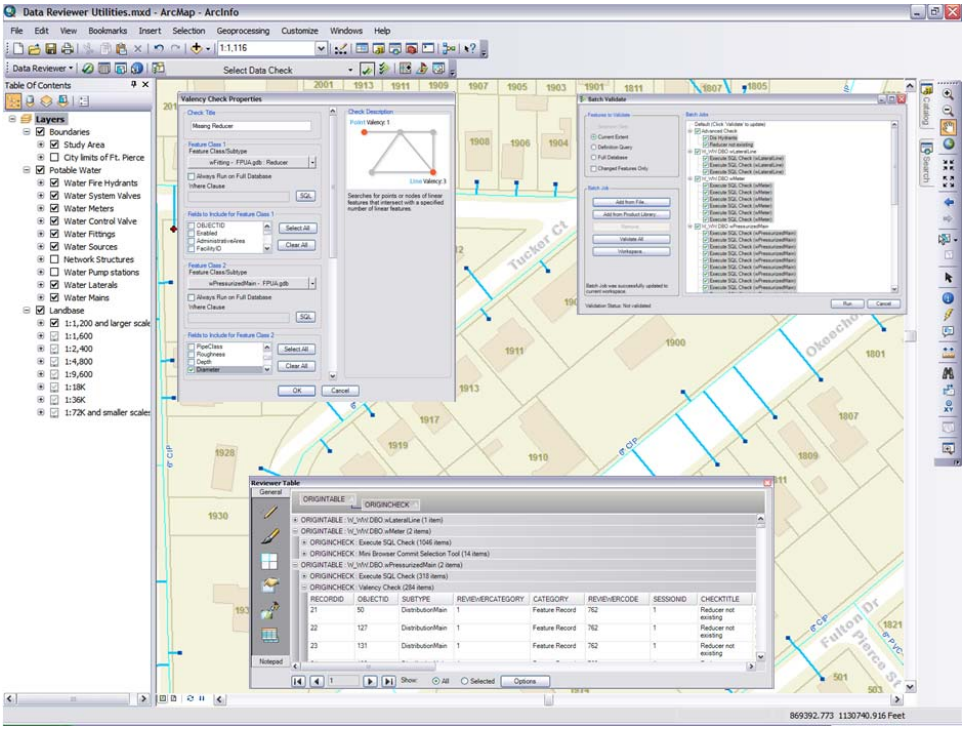
This sample geoprocessing model finds valves whose diameters do not match the diameters of the connecting mains.

Automated Quality Control

While it is possible to automate some of the quality control tasks with geoprocessing models or scripts, you may have to develop and maintain a number of models/scripts to cover all your data validation rules. The ArcGIS Data Reviewer extension to ArcGIS Desktop provides easily configurable, out-of-the-box data checks to automate your quality control process. Over 40 automated data checks can be run ad hoc or as a group in a batch job. Batch jobs are distributable, thereby allowing you to standardize the validation process throughout the utility. This also allows repeatability; that is, you can reuse the batch jobs to validate features as you make updates to them. Features that violate the parameters that you've specified are collected and managed in a central table. You can also schedule the checks to run at set intervals—eliminating the need to manually run the batch job—to further free up resources and reduce the time needed to obtain and maintain high-quality data. Examples of some of the automated checks available are described in the following table.

Check Name	Check Description	Use Case Example
Domain	Validates features where coded value and range domains have been assigned to ensure that all values meet geodatabase attribute domain constraints	These checks are useful when migrating GPS-collected data and data from another format, like CAD, shapefile, or coverage, into the geodatabase.
Subtype	Searches for features with improper or null (optional) subtype code	
Connectivity Rules	Finds features that are part of a geometric network and violate connectivity rules	Connectivity rules are an important aspect of a network. Identifying features that violate geometric network rules and resolving them will enhance data integrity. This helps functions like valve isolation, which uses the geometric network.
Relationship	Searches for records that are orphans or have improper cardinality in a relationship class	Knowing the exact number of pumps in the pump station or treatment plant is important. This check can find orphan pumps that do not have a relationship to a pump station or treatment plant.

Check Name	Check Description	Use Case Example
Duplicate Geometry	Finds features of the same geometry type that are colocated and optionally share attributes (Features can be from either two different feature classes or the same feature class.)	This check can find locations where two or more junctions in a geometric network are on top of each other. Only one of those junctions can actually be connected to the network. These duplicate features can be identified and addressed to ensure proper connectivity of network features.
Geometry on Geometry	Finds features that have a specific spatial relationship, from either two different feature classes or the same feature class, and optionally compares specific attributes	This check can find valves that are connected to mains and compare the diameter. If the diameters do not match, it is reported as an error.
Valency	Validates relationships between point and line features or line features within the same feature class by ensuring that specific patterns of features are met with valency, such as how a point must have a specified number of lines connected to it	This check can validate spatial relationships; for example, tees should be connected to three pipes or a reducer must connect two pipes of different diameters.



Automate and simplify the data review processes with tools to assess, document, correct, and verify the overall quality of water/wastewater data.

You can further simplify your quality control process by integrating the ArcGIS Data Reviewer and ArcGIS Workflow Manager extensions and using the provided custom steps in your workflow. This allows you to run batch jobs as part of your overall workflow and automatically assign an editor to make corrections to the errors found.

Ensuring Issues Are Corrected

While it is important to know where errors have occurred in your data, it is equally important to ensure that the errors are corrected in a timely manner. ArcGIS Data Reviewer tracks errors as they are resolved and verified, storing the errors in a central location, called the Reviewer table, which allows you to monitor and organize the error information. The table can be used to generate reports that represent an accurate picture of the overall quality of your data.

By examining the information stored in the Reviewer table, you continually improve the quality of your data. For example, the table allows you to sort and group error information to see what feature classes are most prone to errors and what types of errors are most prevalent. This can help you make decisions about QA/QC methods that can be introduced to maximize efficiency and ensure high-quality data.

Summary

This paper describes many options that are available to water, wastewater, and stormwater utilities to ensure the quality of their data. The geodatabase offers key features, such as domains and geometric networks that allow you to model real-world features within the database as well as minimize the possibility of introducing errors. Tools like Select By Attributes and Select By Location and geoprocessing models allow you to search for problems in your data. Additional extensions like ArcGIS Data Reviewer, ArcGIS Workflow Manager, and Esri Production Mapping are designed to find errors and ensure consistency of workflows that improve the overall quality of your data.

This paper also outlines various options for QA/QC. To summarize, start by taking a careful look at how the most critical errors in your data are introduced, and then begin implementing ideas that are provided in this paper to address those situations.

More Information

To find out how water, wastewater, and stormwater utilities are using GIS, please visit esri.com/water. For more information on the extensions, please visit esri.com/workflowmanager, esri.com/datareviewer, and esri.com/productionmapping. Access blogs, tutorials, videos, forums, and industry best practice downloads from the ArcGIS Resource Center for the water utilities community.

About Esri

Since 1969, Esri has been helping organizations map and model our world. Esri's GIS software tools and methodologies enable these organizations to effectively analyze and manage their geographic information and make better decisions. They are supported by our experienced and knowledgeable staff and extensive network of business partners and international distributors.

A full-service GIS company, Esri supports the implementation of GIS technology on desktops, servers, online services, and mobile devices. These GIS solutions are flexible, customizable, and easy to use.

Our Focus

Esri software is used by hundreds of thousands of organizations that apply GIS to solve problems and make our world a better place to live. We pay close attention to our users to ensure they have the best tools possible to accomplish their missions. A comprehensive suite of training options offered worldwide helps our users fully leverage their GIS applications.

Esri is a socially conscious business, actively supporting organizations involved in education, conservation, sustainable development, and humanitarian affairs.

Contact Esri

1 800 GIS XPRT (1 800 447 9778)
T 909 793 2853
F 909 793 5953
info@esri.com
esri.com

Offices worldwide
esri.com/locations

esri®
380 New York Street
Redlands, California 92373-8100 USA

Printed in USA