



**Esri International User Conference | San Diego, CA**  
**Technical Workshops | July 13, 2011**

# **ArcGIS Data Reviewer** **An Introduction**

Jay Cary & Michelle Johnson



# Workshop Overview

- **Importance of Data Quality**
- **What is ArcGIS Data Reviewer**
  - **Automated Data Review**
  - **Batch Data Review**
  - **Visual Data Review**
- **Managing Errors and Reporting Data Quality**
- **Summary/Resources**

# Ensuring Data Quality with ArcGIS Data Reviewer

Defining Quality

Understanding Data Reviewer

An Introduction to Data Reviewer Components



# Defining Quality

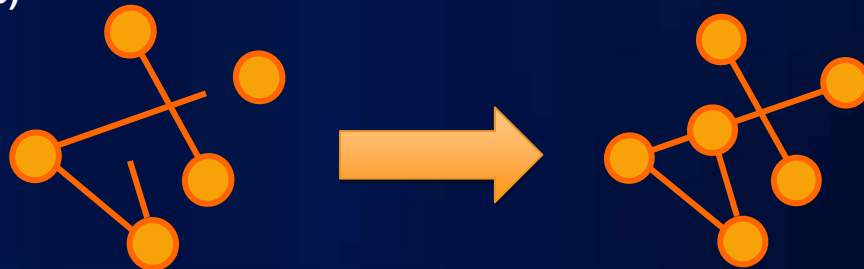
## What is 'good' quality?

- **Positional Accuracy**
- **Attribute Accuracy**
- **Completeness**
- **Logical Consistency**
- **Lineage**

Requirements are defined:

- Intended purpose and usage
- Technical specifications
- Client requirements

(Guptill & Morrison, 1995)



# Defining Quality Control

- **Quality Control**
  - The tasks required to ensure data quality
  - Methods to assess and evaluate data
- **Quality Assurance Plan**
  - Defines desired level of data quality
  - Specifies acceptable and unacceptable levels of error
  - Identifies how quality control will be performed

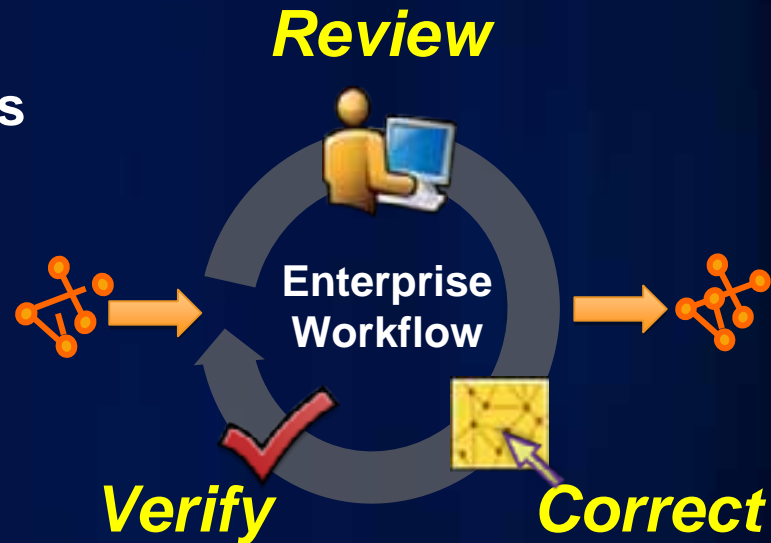
# Data Quality Matters

- **Cost of poor quality can be extreme**
  - \$\$ to fix
  - Impacts users
  - Reputation
- **Consider for a moment**
  - Do you publish a doc without spell checking?
  - Do you operate a PC without a virus scanner?



# ArcGIS Data Reviewer

- Supports data QC process
- Provides
  - Rule based workflows
  - Interactive tools
  - Track errors
- For individuals and enterprise
  - Saves time/money
  - Less rework



*Trusted data through improved quality management*

# Managing Quality Control

- Framework for managing QC lifecycle





## The Quality Control Process with Data Reviewer

## Review Data & Identify Errors

- Heart of the Data Reviewer
- Specialize on data quality control
- Automated checks
- Similar to an audit session
- Results on audit screen
- Tracks all anomalies/errors
- Batch Data Review
- All from a single interface for QC
- Consistent data error for QC purposes
- Get on base
- Visual Review
- Interactive table
- Report data accuracy
- Feature class & tables created automatically

# **Ensuring Data Quality with ArcGIS Data Reviewer**

An Introduction to Data Reviewer



# **Performing Quality Control Automated Data Review**

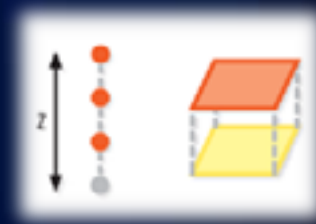
**An Introduction to Automated Data Review**

**Understanding Data Reviewer Checks**



# An Introduction to Automated Data Review

- **Validate data against a set of rules**
- **Quickest and least expensive form of quality control**
- **Validate large amounts of data (100%)**



# Understanding Data Reviewer Checks

## Implement Quality Requirements as Rule-Based Data Reviewer Checks



- 42 configurable data checks
- Grouped into 11 functional areas

# Table Checks



## Execute SQL Check



Mains installed after 1 January 2000. should only be Cast Iron, Ductile Iron, or PVC

The screenshot displays the 'Table Checks' application interface. A red rectangle highlights the 'Execute SQL' section, which contains the text: 'Find features based on a SQL query WHERE clause:'. Below this, the 'Execute SQL Check Properties' dialog box is open. The 'Check Title' field contains the text: 'Material should be either CI, DI or PVC'. The 'Feature Class/Subtype' dropdown is set to 'wPressurizedMain - FPUA.gdb'. The 'Where Clause' field contains the text: 'Material not in ('CI', 'DI', 'PVC') and InstallDate >= date '2000-01-01 00:00:00''. The 'Reviewer Remarks' section has a 'Notes' field with the text: 'Material should be either CI, DI or'. The 'Severity' dropdown is set to '5'. The 'Check Description' panel on the right shows a diagram of a water main with a yellow label 'TYPE:4'. In the background, an 'Attributes' table is visible with the following data:

Attribute	Value
Water Main	1/1/2000
InstallDate	1/1/2000
OperationalArea	District 3
LifecycleStatus	Active
Subtype	DistributionMain
WorkorderID	
FlowMeasurementID	<null>
WaterType	Potable Water
Material	PE



# Feature on Feature Checks

## Geometry on Geometry Check



### Feature on Feature Checks

Hydrants must be connected to a Hydrant Lateral

The screenshot displays the 'Geometry on Geometry Check Properties' dialog box, which is used to configure feature-on-feature checks. The dialog is divided into two main sections, each representing a different check.

**Check 1: Hydrant not connected to Hydrant Lateral**

- Check Title:** Hydrant not connected to Hydrant Lateral
- Feature Class:** wHydrant - FP10A.gdb
- Feature Class/Subtype:** wHydrant - FP10A.gdb
- Always Run on Full Database:** ☐
- Where Clause:** (empty)
- Check Description:** Returns relationship specific information for features in Feature Class 1 that are not connected to features in Feature Class 2. Relationships include crosses, intersects, touches, and others. Any can be used containing tolerance or relationship classes.
- Spatial Reasoning Check:** Type: Intersects; ☒ Not - find features not in the relationship
- Tolerance:** 0 Meters
- Feature Class 1:** wHydrant - FP10A.gdb
- Feature Class 2:** wHydrant Lateral - FP10A.gdb
- Rules:** Hydrant is not connected to a
- Severity:** 5 (High)

**Check 2: Hydrant not connected to a Water Main**

- Check Title:** Hydrant not connected to a Water Main
- Feature Class:** wHydrant - FP10A.gdb
- Feature Class/Subtype:** wHydrant - FP10A.gdb
- Always Run on Full Database:** ☐
- Where Clause:** (empty)
- Check Description:** Returns geometries for features in Feature Class 1 that are not connected to features in Feature Class 2 with the intersections of features in Feature Class 2.
- Spatial Reasoning Check:** Type: Intersects; ☒ Not - find features not in the relationship
- Tolerance:** 0 Meters
- Feature Class 1:** wHydrant - FP10A.gdb
- Feature Class 2:** Water Main - FP10A.gdb
- Rules:** Hydrant is not connected to a
- Severity:** 5 (High)

The 'Spatial Reasoning Check' dialog box is also visible, showing the same settings for both checks.

# Advanced Checks

10

## Valency Check



Reducer must connect two mains of a different diameter



# New Data Reviewer Checks at 10



- **Metadata Check**

- Validate metadata against published standards or custom expressions



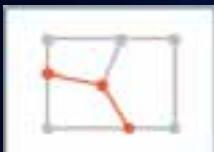
- **Valency Check**

- Validate conditions between line and point features



- **Custom Check**

- Validate data using custom code

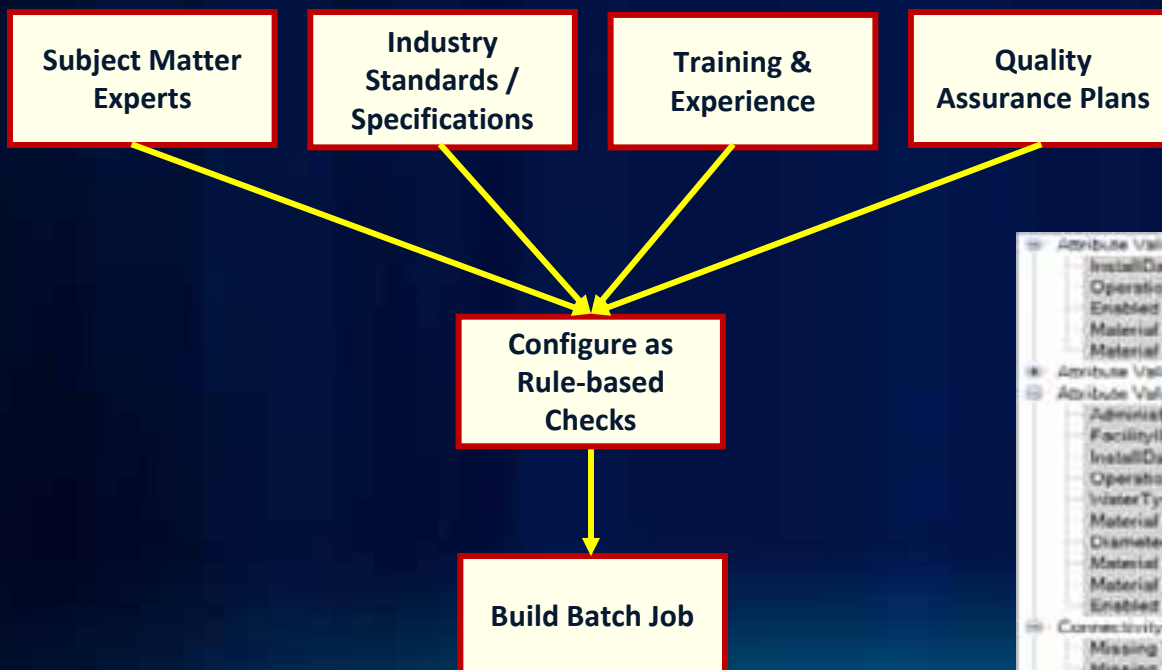


- **Topology Rules Check**

- Validate data against topology rules

# Combine Multiple Rules for Batch Check

- Encapsulate workflow / process
- Designed once and executed many times
- Complete spec check



```
Attribute Validation (w/LateralLine)
  InstallDate should be populated
  OperationalArea should be populated
  Enabled should be set to False if abandoned
  Material should be either CI, DI or PVC (post-2000 installation)
  Material cannot be Wood or Clay (post-2000 installation)
Attribute Validation (w/Meter)
  Attribute Validation (w/PressurizedMain)
    AdministrativeArea should be populated
    FacilityID should be NULL
    InstallDate should be populated
    OperationalArea should be populated
    WaterType should be populated
    Material should be populated
    Diameter should be populated
    Material should be either CI, DI or PVC
    Material cannot be Wood or Clay
    Enabled should be set to False
Connectivity Validation
  Missing Reducer (w/Fittings)
  Missing Tee (w/Fittings)
Disconnected Features
  Disconnected hydrants
```

# Batch Validation

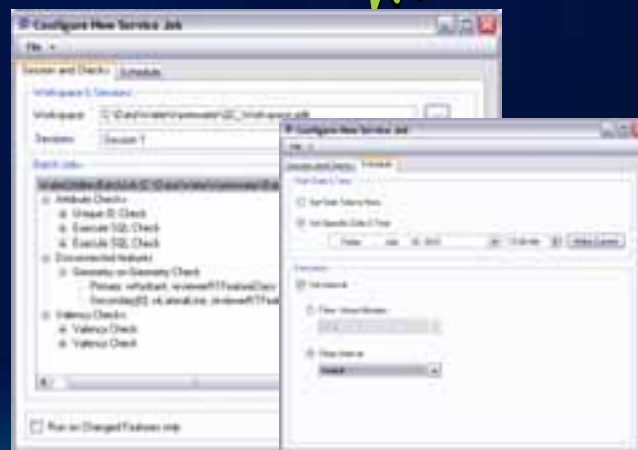


- Run multiple data checks
- Checks are organized into groups

## Additional Options for Batch Validation

## Run data validation outside ArcMap in several ways:

- **Reviewer Service (Windows Service)**
- **Reviewer Console (Command Line)**
- **From ArcToolbox**
- **Within a Python script**
- **As a step in ArcGIS Workflow Manager**



# Automated Data Review

Configuring Data Reviewer Checks  
Executing Automated Data Validations

# Performing Quality Control Visual Data Review

An Introduction to Visual Data Review

Structuring the Visual Data Review

Writing Records and Capturing Missing Features

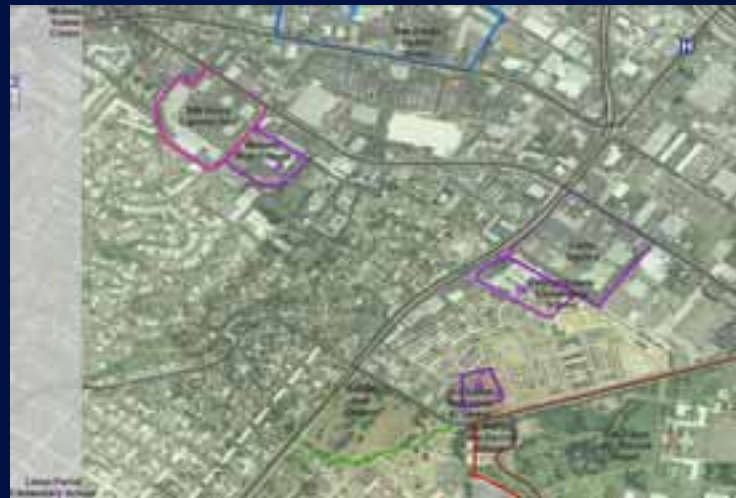
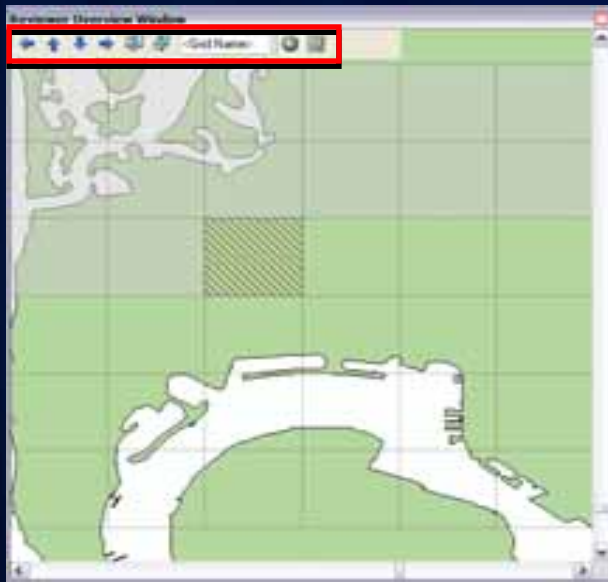


# **An Introduction to Visual Data Review**

- **Visual inspection of the data**
  - **Level of judgment/discretion**
  - **Compare between data sources**
  - **Requires strong knowledge of the data & ability to make consistent decisions**
- **Simplify visual review with Data Reviewer**
  - **Organize and structure the visual review process**
  - **Manage error information in the Data Reviewer table**



# Working with the Reviewer Overview Window

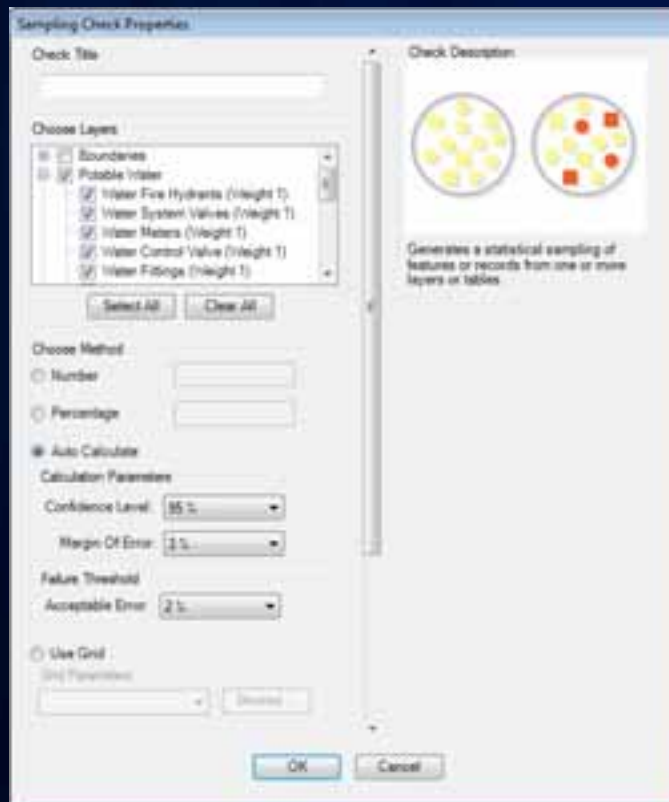


- Structure review by geographic extent
- Navigate to cells & track review progress



# Sampling Data

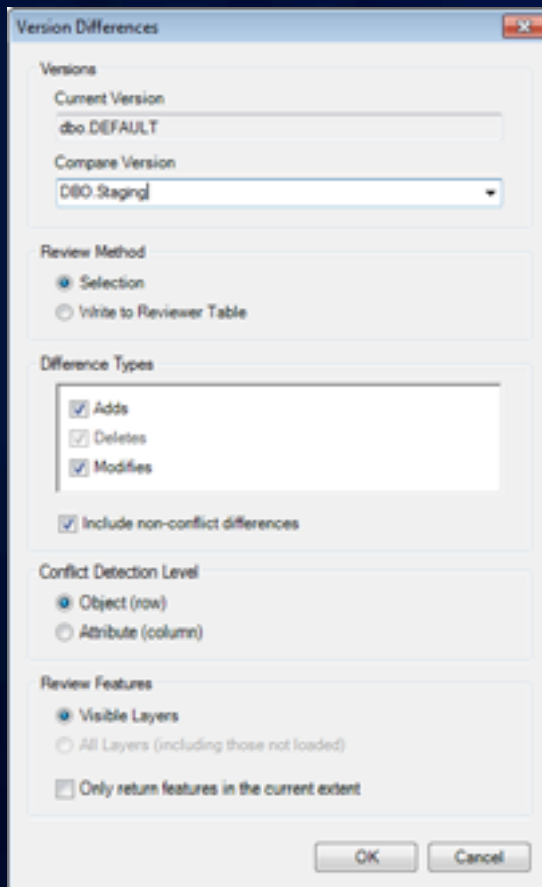
10



- **Determine a subset of features**
  - Fixed number of features
  - Percentage of features
  - Statistically significant sample
  - By polygon grid

# Version Differences Tool

10



The screenshot shows the 'Version Differences' dialog box with the following settings:

- Versions:**
  - Current Version:
  - Compare Version:
- Review Method:**
  - ☒ Selection
  - ☐ Write to Reviewer Table
- Difference Types:**
  - ☒ Adds
  - ☒ Deletes
  - ☒ Modifies
  - ☒ Include non-conflict differences
- Conflict Detection Level:**
  - ☒ Object (row)
  - ☐ Attribute (column)
- Review Features:**
  - ☒ Visible Layers
  - ☐ All Layers (including those not loaded)
  - ☐ Only return features in the current extent

Buttons: OK, Cancel

- Compare two SDE versions
  - Parent/Child, Child/Child
- Identify adds/modifies/deletes
- Browse results or write to Reviewer Table

# Commit to Reviewer Table



- **Write selected record(s) to the Reviewer Table**
  - Features selected interactively
  - Results of automated processes

**Commit to Reviewer Table**

Enter Review Status

Custom  
Add Feature  
Delete Feature  
**Move Feature**  
Reshape Line  
Reshape Area  
Change Subtype to:

Description: Move Feature

Additional Information

Notes

Severity 5 1 - High  
5 - Low

OK Cancel

# Capturing Missing Features



- Capture x,y location or complete geometry
- Identify type of missing feature

**Flag Missing Feature Details**

Feature Class / Subtype  
ParksPreserves - FPUA.gdb

Feature Class:  
ParksPreserves

Subtype:

Enter Review Status  
Add

Description: Add

Notes  
Athletic Field

Severity: 5

☒ Display incomplete dialog warning.

OK Cancel

# Visual Data Review

Managing Visual Review Process

Creating and Using a QC Grid

Identifying Errors Visually



# Measuring and Reporting Data Quality

Working with Reviewer Table Records

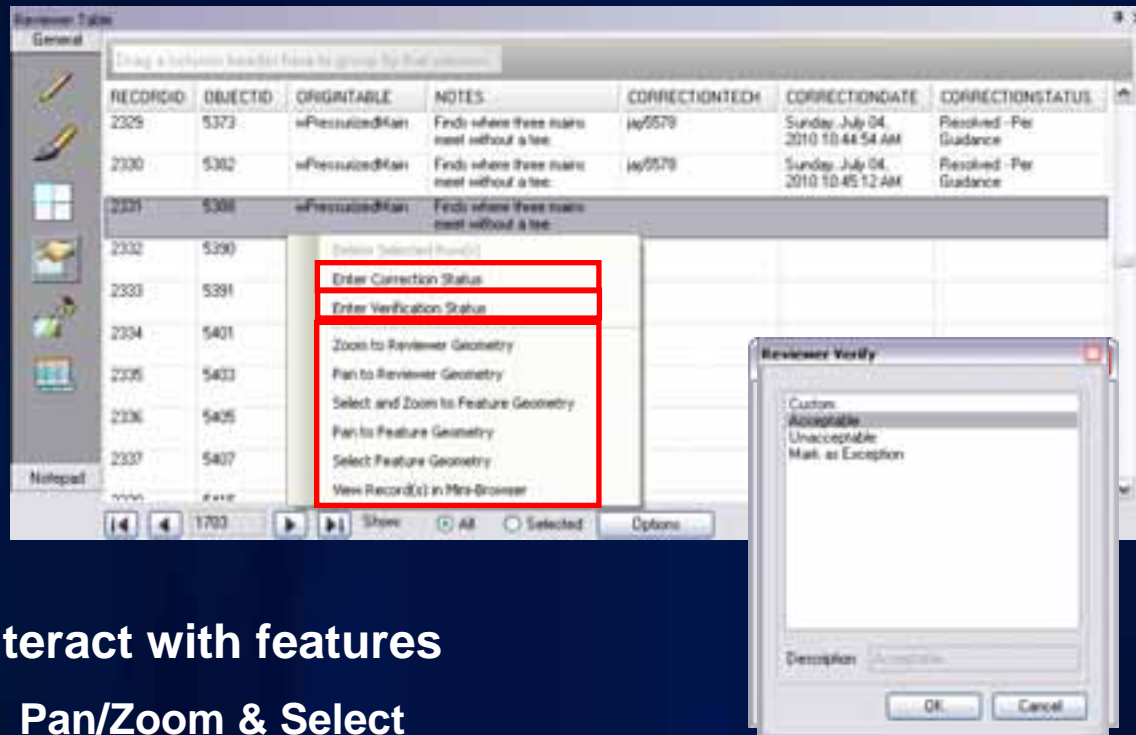
Summarizing Information

Reporting





# Working with Reviewer Table Records



- Interact with features
  - Pan/Zoom & Select
- Enter correction and verification status

# Summarizing Information

The screenshot shows the 'Reviewer Table' application window. At the top, there are two dropdown menus: 'ORIGINTABLE' and 'REVIEWSTATUS', both highlighted with a red rectangle. Below these, a list of summary items is shown, including counts for various error types. The bottom section displays a detailed table of records.

RECORDID	OBJECTID	SUBTYPE	ORIGINCHECK	NOTES	SEVERITY	REVIEWTECH
935	2	DistributionMain	Valency Check	Finds where three mains meet without a tee.	5	wend5397
936	5	DistributionMain	Valency Check	Finds where three mains meet without a tee.	5	wend5397
937	8	DistributionMain	Valency Check	Finds where three mains meet without a tee.	5	wend5397

- Sort and group by table fields
- Summarize quality control information
  - Identify most common data errors
  - View progress of quality control



# Reporting



- Automated reporting of quality control results
- Available Reports
  - Automated Check by Origin Table
  - Automated Check by Subtype
  - Automated Check by Group
  - Total Record Count
  - Sampling

Auto Calculate Sampling Report

Report generated on 7/4/2010 11:10 AM by user:JST  
 Review Worksheet Location: E:\Data\New\Newwater\QC\Workbooks\glt  
 Scenario(s): Scenario 1 - Scenario 2

Client Site	Date/Time	Origin Table	Subtype	Total Records	Confidence Level	Margin of Error	Sampling Number	Acceptable Error Percentage	Acceptable Error	Number of Errors	Pass/Fail	QC Complete
Sample												
	7/4/2010 11:00 AM											
		Address		113	95%	3%	3	2%	3	0	Pass	No
		Address		0			0		0	0		
		Address		30	95%	3%	3	2%	4	1	Pass	No
		Address		3			3		0	0		
		Address		4			4		0	0		
		Address		5			5		0	0		
		Address		287	95%	3%	3	2%	3	0	Fail	No
		Address		2			2		0	0		
		Address		2			2		0	0		
		Address		1010	95%	3%	104	2%	11	0	Pass	No
		Address		10			10		0	0		
		Address		610			610		0	0		
		Address		1			1		0	0		
		Address		1110	95%	3%	114	2%	11	0	Pass	No
		Address		114			114		0	0		
		Address		7	95%	3%	7	2%	0	0	Pass	No
		Address		7			7		0	0		
		Address		1010	95%	3%	104	2%	11	0	Pass	No
		Address		279			279		0	0		
		Address		8			8		0	0		
		Address		10			10		0	0		
		Address		9	95%	3%	9	2%	0	0	Pass	No
		Address		1			1		0	0		
		Address		101	95%	3%	104	2%	0	0	Pass	No
		Address		1			1		0	0		
		Address		1			1		0	0		
		Address		130			130		0	0		
Total							104		12			

# Measuring and Reporting Data Quality

Working with Reviewer Table Records  
Summarizing Information  
Reporting



# Summary: ArcGIS Data Reviewer

Automate, Simplify, and Improve your Quality Control Process

- Standard extension to Desktop
- Supports data QC process
- Provides
  - Rule based workflows
  - Interactive tools
  - Track errors
- For individuals and enterprise
  - Saves time/money
  - Less rework



# Resources

- Product Information [www.esri.com\datareviewer](http://www.esri.com/datareviewer)
  - 60-day Free Evaluation
  - ArcGIS Data Reviewer Checks poster
- Training [www.esri.com/training](http://www.esri.com/training)
  - Instructor Led Training: Data QC Using GIS Data Reviewer
  - Virtual Campus:
    - Introduction to GIS Data Reviewer 
    - Using ArcGIS Data Reviewer to Assess Data Quality 
- Data Reviewer Resource Center
  - [resources.arcgis.com](http://resources.arcgis.com)
- Email questions and comments to [DataReviewer@esri.com](mailto:DataReviewer@esri.com)

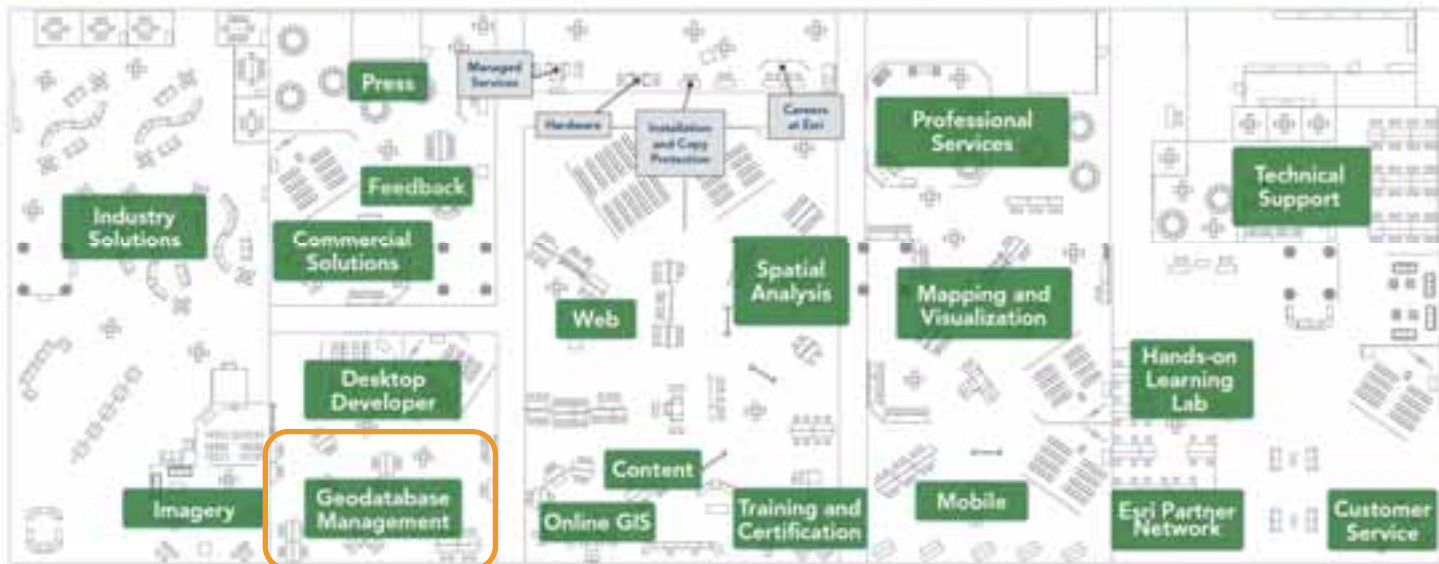
# Other Data Reviewer Sessions

Day and Description	Type	Time	Location
<b>Tuesday July 12</b>			
Assessing positional accuracy using ArcGIS Data Reviewer	Demo Theater	11:00 – 11:30 AM	Geodatabase Mgmt Theater
Integrating ArcGIS Workflow Manager and ArcGIS Data Reviewer to automate quality control workflows	Demo Theater	1:00 – 1:30 PM	Geodatabase Mgmt Theater
ArcGIS Data Reviewer - Extending Validation Framework	Demo Theater	5:30 – 6:00 PM	Geodatabase Mgmt Theater
<b>Wednesday July 13</b>			
ArcGIS Data Reviewer - Metadata quality control	Demo Theater	10:30 – 11:00 AM	Geodatabase Mgmt Theater
<b>ArcGIS Data Reviewer: An Introduction</b>	<b>Tech Workshop</b>	<b>1:30 – 2:45 PM</b>	<b>15B</b>
Assessing positional accuracy using ArcGIS Data Reviewer	Demo Theater	3:00 – 3:30 PM	Geodatabase Mgmt Theater
<b>Thursday July 14</b>			
Road Ahead - Data Reviewer	Tech Workshop	8:30 – 8:50 AM	6A
ArcGIS Data Reviewer: An Introduction	Tech Workshop	10:15 – 11:30 AM	15B
ArcGIS Data Reviewer - Extending Validation Framework	Demo Theater	12:00 – 12:30 PM	Geodatabase Mgmt Theater

# Visit the Geodatabase Management Island

## Esri Product Showcase

2011 Esri International User Conference



Copyright © 2011 Esri. All rights reserved. Esri, the Esri logo, ArcView, ArcGIS, and other names are trademarks of Esri. All other names are trademarks of their respective owners. Esri is not responsible for any errors or omissions in this document. Esri reserves the right to change the information in this document without notice.



# Session Evaluations

- Session feedback has gone digital!

[www.esri.com/sessionevals](http://www.esri.com/sessionevals)





esri