The Importance of Data Quality in your Enterprise
Heather Murnane & Sherry Fagin
Workshop Agenda

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

Delivering high-quality data & products
Defining quality
A business perspective

Executive
Confidently make decisions

Manager
Effective data stewardship

User
Confidence in data
Defining Quality
A Technical Perspective

- Spatial Accuracy
- Thematic Accuracy
- Completeness
- Logical Consistency
- Temporal Quality
- Usability

ISO-19157:2013 Geographic Information – Data Quality
What is ArcGIS Data Reviewer?
Data Quality Management for ArcGIS

• Provides
  - Rule-based validation
  - Interactive tools
  - Track errors

• For individuals and enterprise
  - Saves time/money
  - Less rework

• For multiple domains
  - Configurable
  - Extendable
Automated Review

Streamlining data validation
Managing Quality Control

Quality Control Processes
Defining Quality

Sources of data quality requirements

- Industry standards / Specifications
- Subject matter experts
- Training and experience
- Quality assurance plans
Automating Data Validation

- Implementing quality requirements
  - 40+ configurable checks
  - Attribute
    - Feature and table values
  - Spatial
    - Spatial relationships
  - Feature integrity
    - Collection rules
  - Metadata
    - Completeness/Content
Batch Validation
Implementing Cumulative Review

- Encapsulate quality rules
- Configured from 40+ automated checks
- Designed once and executed many times

Industry standards / Specifications
Subject matter experts
Training and experience
Quality assurance plans
Methods for executing data validation

Execute data validation using
- ArcMap
Methods for executing data validation

Execute data validation using

- ArcMap
- Model/Python script
Methods for executing data validation

Execute data validation using

- ArcMap
- Model/Python script
- ArcGIS Workflow Manager
Methods for executing data validation

Execute data validation using
- ArcMap
- Model/Python script
- ArcGIS Workflow Manager
- ArcGIS for Server
Methods for executing data validation

Execute data validation using

- ArcMap
- Model/Python script
- ArcGIS Workflow Manager
- ArcGIS for Server
- ArcGIS Pro
Authoring Data Quality Rules & Executing Batch Jobs
Sherry Fagin
Single Automated Check
Configure a Batch Job
Execute a Batch Job in ArcGIS Pro
Visual Review  
Discovering patterns & missing elements
Managing Quality Control

Quality Control Processes

Visual Review

Automated Review

Reviewer Results

Quality Reporting
Value of Performing Visual Review

- Discover Patterns
- Find missing features
- Compare to trusted sources
Tools supporting
- Selecting/browsing features
- Redlining missing features
- Flagging features in error
- Assessing positional accuracy
- Comparing geodatabase versions
- Generating random samples
Visual Review
Leveraging ArcGIS for Server

- Extending quality control workflows into other communities
  - QC review across ArcGIS platform
  - Simple to use tools for error identification
  - Manual QC workflow “automation”
Performing Visual Review
Sherry Fagin
Report Feature Widget
Managing and Reporting Quality Results

Communicating data quality
Managing Quality Control

Quality Control Processes

Automated Review

Visual Review

Reviewer Results

Quality Reporting
Managing Quality Control
- QC lifecycle management

Diagram:
- Review
- Correct
- Verify

Diagram elements:
- Reviewer Results
Data Quality Reporting
ArcGIS for Desktop

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

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Origin Table</th>
<th>Subtype</th>
<th>Total Records</th>
<th>Confidence Level</th>
<th>Margin of Error</th>
<th>Sampling Number</th>
<th>Acceptable Error Percentage</th>
<th>Acceptable Error</th>
<th>Number of Errors</th>
<th>Pass/Fail</th>
<th>QC Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/4/2018 11:05:59 AM</td>
<td>AtlasMap</td>
<td></td>
<td>153</td>
<td>98%</td>
<td>3%</td>
<td>0</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>wControlValve</td>
<td></td>
<td>302</td>
<td>98%</td>
<td>3%</td>
<td>2</td>
<td>2%</td>
<td>1</td>
<td>1</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wFitting</td>
<td>997</td>
<td>98%</td>
<td>3%</td>
<td>5</td>
<td>2%</td>
<td>1</td>
<td>2</td>
<td>Fail</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wLateralLine</td>
<td>18600</td>
<td>98%</td>
<td>3%</td>
<td>644</td>
<td>2%</td>
<td>17</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Data Quality Reporting
ArcGIS for Server

Dashboard
- Enabling transparency in data quality
- Better decision making by communicating data quality across stakeholders
- Open quality reporting
- Shared across ArcGIS system
- Tools and methods to communicate quality
Reporting Data Quality
Sherry Fagin
## Correct and Verify

### Reviewer Table

<table>
<thead>
<tr>
<th>Status</th>
<th>Source</th>
<th>Source Subtype</th>
<th>ID</th>
<th>Check Title</th>
<th>Check Notes</th>
<th>Severity</th>
<th>Description</th>
<th>Created On</th>
<th>Corrected On</th>
<th>Correction Notes</th>
<th>Verified On</th>
<th>Verification Notes</th>
</tr>
</thead>
</table>
# Excel Reporting

## Automated Check Report By Origin Table

Report generated on 2/10/2015 7:42 AM
Reviewer Workspace Location: C:\DemoFedGIS\Top10\FedGIS_Top10_Final\ReviewerWorkspaceTop10.gdb
Session(s): Session 8 : QC Tech

<table>
<thead>
<tr>
<th>Batch Job Name</th>
<th>Date/Time</th>
<th>Origin Table</th>
<th>Check Type</th>
<th>Check Title</th>
<th>Severity</th>
<th>Records Validated</th>
<th>Total Results</th>
<th>Total Percent Accuracy</th>
<th>Distinct Results</th>
<th>Distinct Percent Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>C:\Users\Documents\CTM_FedGIS_Checks_v1.1.rb</td>
<td>2/10/2015 12:38:17 PM</td>
<td>StructurePnt</td>
<td>Duplicate Geometry Check</td>
<td>Duplicate Geometry Check - Duplicate building points - Building_P</td>
<td>3</td>
<td>1455</td>
<td>40</td>
<td>97.26%</td>
<td>24</td>
<td>98.35%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>StructurePnt</td>
<td>Duplicate Geometry Check</td>
<td>StructurePnt</td>
<td>60</td>
<td>4</td>
<td>4</td>
<td>93.33%</td>
<td>4</td>
<td>93.33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>StructureSrf</td>
<td>Invalid Geometry Check</td>
<td>Invalid Geometry Check (StructurePnt)</td>
<td>5</td>
<td>496</td>
<td>36</td>
<td>92.73%</td>
<td>20</td>
<td>95.96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>StructureSrf</td>
<td>Cutbacks Check</td>
<td>Cutbacks Check - Angle is too sharp - Building_S</td>
<td>2</td>
<td>455</td>
<td>36</td>
<td>92.73%</td>
<td>20</td>
<td>95.96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>StructureSrf</td>
<td>Duplicate Geometry Check</td>
<td>Duplicate Geometry Check - Duplicate building surfaces - Building_S</td>
<td>2</td>
<td>55</td>
<td>8</td>
<td>85.45%</td>
<td>8</td>
<td>85.45%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Execute SQL Check</td>
<td>SQL Check - Controlling Authority Can't be 'No Information' - Building_S</td>
<td>2</td>
<td>55</td>
<td>4</td>
<td>92.73%</td>
<td>4</td>
<td>92.73%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Execute SQL Check</td>
<td>SQL Check - Feature Function Can't be 'No Information' - Building_S</td>
<td>1</td>
<td>55</td>
<td>1</td>
<td>88.18%</td>
<td>1</td>
<td>88.18%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Execute SQL Check</td>
<td>SQL Check - Roof Shape is not defined - Building_S</td>
<td>2</td>
<td>55</td>
<td>4</td>
<td>92.73%</td>
<td>4</td>
<td>92.73%</td>
</tr>
</tbody>
</table>
Dashboard Reporting
Case Study
Air Combat Command
Geospatial Enterprise
Data Quality

Richard Olivieri
ACC/A7Z
10 February 2015

This Briefing is:
UNCLASSIFIED
The Common Installation Picture (CIP) is the common base map maintained by CE and shared widely across the base LAN. Mission Data Layers (MDL) are functionally specific map layers that are stored, where possible, with the CIP and are shared on a need-to-know basis as determined by functional data steward. Functional communities generate fused installation pictures with CIP and select MDL in functional IT solutions. The CIP is the common base map, classified or unclassified as determined by functional user, and The CIP is maintained by CE and shared widely across the base LAN.
Direct Warfighter Support

- 24x7x365 worldwide access to data and maps
- Data is authoritative and timely (days-old)
- Web services support SOA
- Substantial reduction in licensing costs
- Increased uptime by centralizing management
- System realizes a true enterprise spatial data infrastructure framework
History

- Lengthy manual process
- We created CIP Checker because Data Reviewer was bundled with PLTS and did meet requirements
- CIP Checker was great but it became outdated for GeoBase requirements
• Utilize Data Reviewer for QA/QC.
• Checks run monthly on each base
• Data quality reports online
Key Checks to use on all Feature Classes in your Database:

- Invalid Geometry
- Duplicate Geometry
- Domain
- Polygon Overlap (for obvious features like buildings and parking lots)
Mission Impact

- Improve quality of (and trust in) data
- Helps other Air Force programs capitalize on the data
- Reduce manpower required to manage the data
- Improve utilization of webtools
- Standard schema CRUCIAL for Enterprise GIS
Mission Impact Examples

Web Map Viewer and Services
Mission Impact Examples

Expeditionary Airfield Management and Planning Tool
Visualizing Data Reviewer Results

Mike Cannon
USAF ACC GeoBase
Data Reviewer
ESRI Data Reviewer Reports

• Used to perform Quality Checks against Geo-Spatial data
• Determine valid and invalid geometries
• Batch multiple checks
• Users can create custom attribution checks
  • Checks based on defined standards
• Provides detail information on Data
• GeoMetrix takes generated reports from ESRI Data Reviewer Reports and provides a visual interface that provides a summarized view of Data Quality.
  • Visualize Data Reviewer reports
  • High level view of data quality
    • MAJCOM Summary
    • Base/Wing Summary
  • Drill down details
    • Geometry, Attribution, Domain and Custom Checks
  • View Trends over time
  • Promotes awareness of data quality
  • Helps identify areas for improvement
The Results – Data Quality Improves

• Since GeoMetriX has been implemented at USAF ACC GeoBase data quality has improved.
  • Streamline processing of Data Reviewer Results
  • Better visualization and transparent reporting
  • Clear Definition on what needs to be improved/corrected
  • Leadership at both Wing and MAJCOM can see the same picture
Demo
Questions?

Richard Olivieri
Michael Cannon
ACC Geospatial Information Office
ACC/A7ZG

Richard.Olivieri.ctr@us.af.mil
Michael.Cannon.11.ctr@us.af.mil

DSN 574-4184
Comm 757-764-4184
Workshop Review

- Importance of data quality
- Components of data quality
- Forms of data quality control
  - Automated review
  - Visual/Manual review
- ArcGIS Data Reviewer
  - Automated validation
  - Visual QC tools
  - Reporting
  - Error lifecycle management
Resources

Product page

Resource Center

Training
- www.esri.com/training

Questions & comments
- datareviewer@esri.com
Don’t forget to complete a session evaluation form!