Consistency and Quality of INSPIRE & ELF Data, using GIS Tools

Presentation to: Workshop on Spatial Data and Map Quality

Author: Paul Hardy (phardy@esri.com), Jay Cary, Roberto Lucchi, Rene Hinojosa

Date: 20 Jan 2015, Malta
ESRI Quality Solution

★ Geodatabase Integrity
- Schema constraints (Attribute Domains etc)
- Rich data types (networks, etc)
- Geoprocessing tools (Check Feature etc)
- Data Load Checks (SDE)
- Versioning to protect integrity of long transactions

★ Topology
- Topology Rules and error handling

★ ArcGIS Data Reviewer
- Primary Quality Assessment tool

★ ArcGIS for INSPIRE
- Pre-prepared INSPIRE/ELF geodatabase data models
- Quality rule batch jobs
Data Loading and Checking

- ArcGIS reading of shapefiles etc is tolerant of errors
  - Null geometry, Self intersections, Invalid holes
- So two geoprocessing tools to detect and repair
  - Check_Features, Repair_Features
- When data stored in SDE geodatabase – less tolerant
  - SDE will clean geometry on load, to canonical forms
- Geoprocessing framework and model builder
  - 1000+ geoprocessing tools, including ones for data loading and restructuring
- Data Interoperability Extension
  - 70+ formats read
  - Safe Software FME and workbench
Topology Rules

ArcGIS® Geodatabase Topology Rules

- Must not overlap
- Must not have gaps
- Contains point
- Boundary must be covered by
- Must be covered by feature class of
- Must not overlap with
- Area boundary must be covered by boundary of
- Must be properly inside polygons
- Must be larger than cluster tolerance
- Endpoint must be covered by
- Point must be covered by line
- Point must be covered by endpoint of
- Point must be covered by boundary of
- Line must not intersect or touch interior
- Line must overlap with
- Line must not have dangles
- Line must not self overlap
- Line must not have pseudonodes
- Line must not have multiplicity
- Polygon must not have gaps
- Polygon must be covered by
- Polygon must not overlap with
- Polygon must cover each other
- Polygon must be covered by boundary of
- Polygon must be covered by feature class of
- Polygon must not overlap with
- Polygon must not have dangles
- Polygon must not have pseudonodes
- Polygon must not have multiplicity

Topology rules ensure data quality and accuracy in geospatial applications.
ArcGIS Data Reviewer Checks

Polygon Checks

- **Evaluate Polygon Perimeter and Area**: Searches for polygon, part, ring, or segment features whose area or perimeter is within a specified range.

- **Invalid Hole Feature**: Finds features that intersect polygon feature holes.

- **Polygon Sliver**: Finds polygons below a specified thinness ratio (t) and optionally whose area is within a specified threshold.

Feature on Feature Checks

- **Geometry on Geometry**: Searches for features from two different feature classes or within the same feature class that spatially interact (e.g., intersect) or are within a tolerance of each other.
  - **Comparison to Topology**: Line—Must not overlap, intersect, or overlap with Polygon—Must not overlap, contain point, or overlap with

- **Intersection on Geometry**: Returns geometries for features in Feature Class 1 that intersect with the intersections from features from Feature Class 2 and 3.

- **Polygon Overlap/Gap Is Sliver**: Returns overlap/gap geometries between polygon features from two feature classes that have a thinness ratio beneath a user-specified threshold; optionally requires that the overlap/gap polygons be beneath a maximum area threshold.
  - **Comparison to Topology**: Polygon must not have gaps.
ArcGIS Data Reviewer Checks

Duplicate Geometry Checks

- Duplicate Geometry: Finds features of the same geometry type that are collocated.

Duplicate Vertex: Searches for vertices in selected polyline or polygon feature classes that are within a specified tolerance of each other.

Polyline Checks

- Cutbacks: Identifies segments where the angle between segments in a polygon or polyline is below a specified minimum value.

- Evaluate Polyline Length: Searches for segment, part/path, and polyline features whose length is within a specified range.
ArcGIS Data Reviewer Checks

**Database Validation Checks**

- **Connectivity Rules**: Returns geometries for features that violate the geometric network connectivity rules.

- **Domain**: Validates coded value and range domains to ensure that all values meet domain constraints.

- **Relationships**: Searches for records that are orphans or have improper cardinality in a relationship class.

- **Subtype**: Searches for feature classes with improper or null subtypes.

**Table Checks**

- **Execute SQL**: Finds features based on a SQL query WHERE clause.

- **Regular Expression**: Finds features with attribute values that violate the regular expression.

- **Table to Table Attribute**: Returns rows whose attributes match those of a feature class or table and/or comply with a user-defined WHERE clause comparing the attributes between feature classes and/or tables.

- **Unique ID**: Checks the values of a set of fields across a set of tables and feature classes for uniqueness within a given workspace.
ArcGIS Data Reviewer Checks

Z-Value Checks

Adjacent Vertex Elevation Change
Finds vertices for polyline or polygon features with elevation (z-value) changes greater than the specified tolerance.

Different Z at Intersection
Finds two intersecting line features whose z-value difference is within the minimum/maximum specified tolerance values at the point where they intersect.

Evaluate Z-Values
Searches for features whose z-values are within a specified range.

Polygon/Ring Closed
Searches for unclosed rings in polygons based on the x-value, y-value, and z-value.

Slope Direction Change Monotonicity
Searches for polylines with vertices that change slope direction according to specified conditions.

Topology Checks

Find Dangles
Finds polyline features with nodes that are within a user-defined tolerance but not connected to other polyline or polygon features.

Orphan
Finds single polyline features that are not connected in the database topology.

Unnecessary Nodes
Finds features that share a node and have identical attributes in common fields. Comparison to Topology: Line must not have pseudo nodes.

Unnecessary Polygon Boundaries
Finds adjacent polygon features that share a boundary and have identical attributes in editable fields.

Topology Rules
Returns the geometry of features that violate the topology rules that have been defined for a feature dataset in the geodatabase.

Default Checks

Invalid Geometry
Finds features whose geometry is empty, nothing, or not simple, as well as features with empty envelopes.

Multipart Line
Searches for polyline features with more than one part. Comparison to Topology: Line must be single part.

Multipart Polygon
Finds polygon features with more than one part and polygon features with holes.

Nonlinear Segment
Searches for nonlinear segments, such as arcs and curves, in line and polygon features.

Polyline or Path Classes on Self
Finds paths and lines in polylines that touch or cross themselves. Comparison to Topology: Line must not self intersect or self overlap.
ArcGIS Data Reviewer Checks

### Spatial Parameter Evaluation Checks

**Evaluate Extent**
- Returns features where the extent properties (x and y) are within specified parameters

**Evaluate Intersection Count**
- Finds vertices for polyline features in one feature class that intersect polyline or polygon features in a second feature class a specified number of times

**Evaluate Part Count**
- Finds features with a part count that is within a specified range

**Evaluate Vertex Count**
- Searches for features with a vertex count that is within a specified range

### Advanced Checks

**Composite**
- Searches for features that satisfy combined Geometry on Geometry and/or Table to Table attribute checks by feeding the results of one check into the next check

**Custom**
- Returns records that meet the validation conditions of a custom check you developed to meet your specific requirements

**Metadata**
- Validates that the specified GIS item descriptions have been populated to comply with either a standards-based metadata schema or user-specified content requirements

**Sampling**
- Generates a statistical sampling of features or records from one or more layers or tables

**Valency**
- Searches for points or nodes of linear features that intersect with a specified number of linear features
Data Reviewer – Manually trigger checks
Combine Multiple Rules for Batch Check

- Implementing cumulative review
  - Encapsulate QC model
  - Designed once and executed many times
  - Complete specification check

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

Configure as rule-based checks
Build Reviewer Batch Job
ELF Quality Rules

- Being created in RuleSpeak by WP2
  - For Regional/Global scale initially, then master levels
- Esri has gone through spreadsheet
  - Added column with ADR check for each rule
  - Now implementing rules as an ADR batch job
  - Has been lack of test data
- Spain NMA and Cadastre offered help
<table>
<thead>
<tr>
<th>RuleID</th>
<th>Quality Element</th>
<th>Feature type</th>
<th>RuleSpeak rule</th>
<th>DR Check/ ArcToolbox</th>
<th>Regional</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYD01</td>
<td>completeness commission</td>
<td>Island</td>
<td>The area of a surface feature must be equal or greater than the target area size.</td>
<td>Evaluate Polygon Perimeter and Area &lt; 2 km²; Target area size = 0,4 km²</td>
<td>Target area size</td>
<td>3 km²</td>
</tr>
<tr>
<td>HYD02</td>
<td>completeness commission</td>
<td>Wetland</td>
<td>The area of a surface feature must be equal or greater than the target area size.</td>
<td>Evaluate Polygon Perimeter and Area &lt; 2 km²; Target area size = 0,4 km²</td>
<td>Target area size</td>
<td>(not included in global)</td>
</tr>
<tr>
<td>HYD03</td>
<td>completeness commission</td>
<td>StandingWater</td>
<td>The area of a surface feature must be equal or greater than the target area size.</td>
<td>Evaluate Polygon Perimeter and Area &lt; 2 km²; Target area size = 0,4 km²</td>
<td>Target area size</td>
<td>0,5 km²</td>
</tr>
<tr>
<td>HYD04</td>
<td>Completeness Omission</td>
<td>SeaArea, Shore, StandingWater, LandWaterBoundary, DamOrWeir, Watercourse,</td>
<td>A feature type that is not voidable must be included in the data set.</td>
<td>Schema Compare Tool &lt; 2 km²</td>
<td>(Feature type Wetland not included)</td>
<td></td>
</tr>
<tr>
<td>HYD05</td>
<td>Logical consistency</td>
<td>HydrogeologicalObjectNatural (voidable), DamOrWeir, PumpingStation (voidable),</td>
<td>A point feature must be connected to a Watercourse line feature or WatercourseLink</td>
<td>Geometry on Geometry &lt; 2 km²</td>
<td>(Feature type Pumpingstation not included)</td>
<td></td>
</tr>
<tr>
<td>HYD06</td>
<td>topological consistency</td>
<td>Watercourse</td>
<td>A line feature of Lock and DamOrWeir must lie on the boundary of a Watercourse surface feature or of StandingWater surface feature</td>
<td>Execute SQL: [Width] &gt; 125 m; the minimum width = 125 m</td>
<td>the minimum width = 500 m</td>
<td></td>
</tr>
<tr>
<td>HYD07</td>
<td>Logical consistency</td>
<td>Lock, DamOrWeir</td>
<td>A point feature of Lock and DamOrWeir must lie on the endpoint of a Watercourse line feature</td>
<td>Geometry on Geometry &lt; 2 km²</td>
<td>(Feature type Lock not included)</td>
<td></td>
</tr>
<tr>
<td>HYD08</td>
<td>topological consistency</td>
<td>Lock, DamOrWeir</td>
<td>A Watercourse surface feature must contain at least one WatercourseLink if all of the following are true: it has at least one ingoing watercourse it has at least one outgoing watercourse</td>
<td>Geometry on Geometry &lt; 2 km²</td>
<td>(Feature type Lock not included)</td>
<td></td>
</tr>
<tr>
<td>HYD09</td>
<td>Logical consistency</td>
<td>Watercourse</td>
<td>A StandingWater surface feature must contain at least one WatercourseLink if all of the following are true: it has at least one ingoing watercourse it has at least one outgoing watercourse</td>
<td>Geometry on Geometry/Composite &lt; 2 km²</td>
<td>(Feature type Lock not included)</td>
<td></td>
</tr>
<tr>
<td>HYD10</td>
<td>topological consistency</td>
<td>WatercourseLink</td>
<td>A Wetland surface feature must contain at least one WatercourseLink if all of the following are true: it has at least one ingoing watercourse it has at least one outgoing watercourse</td>
<td>Geometry on Geometry/Composite &lt; 2 km²</td>
<td>(Feature type Lock not included)</td>
<td></td>
</tr>
<tr>
<td>HYD11</td>
<td>Logical consistency</td>
<td>Wetland</td>
<td>A Wetland surface feature must contain at least one WatercourseLink if all of the following are true: it has at least one ingoing watercourse it has at least one outgoing watercourse</td>
<td>Geometry on Geometry/Composite &lt; 2 km²</td>
<td>(Feature type Lock not included)</td>
<td></td>
</tr>
<tr>
<td>HYD12</td>
<td>topological consistency</td>
<td>ShorelineConstruction, StandingWater, LandWaterBoundary, Falls (voidable),</td>
<td>If two or more line features intersect or touch there must be a node</td>
<td>Topology Rules: Must Not Have Pseudo Nodes &lt; 2 km²</td>
<td>(Feature type Wetland not included)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Embankment (voidable), DamOrWeir, PumpingStation (voidable), Watercourse,</td>
<td></td>
<td></td>
<td>(Feature type Wetland not included)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WatercourseLink, Crossing (voidable), WatercourseLinkSequence,</td>
<td></td>
<td></td>
<td>(Feature type Wetland not included)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WatercourseSeparateCrossing (voidable), Lock (voidable)</td>
<td></td>
<td></td>
<td>(Feature type Wetland not included)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HydrogeologicalObjectNatural (voidable), Falls (voidable), Embankment</td>
<td>A point feature must not be inside one of the following: StandingWater</td>
<td>Topology Rules: Must Not Have Pseudo Nodes &lt; 2 km²</td>
<td>(Feature type Falls, Embankment, PumpingStation, Crossing, Lock not included)</td>
<td></td>
</tr>
</tbody>
</table>
Data Reviewer - Batch Job Manager

- Design Batch Jobs in Batch Job Manager
- Run job manually or automatically
ArcGIS Data Reviewer – more complex check example – water utilities
## Reporting

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

### Reporting Details

<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>7/4/2010 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>wControlValv</td>
<td></td>
<td>362</td>
<td>98%</td>
<td>3%</td>
<td>0</td>
<td>2%</td>
<td>1</td>
<td>1</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>AirRelease</td>
<td></td>
<td>2</td>
<td>98%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BackflowContr</td>
<td></td>
<td>4</td>
<td>98%</td>
<td>1%</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>BlowOff</td>
<td></td>
<td>3</td>
<td>98%</td>
<td>2%</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>wFitting</td>
<td></td>
<td>297</td>
<td>98%</td>
<td>3%</td>
<td>0</td>
<td>2%</td>
<td>1</td>
<td>1</td>
<td>Fail</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Reducer</td>
<td></td>
<td>7</td>
<td>98%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Tap</td>
<td></td>
<td>2</td>
<td>98%</td>
<td>2%</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>wLateralLine</td>
<td></td>
<td>18690</td>
<td>98%</td>
<td>3%</td>
<td>0</td>
<td>2%</td>
<td>17</td>
<td>9</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Hydrant Lateral</td>
<td></td>
<td>50</td>
<td>98%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Service Lateral</td>
<td></td>
<td>499</td>
<td>98%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Service Lateral To Meter</td>
<td></td>
<td>1</td>
<td>98%</td>
<td>2%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Pass</td>
<td>Yes</td>
</tr>
</tbody>
</table>
ArcGIS 10.3 - Improved Check Performance

Goals

★ Improve stability
★ Reduce validation time in commonly-used checks
★ Maintain equivalency with prior releases

<table>
<thead>
<tr>
<th>Relationship Type</th>
<th>Feature Class 1 (Feature Count)</th>
<th>Feature Class 2 (Feature Count)</th>
<th>Times Faster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intersects</td>
<td>Polygon (37921)</td>
<td>Polygon (67432)</td>
<td>4.1</td>
</tr>
<tr>
<td>Intersects</td>
<td>Line (105948)</td>
<td>Line (348993)</td>
<td>88.2</td>
</tr>
<tr>
<td>Intersects</td>
<td>Point (134429)</td>
<td>Point (35432)</td>
<td>117.6</td>
</tr>
<tr>
<td>Intersects</td>
<td>Point (134429)</td>
<td>Line (348993)</td>
<td>102.5</td>
</tr>
<tr>
<td>Intersects</td>
<td>Point (134429)</td>
<td>Polygon (67432)</td>
<td>59.2</td>
</tr>
<tr>
<td>Intersects</td>
<td>Line (348993)</td>
<td>Polygon (67432)</td>
<td>82.7</td>
</tr>
</tbody>
</table>

* File Geodatabase
ArcGIS Pro

- New UI, 64-bit, multi-threaded, fast graphics
- Integrated Data Quality management
Questions?

phardy@esri.com