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

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

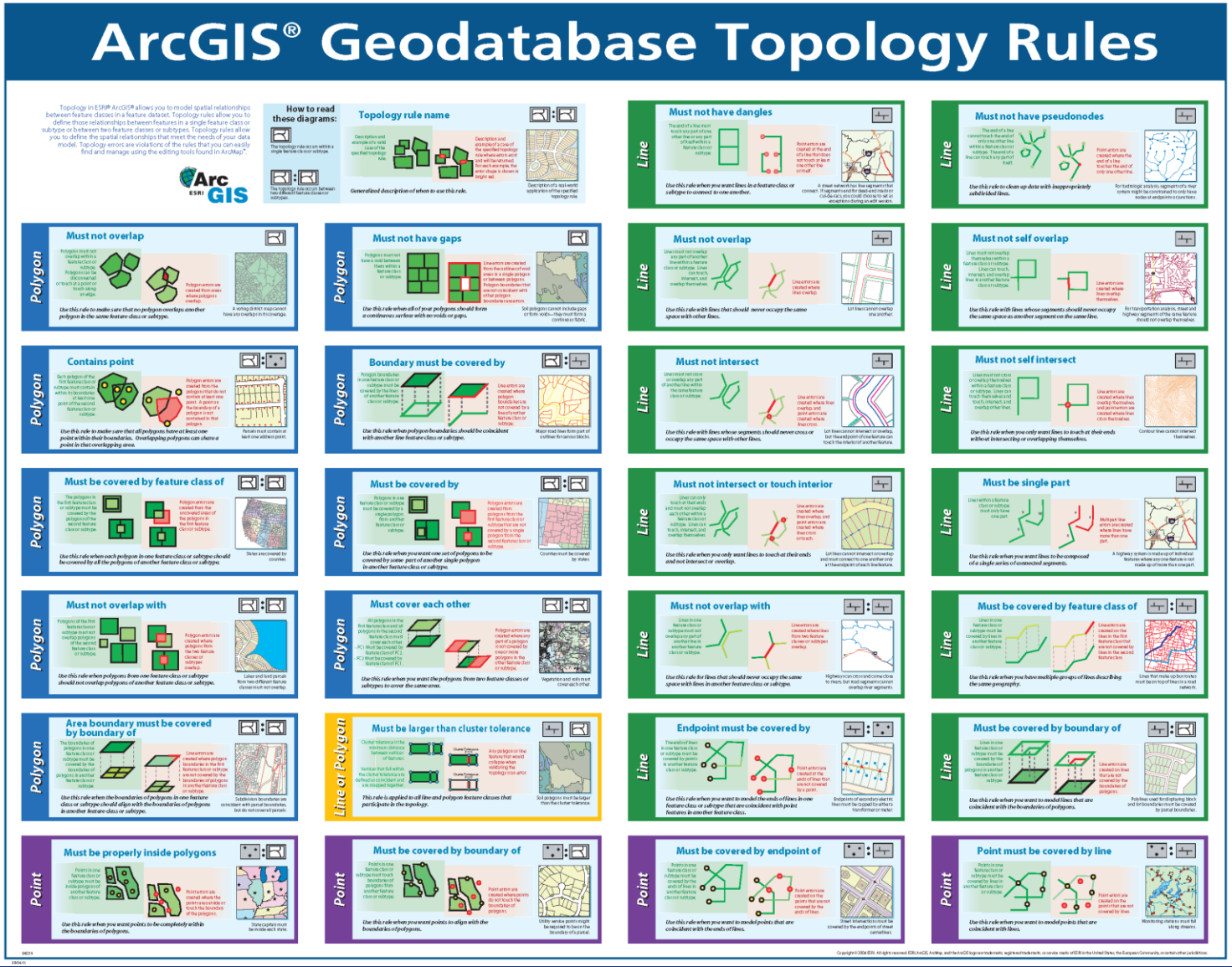
ESRI Quality Solution

- 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

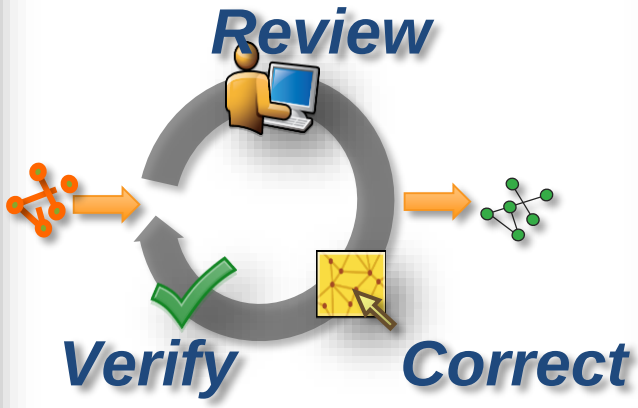
Topology Rules



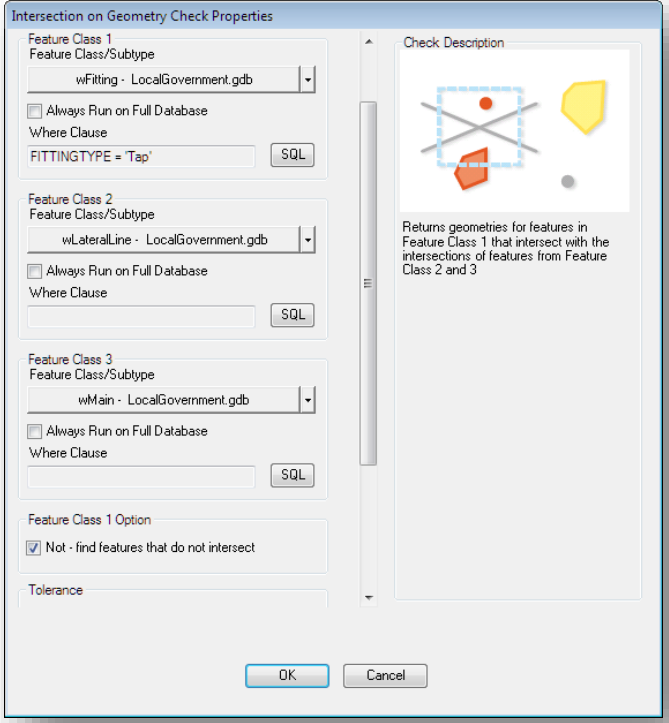
the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

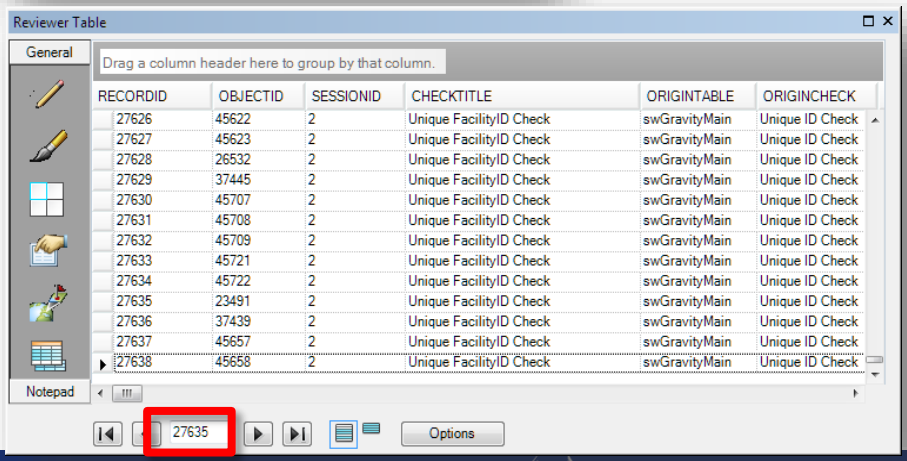
ArcGIS Data Reviewer

Many Automated checks

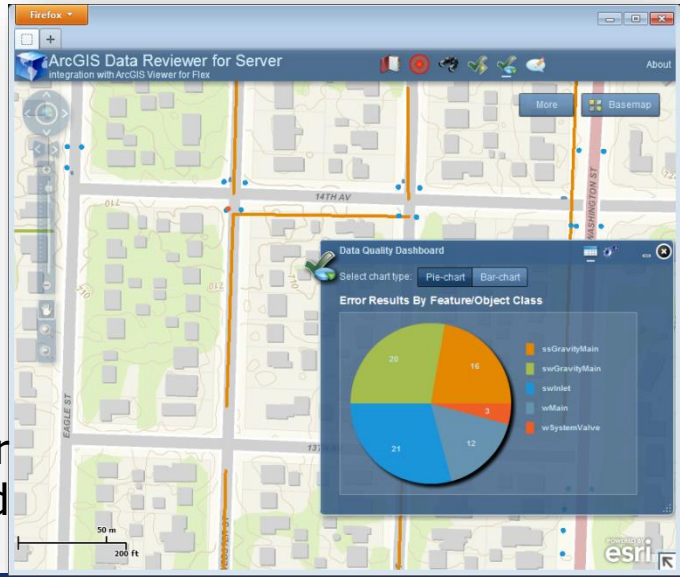


Checks configurable





Errors table Integrated



Server Dashboard

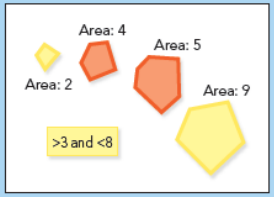
the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

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

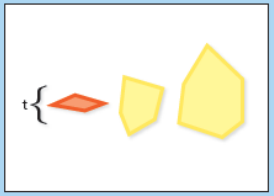


Invalid Hole Feature

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

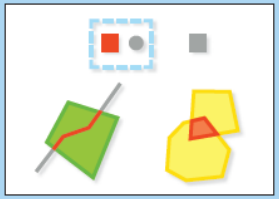


Feature on Feature Checks

Geometry on Geometry

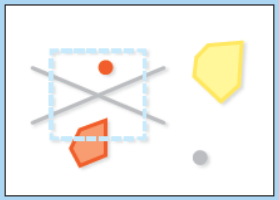
Searches for features from two different feature classes or within the same feature class that spatially interact (e.g., intersect) or are within a tolerance of each other

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

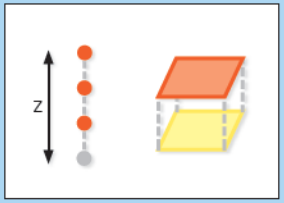
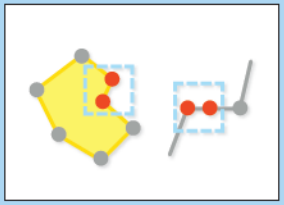
Returns overlap/gap geometries between polygon features from two feature classes that have a thinness ratio beneath a user-specified threshold; optionally requires that the overlap/gap polygons be beneath a maximum area threshold

Comparison to Topology:
Polygon must not have gaps

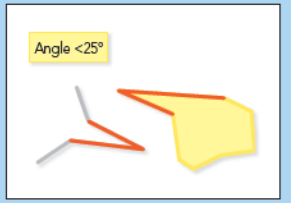
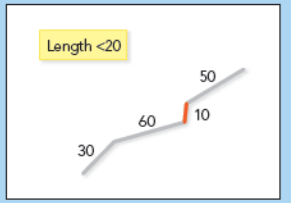
the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

ArcGIS Data Reviewer Checks

Duplicate Geometry Checks

Duplicate Geometry	Finds features of the same geometry type that are colocated	
Duplicate Vertex	Searches for vertices in selected polyline or polygon feature classes that are within a specified tolerance of each other	

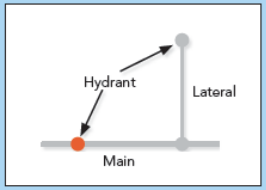
Polyline Checks

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	

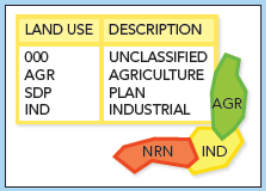
ArcGIS Data Reviewer Checks

Database Validation Checks

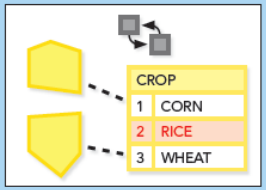
Connectivity Rules — Returns geometries for features that violate the geometric network connectivity rules



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

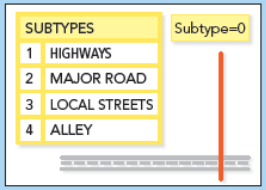
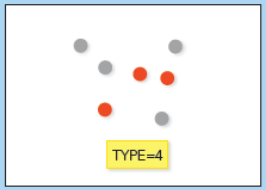


Table Checks

Execute SQL — Finds features based on a SQL query WHERE clause



Regular Expression — Finds features with attribute values that violate the regular expression

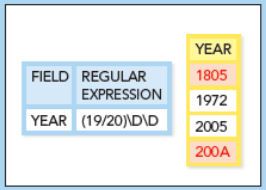
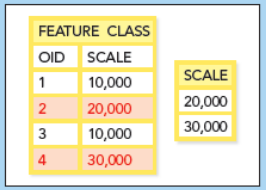
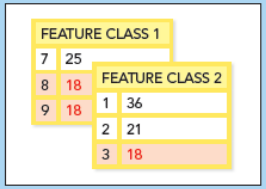


Table to Table Attribute — Returns rows whose attributes match those of a feature class or table and/or comply with a user-defined WHERE clause comparing the attributes between feature classes and/or tables



Unique ID — Checks the values of a set of fields across a set of tables and feature classes for uniqueness within a given workspace



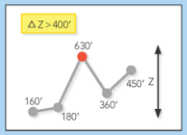
the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

ELF EUROPEAN LOCATION FRAMEWORK

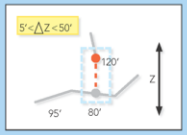
ArcGIS Data Reviewer Checks

Z-Value Checks

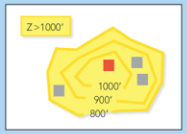
Adjacent Vertex Elevation Change
Finds vertices for polyline or polygon features with elevation (z-value) changes greater than the specified tolerance



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

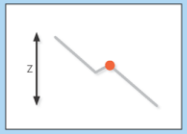


Evaluate Z-Values
Searches for features whose z-values are within a specified range



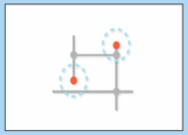
Polygon/ Ring Closed
Searches for unclosed rings in polygons based on the x-value, y-value, and z-value

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



Topology Checks

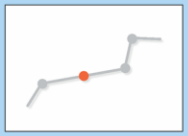
Find Dangles
Finds polyline features with nodes that are within a user-defined tolerance but not connected to other polyline or polygon features



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

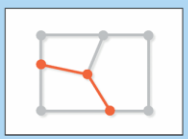
Unnecessary Nodes
Finds features that share a node and have identical attributes in editable fields

Comparison to Topology:
Line must not have pseudo nodes



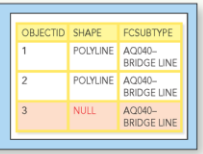
Unnecessary Polygon Boundaries
Finds adjacent polygon features that share a boundary and have identical attributes in editable fields

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

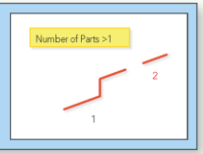
Invalid Geometry
Finds features whose geometry is empty, nothing, or not simple, as well as features with empty envelopes



OBJECTID	SHAPE	FCSUBTYPE
1	POLYLINE	AQ040–BRIDGE LINE
2	POLYLINE	AQ040–BRIDGE LINE
3	NULL	AQ040–BRIDGE LINE

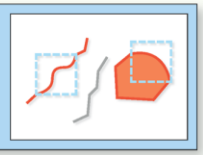
Multipart Line
Searches for polyline features with more than one part

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 Closes on Self
Finds paths and lines in polylines that touch or cross themselves

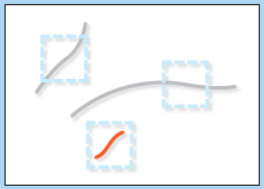
Comparison to Topology:
Line must not self intersect or self-overlap

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

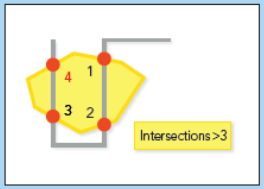
ArcGIS Data Reviewer Checks

Spatial Parameter Evaluation Checks

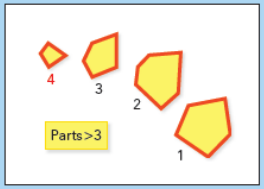
Evaluate Extent
Returns features where the extent properties (x and y) are within specified parameters



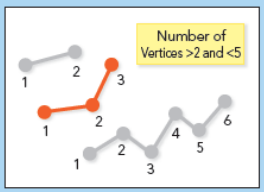
Evaluate Intersection Count
Finds vertices for polyline features in one feature class that intersect polyline or polygon features in a second feature class a specified number of times



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

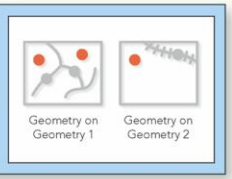


Evaluate Vertex Count
Searches for features with a vertex count that is within a specified range

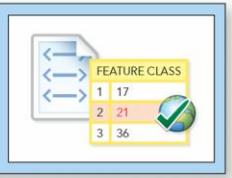


Advanced Checks

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

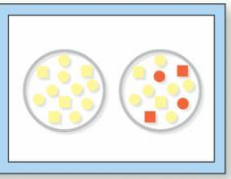


Custom
Returns records that meet the validation conditions of a custom check you developed to meet your specific requirements

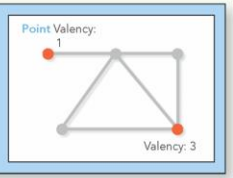


Metadata
Validates that the specified GIS item descriptions have been populated to comply with either a standards-based metadata schema or user-specified content requirements

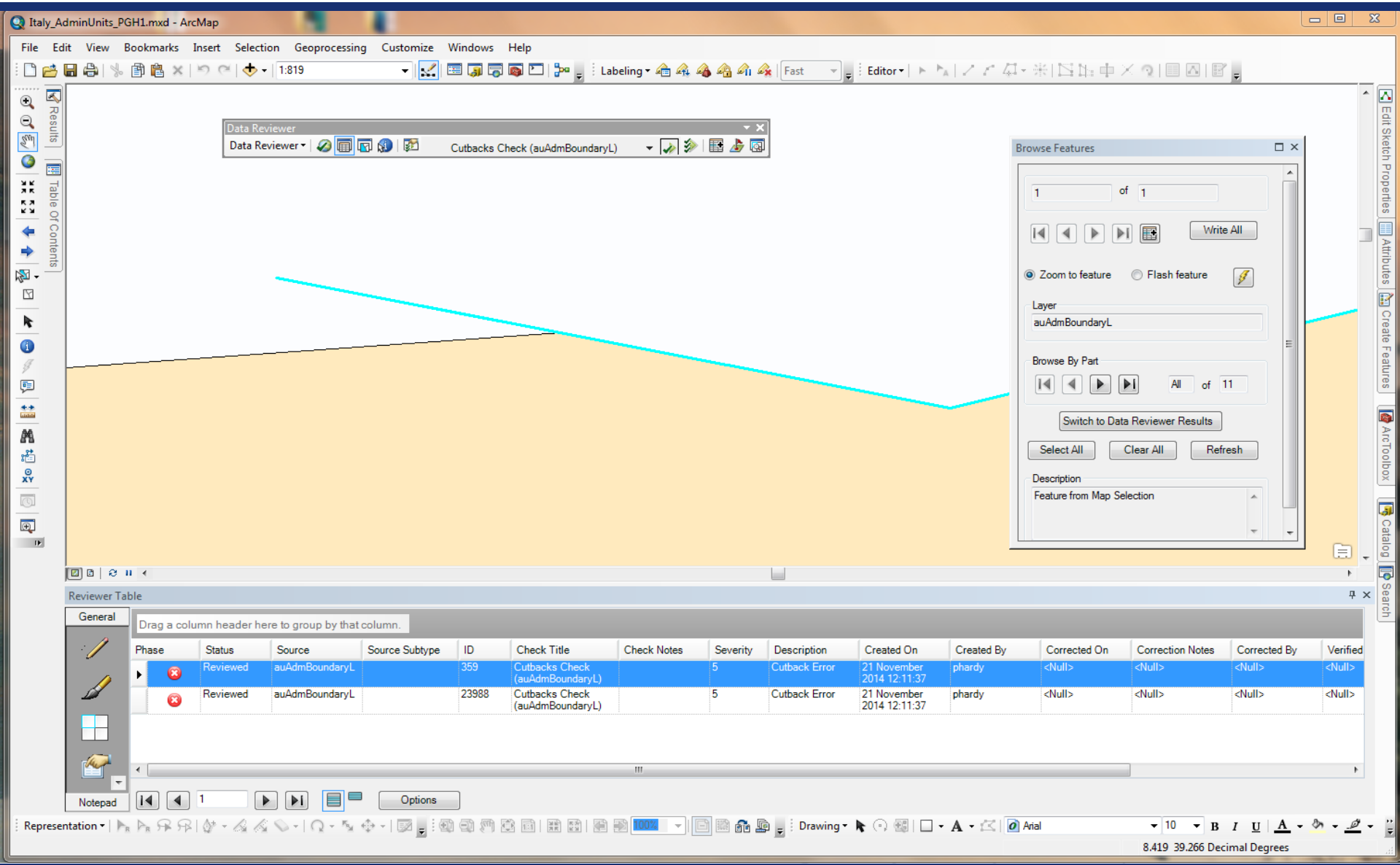
Sampling
Generates a statistical sampling of features or records from one or more layers or tables



Valency
Searches for points or nodes of linear features that intersect with a specified number of linear features



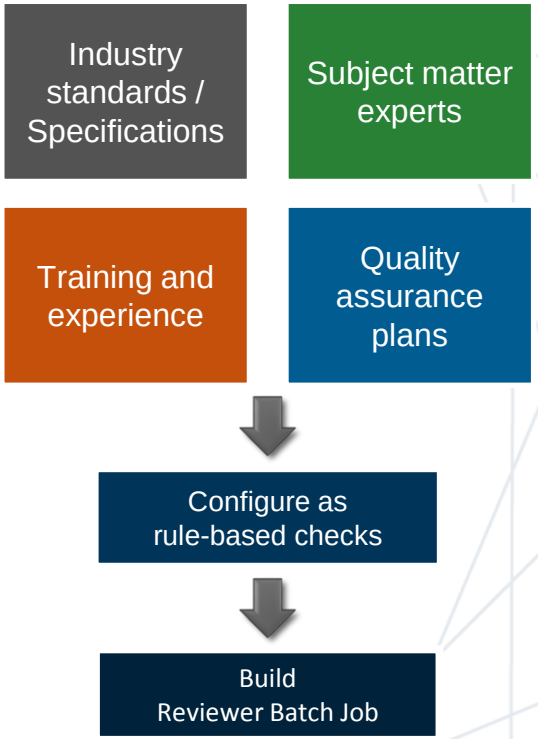
Data Reviewer – Manually trigger checks



the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Combine Multiple Rules for Batch Check

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



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

RuleSpeak to Reviewer Check

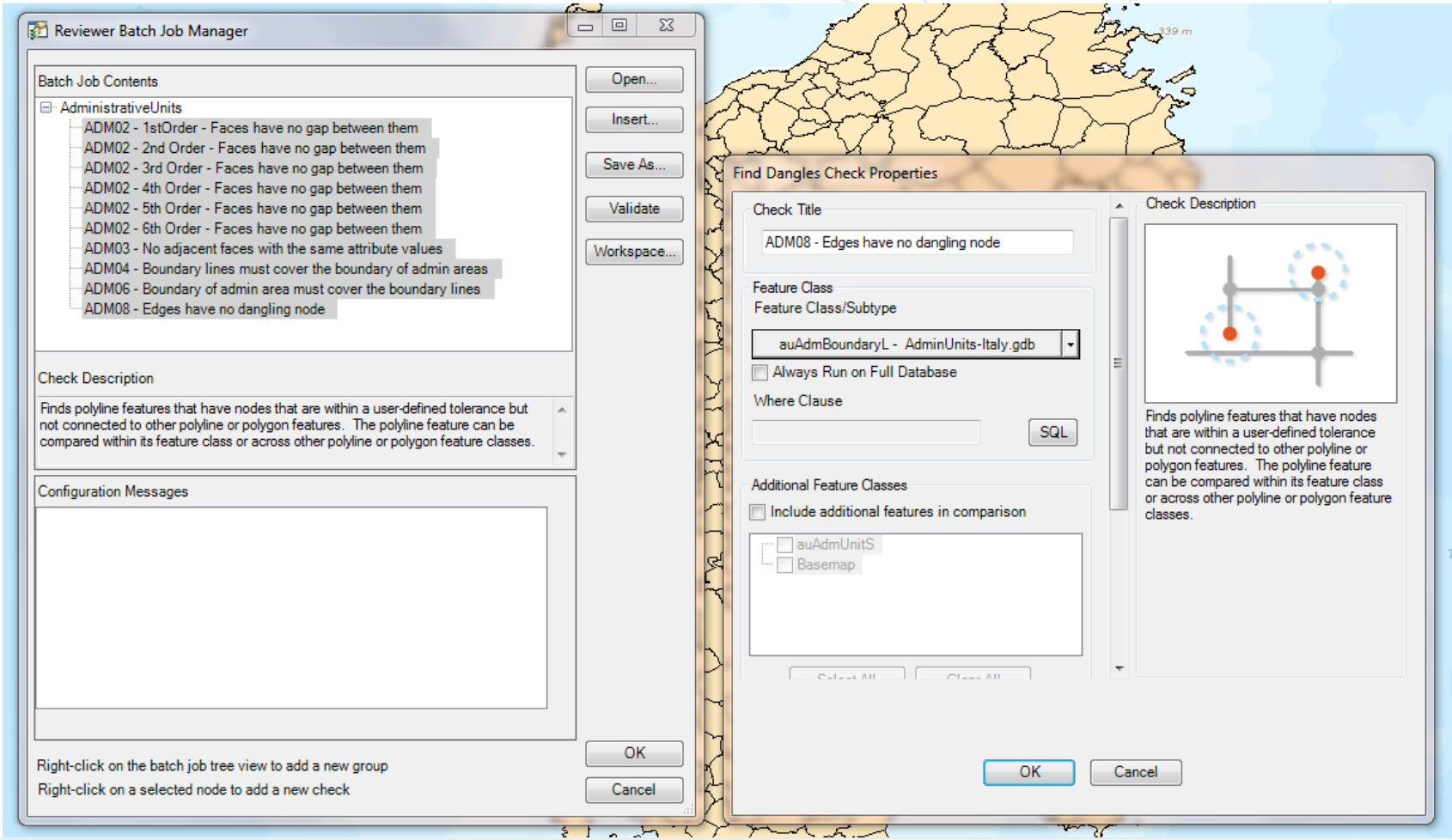
Microsoft Excel - D2 3 6_Quality_RegionalGlobal_v0 1-sanitized.xls [Compatibility Mode]

	A	B	C	D	E	F	G
1	**RuleID**	**Quality Element**	**Feature type**	**RuleSpeak rule**	**DR Check/ ArcToolbox**	**Regional**	**Global**
2	HYD01	completeness commission	Island	The area of a surface feature must be equal or greater than the **target area size.**	Evaluate Polygon Perimeter and Area	target area size = 0,4km²	Target area size = 3km²
3	HYD02	completeness commission	Wetland	The area of a surface feature must be equal or greater than the **target area size.**	Evaluate Polygon Perimeter and Area	target area size = 0,4km²	(not included in global)
4	HYD03	completeness commission	StandingWater	The area of a surface feature must be equal or greater than the **target area size.**	Evaluate Polygon Perimeter and Area	target area size = 0,4km²	Target area size = 0,5km²
5	HYD04	Completeness Ommission	SeaArea, Shore, StandingWater LandWaterBoundary, DamOrWeir, Watercourse, Wetland, Island, WatercourseLink, HydroNode WatercourseLinkSequence	A feature type that is not voidable must be included in the data set.	Schema Compare Tool		(Feature type Wetland not included)
6	HYD05	Logical consistency topological consistency	HydrogeologicalObjectNatural (voidable), DamOrWeir, PumpingStation (voidable), Lock (voidable), Watercourse, WatercourseLink	A point feature must be connected to a Watercourse line feature or WatercourseLink	Geometry on Geometry		(Feature type Pumpingstation not included)
7	HYD06	Logical consistency conceptual consistency?	Watercourse	A surface feature must have a average width equal or larger than the **minimum width**	Execute SQL: [Width] ≥ 125m	the minimum width= 125m	the minimum width= 500m
8	HYD07	Logical consistency topological consistency	Lock, DamOrWeir Watercourse StandingWater	A line feature of Lock and DamOrWeir must lie on the boundary of a Watercourse surface feature or of StandingWater surface feature	Geometry on Geometry		(Feature type Lock not included)
9	HYD08	Logical consistency topological consistency	Lock, DamOrWeir Watercourse	A point feature of Lock and DamOrWeir must lie on the endpoint of a Watercourse line feature	Geometry on Geometry		(Feature type Lock not included)
10	HYD09	Logical consistency topological consistency	Watercourse WatercourseLink	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	Geometry on Geometry/Composite		
11	HYD10	Logical consistency topological consistency	StandingWater WatercourseLink	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	Geometry on Geometry/Composite		
12	HYD11	Logical consistency topological consistency	Wetland WatercourseLink	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	Geometry on Geometry/Composite	(applys only to regional LoD)	(Feature type wetland not included)
13	HYD12	Logical consistency topological consistency	ShorelineConstruction (voidable), StandingWater, LandWaterBoundary, Falls (voidable), Embankment (voidable), DamOrWeir, PumpingStation (voidable), Watercourse, WatercourseLink, Crossing (voidable) WatercourseLinkSequence, WatercourseSeparateCrossing (voidable), Lock (voidable)	If two or more line features intersect or touch there must be a node	Topology Rules: Must Not Have Pseudo Nodes		(Feature type Falls, Embankment, PumpingStation, Crossing, Lock not included)
			HydrogeologicalObjectNatural (voidable), Falls (voidable), Embankment (voidable), DamOrWeir,	A point feature must not be inside one of the following: StandingWater			(Feature type Falls,

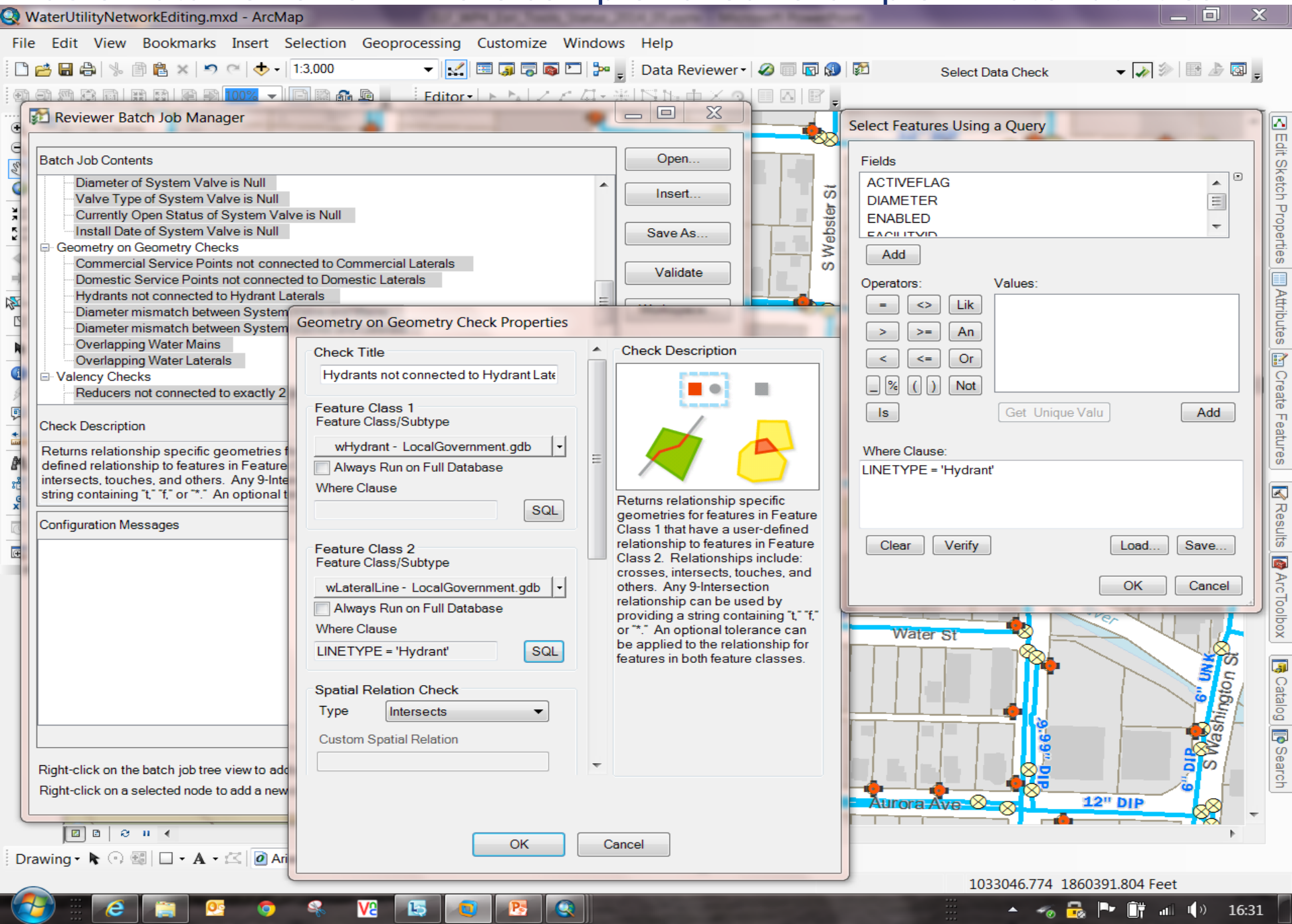
Building | GeographicalNames | Hydrography | TransportNetwork | ProtectedSites | LandCover | Miscellaneous | Cross-theme

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

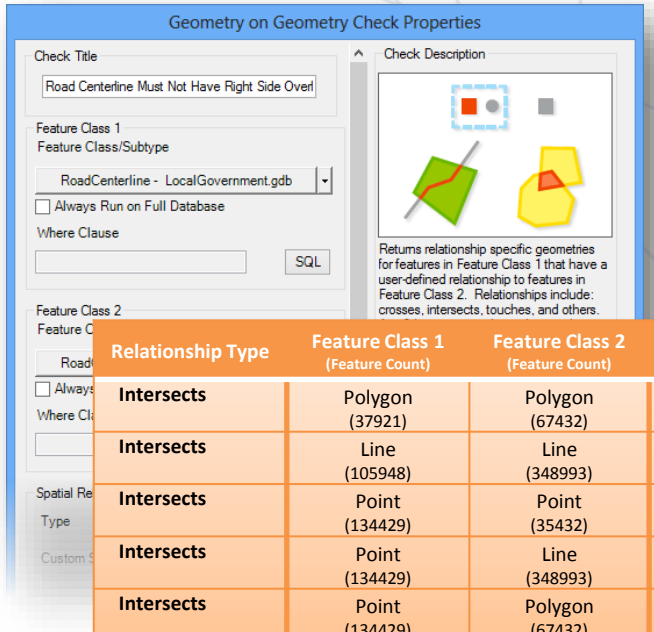
- Automated reporting of quality control results
- Available Reports
  - Automated Check (Origin Table, Subtype, Check Group)
  - Total Record Count
  - Sampling

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
7/4/2010 11:05:59 AM											
	AtlasMap		153	98%	3%	0	2%	0	0	Pass	Yes
						0			0		
	wControlValve		302	98%	3%	9	2%	1	1	Pass	Yes
		AirRelease				2			0		
		BackflowControl				4			0		
		BlowOff				3			1		
	wFitting		297	98%	3%	9	2%	1	2	Fail	Yes
		Reducer				7			0		
		Tap				2			2		
	wLateralLine		18690	98%	3%	544	2%	17	9	Pass	Yes
		Hydrant Lateral				50			0		
		Service Lateral				493			8		
		Service Lateral To Meter				1			1		

ArcGIS 10.3 - Improved Check Performance

Goals

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

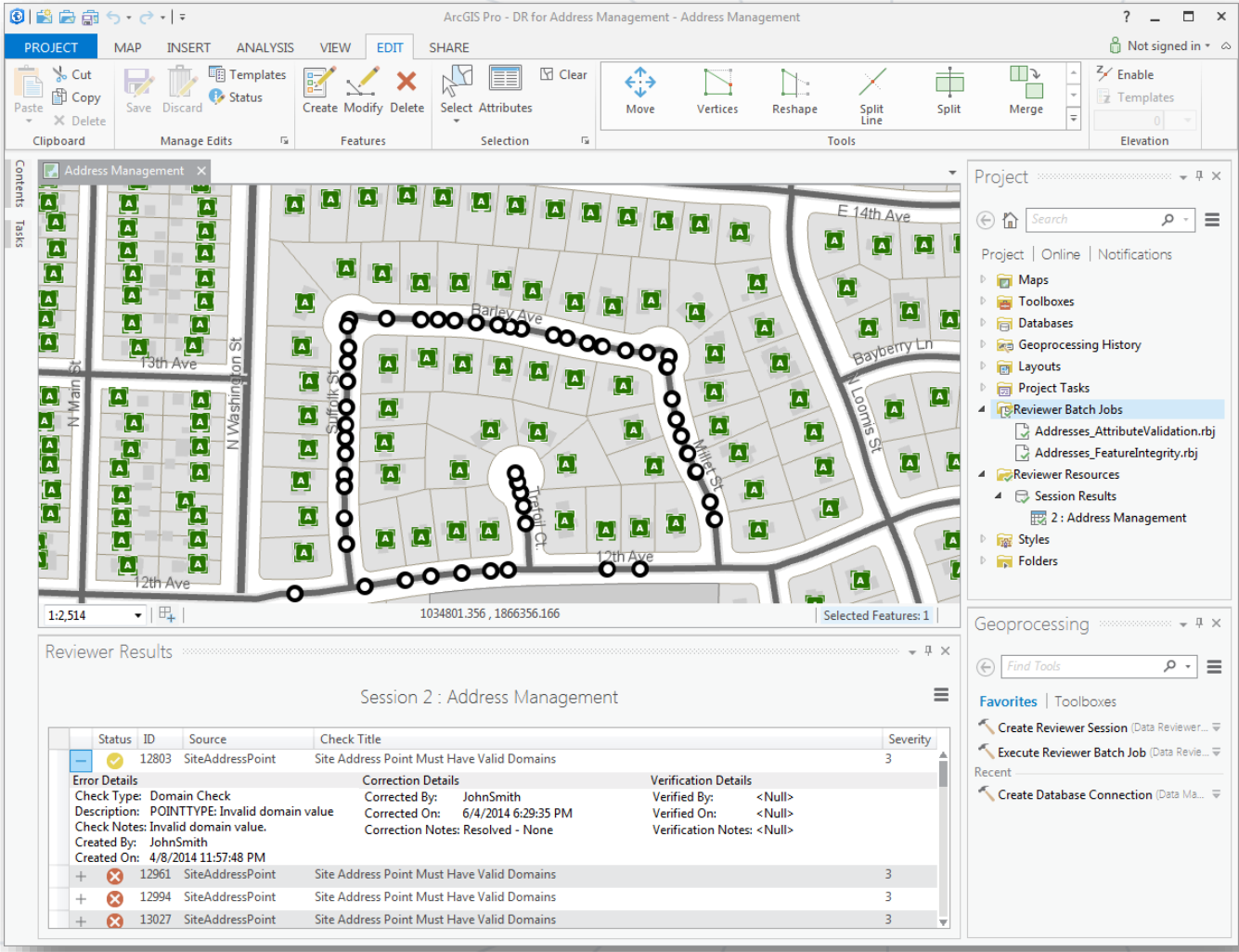


Relationship Type	Feature Class 1 (Feature Count)	Feature Class 2 (Feature Count)	Times Faster
Intersects	Polygon (37921)	Polygon (67432)	**4.1**
Intersects	Line (105948)	Line (348993)	**88.2**
Intersects	Point (134429)	Point (35432)	**117.6**
Intersects	Point (134429)	Line (348993)	**102.5**
Intersects	Point (134429)	Polygon (67432)	**59.2**
Intersects	Line (348993)	Polygon (67432)	**82.7**

* File Geodatabase

ArcGIS Pro

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



the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)

Questions?

phardy@esri.com

the Competitiveness and Innovation framework Programme (CIP)
ICT Policy Support Programme (PSP) Call 6 (Grant 325140)