ELF WP4 Data Reviewer tests

Celia Sevilla Sánchez
Gunhild Lönnberg

Contents

- Introduction
- ELF Quality Rules
- ArcGIS Data Reviewer
  - IGN Data
  - Process
  - Results
- Problems
- Feedback to WP2
- Conclusions

Introduction

- The goal of ELF project is to deliver the European Location Framework required to provide up-to-date, authoritative, interoperable, cross-border, reference geo-information for use by the European public and private sectors (2013-2016)
- There are 30 companies working together into a consortium: mapping and cadastral agencies, software developers, academia, etc.
- The project is organized in 7seven working packages (WP)
- WP4: Geo-tools -> to integrate reference data and harmonize them
  - Edge matching
  - Generalization
  - Change detection
  - Quality assurance
    - ESRI: ArcGIS Data Reviewer -> preconfigured batch rules following ELF quality controls (Annex A spreadsheet)
      - IGN-Spain: Test Data Reviewer using own data
      - Cadaster Spain
      - Lantmäteriet: Sweden
    - 1Spatial cloud
    - PP Repair
    - ETF: network services

ELF Quality Rules

- Being created in RuleSpeak by WP2 (specifications)
  - RuleSpeak: a set of practical guidelines for expressing rules in clear, unambiguous, well-structured English.
  - ELF Quality rules for every different level:
    - Master (<100k):
      - LoD0: < 5k
      - LoD1: 5K-25K
      - LoD2: 25K-100K
    - Regional (100k-500k)
    - Global (>500k)
  - Rules for: Administrative Units (7), Buildings (9), Cadastral Parcels (2), Geographical Names (9), Hydrography (28), Transport Network (16), Protected Sites (1), Land Cover (2), Cross-theme (12).
  - Automated rules: completeness, topological consistency and logical consistency.

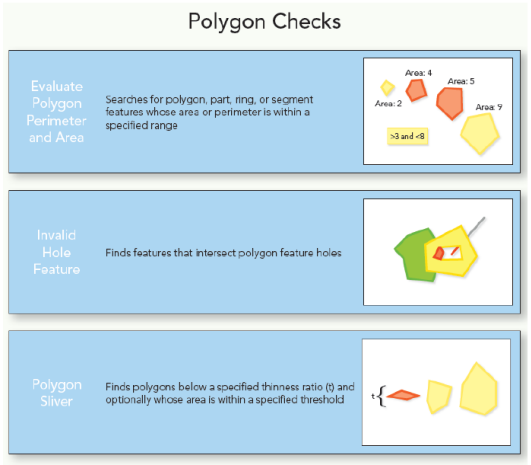
- ESRI has gone through spreadsheet
  - Added a column with the proper check to be used for each rule
  - Implemented some of the rules as batch job for: AU, HY and RT
  - Gave a webinar to explain how to change rules and create new ones

ELF Quality Rules

RuleID	Quality Element	Descripion of rule	Feature type	RuleSpeak rule	DR Check/ ArcToolbox	Master LoD1	Master LoD2	Regional	Global
HYD01	completeness commission	The minimum allowed area size	Island	The area of a surface feature must be greater than or equal to the TargetAreaSize.	Evaluate Polygon Perimeter and Area			TargetAreaSize = 0,4km²	TargetAreaSize = 3km²
HYD02	completeness commission	The minimum allowed area size	Wetland	The area of a surface feature must be greater than or equal to the TargetAreaSize.	Evaluate Polygon Perimeter and Area			TargetAreaSize. = 0,4km²	(not included in global)
HYD03	completeness commission	The minimum allowed area size	StandingWater	The area of a surface feature must be greater than or equal to the TargetAreaSize.	Evaluate Polygon Perimeter and Area			TargetAreaSize. = 0,4km²	TargetAreaSize. = 0,5km²
HYD04.1 HYD04.2	Completeness Ommission	Mandatory feature classes must have features:	SeaArea Shore StandingWater LandWaterBoundary DamOrWeir Watercourse Wetland Island	HYD04.1 A feature type must be considered 'Mandatory' if it is in the MandatoryFeaturesList: HYD04.2 At least one feature of each Mandatory feature type must be in the data set.	Schema Compare Tool	MandatoryFeaturesList: * SeaArea *StandingWater *LandWaterBoundary *DamOrWeir *Watercourse	MandatoryFeaturesList: * SeaArea *StandingWater *LandWaterBoundary *DamOrWeir *Watercourse	MandatoryFeaturesList: * SeaArea *Shore *StandingWater *LandWaterBoundary *DamOrWeir *Watercourse *Wetland *Island	MandatoryFeaturesList: * SeaArea *Shore *StandingWater *LandWaterBoundary *DamOrWeir *Watercourse *Island
HYD05.1 HYD05.2 HYD05.3	Logical consistency topological consistency	Connected nodes must be connected to endpoints of a Watercourse	HydrogeologicalObjectNatural (voidable) DamOrWeir PumpingStation (voidable) Lock (voidable) Watercourse WatercourseLink	HYD05.1 A point feature with type in the WaterPointFeatureList must be considered a Water Point Feature. HYD05.2 A feature with type in the following list must be considered a Watercourse Feature: * Watercourse * WatercourseLink. HYD05.3 A Water Point Feature must intersect at least one Watercourse Feature	Geometry on Geometry	WaterPointFeatureList: * DamOrWeir * Lock.	WaterPointFeatureList: * DamOrWeir * Lock.	WaterPointFeatureList: * HydrogeolocialObjectNatural * DamOrWeir *PumpingStation * Lock.	WaterPointFeatureList: * HydrogeolocialObjectNatural * DamOrWeir * Lock.
				HYD06.1 The Average Width of a surface Watercourse must be computed as					

ArcGIS Data Reviewer

- There are more than 40 configurable checks (automated):
  - Database validation
  - Table validation
  - Spatial parameter evaluation
  - Topology
  - Polygon
  - Polyline
  - Z-Value
  - Feature on feature
  - Duplicate geometry
  - etc.



http://www.esri.com/library/fliers/pdfs/arcgis-data-reviewer-checks.pdf

- ADR is a user-configurable tool with generic checks that can be configured for a wide range of data schemas and data content

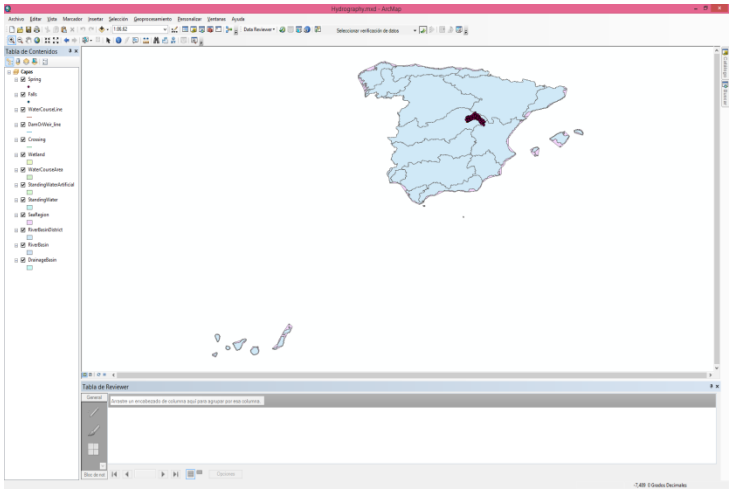
IGN data

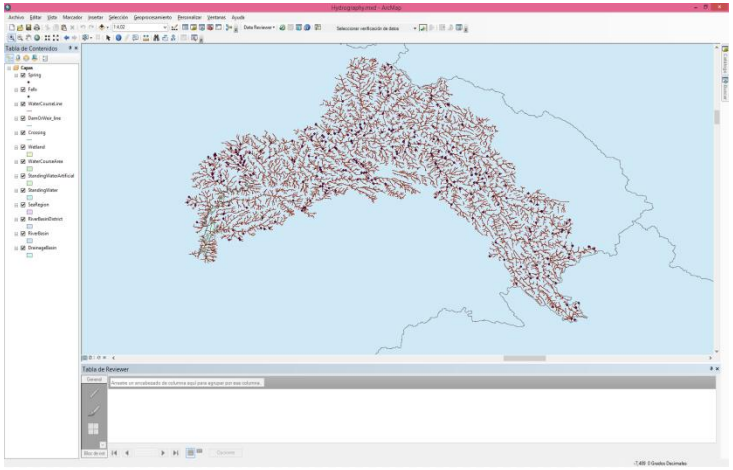
- IGN purpose is to validate the data quality of existing data sets (INSPIRE compliant while ELF compliant are available) according with ELF data quality rules expressed by WP2, using ArcGIS Data Reviewer.
- The objective is to check if ESRI tools are enough to ensure quality according to ELF specifications.
- Data source:
  - **Administrative Units:** data extracted from WFS that contain the administrative units of Spain at 1:25K scale
    - http://www.ign.es/wfs-inspire/unidades-administrativas
  - **Hydrography:** data extracted from the future GRI_HY data base at 1:5K
  - ***Transport Network:** data extracted from WFS at 1:100K
    - http://www.ign.es/wfs-inspire/transportes-btn100
  - ***Geographical Names:** data extracted from WFS:
    - http://www.ign.es/wfs-inspire/ngbe
  - ***Land Cover:** data extracted from WFS
    - http://www.ign.es/wfs-inspire/ocupacion-suelo

* To be done

IGN data: Hydrography

- Falls: 1 entity
- SpringPoint: 266 entities
- WaterCourseLine: 9444 entities
- DamOrWeirLine: 7 entities
- Crossing: 267 entities
- WaterCourseArea: 6 entities
- StandingWaterArtificial: 345 entities
- StandingWater: 8 entities
- Wetland: 2 entities
- DrainageBasin: 118 entities
- RiverBasin: 118 entities
- RiverBasinDistrict: 25 entities
- SeaArea: 22 entities





Process

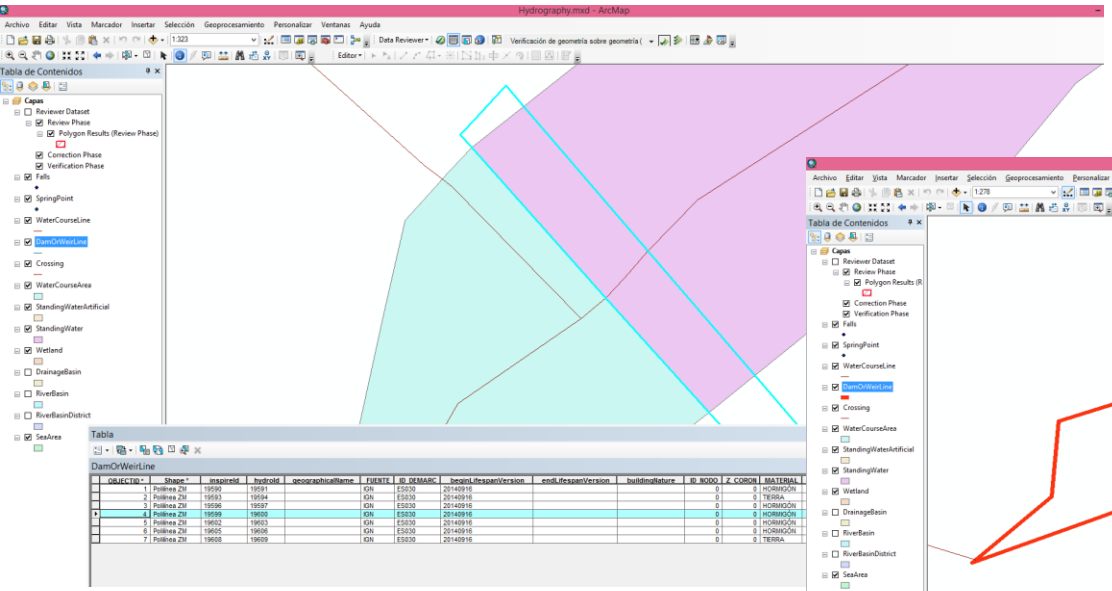
1. Transform IGN data into a geodatabase (*.gdb) and create topology if it is needed (Administrative Units)
2. Create a new project in ArcMap (*.mxd)
3. Create a separate geodatabase for the Reviewer Session: store the results of the checks
4. Open a new or an existing Reviewer Session
5. Open the Batch Rule Manager tool and load the batch rules (*.rbj)
6. Adapt the rules to IGN geographic objects: folder, entitiy name, checks, etc.
7. Run the batch rules chosen for the full database, window extension, the selected object or the object that has changed.
8. Open the reviewer table and organize items
9. Start visual revision of th errors

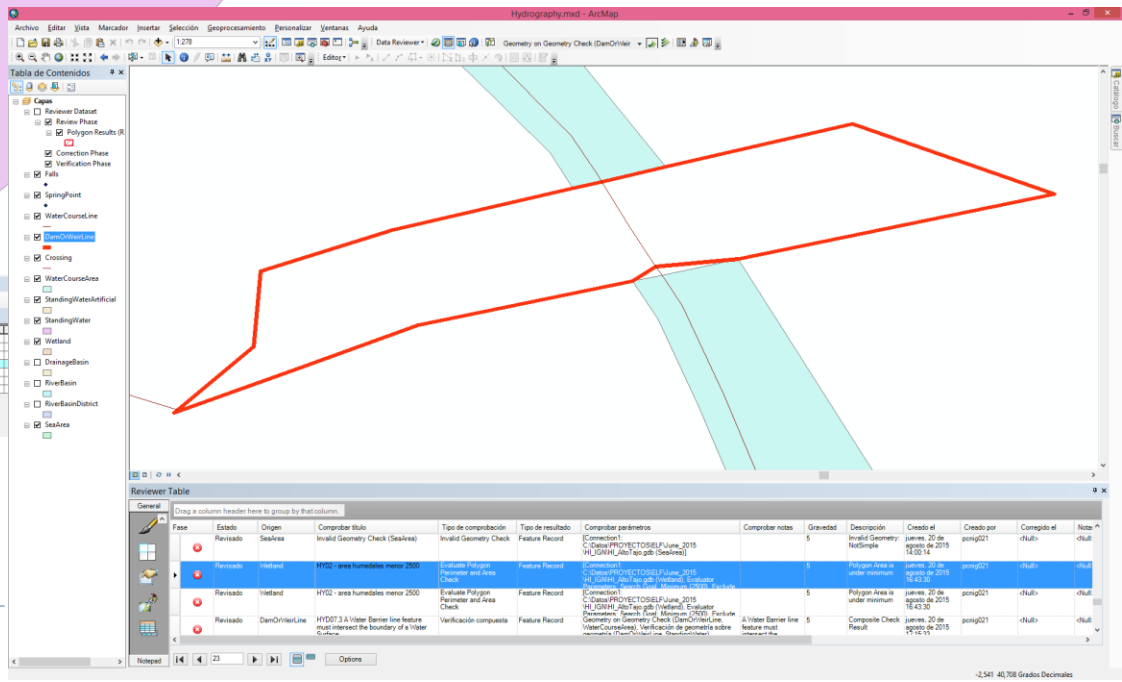
Results HY

- ESRI provides 8 rules
  - 4 new were easily created (10 run as total)
  - 2 gave non expected results (reported to ESRI)
  - 4 needs more development (Model Builder to create series of Geoprocessing Tools). 1 provided as example.
  - 6 does not apply to the LoD0 scale (even we think they were, p.e.: the minimun allowed area size Standing waters)
  - 1 was duplicated
- The first run it took one minute and detected 11 errors
- Afterwords, in order to check some of the rules without errors we created some errors in the gdb and even new entities to check entities in the coast (SeaLevel, ForeShore, etc.)
- Some of the rules were included in our own controls

Results HY

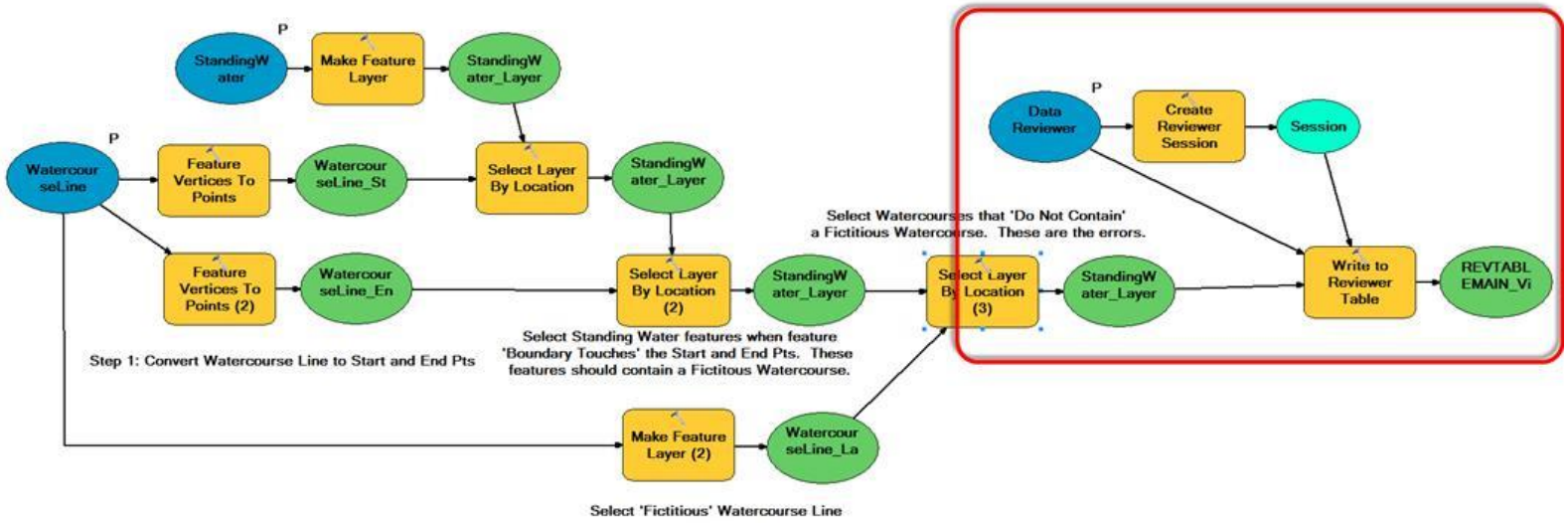
- HYD07. Logical consistency (conceptual consistency) Dam lines must be covered by lake or reservoir boundaries→ wrong results





Results HY

- HYD10. Logical consistency (topological consistency)
  - A StandingWater surface feature must contain at least one WatercourseLink if all of the following are true:
    - it has at least one ingoing watercourse



Results HY

- HYD05 and HYD08 are the same:
  - HYD05: Logical consistency. Topological consistency
    - A Water Point Feature (Lock or DamOrWeir) must intersect at least one Watercourse Feature
  - HYD08: Logical consistency. Topological consistency
    - A Water Barrier point feature (Lock or DamOrWeir) must intersect an endpoint of a Watercourse line feature.

Results HY

- HYD013: We consider it wrong
  - Logical consistency. Topological consistency
    - Nodes shall not be contained by areas HYD08: Logical consistency. Topological consistency -> A point feature from the following: ***Falls**
      *DamOrWeir
      *Crossing
      *Lock
      *Sluice
      *Rapids
      *Ford
    - must not be within one of the following:
      *StandingWater
      *Watercourse
      *SeaArea

Results HY

- This ELF control has not been found in INSPIRE specifications and does not make sense because of the following:
  - Falls and Rapids will be located inside WaterCourseAreas or on WaterCourseLines
  - Crossing: nodes will be located just on the WaterCourseLines and inside WaterCourseAreas
  - DamOrWeir, Lock, Sluice: the node will be located in the boundary of the WaterCourseArea.

Results HY

- Rules that should be applied to LoD0:
  - The minimum allowed area size for Wetland, StandingWater and Island
  - There must be a node at each point of intersection between an intersecting pair of features from LineFeaturesList
  - Coastline must be covered by the boundary of sea area
  - Coastline must not have gaps
  - Foreshore areas must overlap either sea area or watercourse area
  - Foreshore area must not overlap with island

Problems

- Rules were provided in ArcGIS 10.3 and they didn't work in 10.2 version -> update licence services
- New tool: learning time
- Geoprocessing tools requiere more knowledge
- It is necessary to create a new geodatabase (*.gdb) from the existing data and also create topology -> a problem if the data are in another format
- It takes long or it does not work with the full database
  - It is usefull for production units
- In the case of data obtained from WFS there weren't errors so we have to create them to check if the tool was able to detect them

Feedback to WP2

- New rules to be considered:
  - Network continuity
  - Direction of the water flow
  - DTM consistent
- Rules already existing that we consider must be applied to LoD0

Conclusions

- Good experience from our point of view to test other tools and to include rules in our own quality controls
- Data Reviewer is recommended if you are normally working with ArcGIS:
  - Known enviroment and data already in gdb
- ArcGis Data Reviewer is a good quality check tool for using in the production system:
  - it is possible to store the results and this can be review and justify (results are stored in a gdb)
  - It is also easy to create the most part of the rules, although some of them need more development.
  - And it is useful to have some rules stored that can be share with other agencies or with the companies that produce the data.
- It is fast if the sample of the data is not too big, so it can be applied to production units.
- It is also easy to make some changes in a rule and/or in the data and pass it again and only over the changes or the window extension.