Python – It’s Like Having an Extra Person to QA/QC Data

Jennifer M. Holland, GISP
28 June 2016
Project Details

- U.S. Army Reserve (USAR)
  - Army requirement to have complete data
    - Not met
  - Need completed data to manage their facilities
- Data collection project
  - 473 sites
    - On site data collection
      - GIS (map grade GPS)
      - Space and Facility Utilization
GIS

- SDSFIE version 2.6 data schema
  - 14 feature datasets
    - 28 feature classes
Fort Bliss
QA/QC Process

- Requirements
  - Must QA/QC at least 20% of the GIS data
    - Spatial
    - Attributes
- Interested in automating the process as much as possible
  - Python
Check for Empty Feature Classes

This tool looks for feature classes that are supposed to be present based upon a list of expected feature classes. It works recursively through subfolders and/or feature datasets.
Check for Empty Fields
Check for Geometry Errors

This tool will look for geometry errors and output any errors found to a dbf table. The tool works recursively through sub folders and/or feature datasets.
Check for Missing Layers

This tool looks through layers to see if one is missing from a semi-colon delimited list of expected layers. The tool works recursively through sub folders and/or feature datasets.
Check Spatial Data

This script exports the projection information of a feature or raster dataset to either a text file or Excel file. This script is recursive which means it will look down through folders and subfolders for data.
Check Spatial Location

This tool checks to see if layers fall within a specified spatial location. It works recursively through sub folders and/or feature datasets.
USAR Tool

- Create FGDB
- Create feature dataset
- Create topology
- Add feature classes
- Load topology rules
- Validate topology
Totals

- ~60,000 features
  - Manual review
    - 20% review
    - 333 hours
    - 8.5 weeks
  - Automated review
    - 90% review
    - 400 hours
    - 10 weeks
Lessons Learned

- Data Reviewer
  - Has *most* of the functionality needed out of the box
  - Has some additional functionality that was not needed for this project

- USAR Python Tools
  - Topology
  - Simple
  - Can be targeted to *exactly* what is needed with customization
Conclusion

- Automating the QA/QC process saved:
  - Time
  - Money
  - Data Quality (most important)
    - More data reviewed (+/- 90%)
    - More time spent visually inspecting data
- Great alternative to Data Reviewer
- Tools available for other projects
  - “Toolified”
Questions

Python!
It’s like having an extra person to QA/QC data

Jennifer M. Holland, GISP
Geographer
U.S. Army Corps of Engineers
Fort Worth District
jennifer.m.Holland@usace.army.mil