Python Scripting for Regional Land Use Data Management and QC Workflow

2015 ESRI User Conference | July 21, 2015 | San Diego, CA

Jung Seo, Frank Wen, Simon Choi and Tom Vo

Research & Analysis
Southern California Association of Governments
Southern California Association of Governments (SCAG)
Southern California Association of Governments (SCAG)

Nation’s largest Metropolitan Planning Organization (MPO)

6 counties and 191 cities

18.4 million people within 38,000+ square miles

GRP in 2013: $924 Billion (16th largest economy in the world)
Overview

- Background
- Objectives
- Methodology
- Conclusions
BACKGROUND
2016 RTP/SCS and Senate Bill 375

- 2016-2040 Regional Transportation Plan / Sustainable Communities Strategy (RTP/SCS)
  - A long-range transportation plan
- SB375 – California’s Climate Protection Act
  - Integration of transportation, land use, housing and environmental planning to meet the regional GHG emission reduction targets
Bottom-Up Local Input Process

- Bottom-up local input process
  - Participation and cooperation of all 197 local government partners
- Review by local jurisdictions on SCAG’s land use and resource areas datasets
  - SCAG Data/Map Book
  - One-on-one meeting
- Collect data changes, answer questions, provide technical guidance
Regional Land Use Database

- Development of regional land use database in preparation for the 2016 RTP/SCS
- Updated and reviewed thru the bottom-up local input process
- Base data for integrated growth forecast, scenario planning model, planning and policy analysis, etc.
Regional Land Use Database
(Dataset Overview)

- Land use types
  - General plan land use and zoning (GPZN)
  - Existing land use (LU)
  - Specific plan land use (SP)
- Geographic level
  - Parcel data at city and county level
- Standardized land use code
  - 40 general plan land use classification
  - 110 existing land use classification
### Table 2: 2012 SCAG Existing Land Use Codes - Legend

<table>
<thead>
<tr>
<th>Legend</th>
<th>Land Use Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Single Family Residential" /></td>
<td>1110 Single Family Residential</td>
</tr>
<tr>
<td><img src="image2" alt="Multi-Family Residential" /></td>
<td>1120 Multi-Family Residential</td>
</tr>
<tr>
<td><img src="image3" alt="Mobile Homes and Trailer Parks" /></td>
<td>1130 Mobile Homes and Trailer Parks</td>
</tr>
<tr>
<td><img src="image4" alt="Mixed Residential" /></td>
<td>1140 Mixed Residential</td>
</tr>
<tr>
<td><img src="image5" alt="Rural Residential" /></td>
<td>1150 Rural Residential</td>
</tr>
<tr>
<td><img src="image6" alt="General Office" /></td>
<td>1210 General Office Use</td>
</tr>
<tr>
<td><img src="image7" alt="Commercial and Services" /></td>
<td>1220 Retail Stores and Commercial Services</td>
</tr>
<tr>
<td><img src="image8" alt="Facilities" /></td>
<td>1240 Public Facilities</td>
</tr>
<tr>
<td><img src="image9" alt="Education" /></td>
<td>1260 Educational Institutions</td>
</tr>
<tr>
<td><img src="image10" alt="Military Installations" /></td>
<td>1270 Military Installations</td>
</tr>
<tr>
<td><img src="image11" alt="Industrial" /></td>
<td>1300 Industrial</td>
</tr>
</tbody>
</table>

### Transportation, Communications, and Utilities
- 1400 Transportation
- 1410 Airports
- 1411 Pipelines
- 1412 Railroads
- 1413 Freeways and Major Roads
- 1414 Park and Ride Lots
- 1415 Bus Terminals and Yards
- 1416 Truck Terminals
- 1417 Harbor Facilities
- 1418 Navigation Aids
- 1420 Communication Facilities
- 1430 Utility Facilities
- 1440 Maintenance Yards
- 1441 Bus Yards
- 1442 Rail Yards
- 1450 Mixed Transportation
- 1460 Mixed Transportation and Utility

### Mixed Commercial and Industrial
- 1500 Mixed Commercial and Industrial

### Mixed Residential and Commercial
- 1600 Mixed Residential and Commercial

### Open Space and Recreation
- 1800 Open Space and Recreation
- 1810 Golf Courses
- 1830 Local Parks and Recreation
- 1840 Regional Parks and Recreation
- 1845 Cemeteries
- 1850 Wildlife Preserves and Sanctuaries
- 1860 Specimen Gardens and Arboretums
- 1870 Beach Parks
- 1880 Other Open Space and Recreation

### Agriculture
- 2000 Agriculture
- 2100 Cropland and Improved Pasture Land
- 2110 Irrigated Cropland and Improved Pasture Land
- 2120 Non-Irrigated Cropland and Improved Pasture Land
- 2200 Orchards and Vineyards
- 2300 Nurseries
- 2400 Dairy, Intensive Livestock, and Associated Facilities
- 2500 Facility Operations
- 2600 Other Agriculture
- 2700 Horse Ranches

### Vacant
- 3000 Vacant
- 3010 Vacant Undeveloped
- 3020 Abandoned Orchards and Vineyards
- 3030 Vacant With Limited Improvements
- 3400 Beaches (Vacant)
- 3410 Urban Vacant

### Water
- 4000 Water
- 4100 Water, Undeveloped
- 4200 Harbor Water Facilities
- 4300 Marine Water Facilities
- 4400 Water Within a Military Installation
- 4500 Area of Inundation (High Water)

### Under Construction
- 1700 Under Construction
Regional Land Use Database
(Attribute Information)

- General plan and zoning (GPZN)
  - City’s GP designations
  - City’s zoning codes
  - SCAG’s standardized GP code
  - Residential density (average, min/max)
  - Year of adoption/amendment

- Existing land use (LU)
  - SCAG’s standardized LU code

- SCAGUID12, APN, county, city, acreage, etc.
### GPZN Attribute Table

<table>
<thead>
<tr>
<th>APN</th>
<th>FIPS</th>
<th>X_CENTER</th>
<th>Y_CENTER</th>
<th>Shape_Leng</th>
<th>Shape_Area</th>
<th>CITY</th>
<th>COUNTY</th>
<th>DENSITY</th>
<th>LOW</th>
<th>HIGH</th>
<th>YEAR_ADOPT</th>
<th>ZONE_CODE</th>
<th>CITY_GP_CO</th>
<th>NOTES</th>
<th>ACRES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
</tr>
</tbody>
</table>

### LU Attribute Table

<table>
<thead>
<tr>
<th>FID</th>
<th>Shape</th>
<th>SCAGUID12</th>
<th>APN</th>
<th>FIPS</th>
<th>X_CENTER</th>
<th>Y_CENTER</th>
<th>Shape_Leng</th>
<th>Shape_Area</th>
<th>CITY</th>
<th>COUNTY</th>
<th>LU12</th>
<th>ACRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Polygon</td>
<td>025037125</td>
<td>037-140-011</td>
<td>06925</td>
<td>63898.113907</td>
<td>36685.4931863</td>
<td>3878.012083</td>
<td>3979.691961</td>
<td>Brawney</td>
<td>Imperial</td>
<td>1433</td>
<td>0.00000</td>
</tr>
<tr>
<td>1</td>
<td>Polygon</td>
<td>025037126</td>
<td>037-140-011</td>
<td>06925</td>
<td>64368.927038</td>
<td>36685.4931863</td>
<td>3878.012083</td>
<td>3979.690568</td>
<td>Brawney</td>
<td>Imperial</td>
<td>1433</td>
<td>0.00000</td>
</tr>
<tr>
<td>2</td>
<td>Polygon</td>
<td>025037127</td>
<td>037-140-011</td>
<td>06925</td>
<td>63531.762981</td>
<td>36685.4931863</td>
<td>3878.012083</td>
<td>3979.690568</td>
<td>Brawney</td>
<td>Imperial</td>
<td>1433</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

(continued...)

(contains more rows and columns)
## Regional Land Use Database (Dataset Size)

<table>
<thead>
<tr>
<th>County</th>
<th>Parcel No.</th>
<th>GPZN Size</th>
<th>LU Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperial</td>
<td>85,929</td>
<td>62 MB</td>
<td>33 MB</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>2,092,552</td>
<td>1,620 MB</td>
<td>965 MB</td>
</tr>
<tr>
<td>Orange</td>
<td>661,051</td>
<td>682 MB</td>
<td>471 MB</td>
</tr>
<tr>
<td>Riverside</td>
<td>810,948</td>
<td>749 MB</td>
<td>490 MB</td>
</tr>
<tr>
<td>San Bernardino</td>
<td>804,529</td>
<td>615 MB</td>
<td>358 MB</td>
</tr>
<tr>
<td>Ventura</td>
<td>252,602</td>
<td>238 MB</td>
<td>158 MB</td>
</tr>
<tr>
<td>SCAG Region</td>
<td>4,707,611</td>
<td>3.92 GB</td>
<td>2.41 GB</td>
</tr>
</tbody>
</table>
OBJECTIVES
Objectives

- Development of an effective workflow for huge regional land use database
  - To develop an efficient regional land use data management and QC process
  - To develop a standardized and reliable workflow
METHODOLOGY
### Base Data Development Process

- **Sources for 2012 land use base data**
  - SCAG’s 2008 land use data
  - DMP LPS property data
  - DMP new construction data

- **Data processing & standardization**
  - Property data processing thru Statistical Analysis Software (SAS) by county
  - Geoprocessing thru ArcGIS applications and Python scripting
Base Data Development Process (Data Standardization)

- Data Standardization
  - 2008 land use codes → new 2012 codes
  - Attribute field properties
    - Field value type and lengths
  - Disaggregate to city-level for data entry and update process
  - Merge city-level datasets to county-level
Sample Python Scripts (Data Standardization)

- **SelectLayerByAttribute_management**

  ```python
  # Make a layer from the feature class
  arcpy.MakeFeatureLayer_management(fc, "test")

  # Select features by 'CITY' name
  whereClause = "'" + cityField + "'" + cityValue + "'"
  arcpy.SelectLayerByAttribute_management("test", "NEW_SELECTION", whereClause)

  # Write the selected features to a new feature class
  newFC = "P://=general_plan_2012/shapes/=updates/County/GPZN_County_Feb2015/=breakdown/" + cnty + "/" + cityName + ".GPZN.shp"
  arcpy.CopyFeatures_management("test", newFC)
  ```

- **Merge_management, AddField_management**

  ```python
  # Merge city-level shapefiles to county-level shapefiles
  fcList = arcpy.ListFeatureClasses("*.shp", "")
  fcMerged = "P://=general_plan_2012/shapes/=updates/County/GPZN_County_Feb2015/GeneralPlan_poly/" + cnty + ".2012.shp"
  arcpy.Merge_management(fcList, fcMerged)

  # Add a field to include parcel acreage and calculate acreage
  arcpy.AddField_management(fcMerged, "ACRES", "DOUBLE", 15, 10, ",", "ACRES", "NULLABLE")
  arcpy.CalculateField_management(fcMerged, "ACRES", "!shape.area@acres!", "PYTHON_9.3")
  ```
Data Entry and Update Process

- Manual update on city-level data by staff
  - Inputs received from jurisdictions
  - Edits on field values and parcel shape
- Merge city-level data to county-level data
- County-level update, e.g. open space
- **The Problem** – Potential human mistakes
  - Incorrect attribute field value – land use, city name, density info, etc.
  - Incorrect parcel shape & location
Data Review Process

- QC process for:
  - Data standardization
  - City-level datasets
  - County-level datasets

- Types of QC
  - Attribute field information
  - Feature comparison
  - Spatial match

- Development of Python-based workflow
Data Review Process
(City-Level Datasets)

- QC for City-Level Datasets
  - Geographic comparison
    - Parcel location vs. city boundary
  - Feature comparison
    - Geometry, feature count, attribute field count, etc.
  - Attribute field information
    - Field value accuracy – city name, residential density, null values, etc.
Sample Python Scripts (Feature Comparison)

- **FeatureCompare_Management**

```python
# Set variables for feature comparison
baseFeature = "P:/=general_plan_2012/shapes/=updates/City/" + cnty + "/" + cityName + "_GPZN.shp"
testFeature = "P:/=existing_landuse 2012/shapes/City/=updates/" + cnty + "/" + cityName + "_LU.shp"
sortField = "SCAGUID12"
xyTolerance = "1 METERS"
compareType = "GEOMETRY_ONLY"
continueCompare = "CONTINUECOMPARE"

# Compare feature and print the results
compareResult = arcpy.FeatureCompare_management(baseFeature, testFeature, sortField, compareType, "", xyTolerance, "", "", "", "", continueCompare)
print compareResult
print arcpy.GetMessages()
```

- **output**

```
===== Begin of Feature Comparison for IM =====
Brawley (Total Features: 7319)

Executing: FeatureCompare P:/=general_plan_2012/shapes/=updates/City/IM/Brawley_GPZN.shp "P:/=existing_landuse 2012/shapes/City/=updates/IM/Brawley_LU.shp" SCAGUID12 GEOMETRY_ONLY # 1 Meters 0.001 0.001 # # CONTINUE_COMPARE #
Start Time: Mon Sep 08 17:08:44 2014
Table: Tables have different number of fields (Base: 24, Test: 13).
Table: Table row counts are the same.
SpatialReference: Spatial references are the same.
FeatureClass: Geometries are the same.
Succeeded at Mon Sep 08 17:08:45 2014 (Elapsed Time: 1.00 seconds)
```
Data Review Process (County-Level Datasets)

- QC for County-Level Datasets
  - Feature comparison
    - Feature count
    - New SCAGUID12 for subdivided parcel
  - Attribute field information
    - Field value accuracy – incorrect land use codes, SCAGUID12, etc.
Sample Python Scripts (Attribute Fields)

- **SearchCursor, getValue**

```python
cursor = arcpy.SearchCursor(fc)
for row in cursor:
    citynameRow = row.getValue(cityField)
    citygpRow = row.getValue(citygpField)
    scaggpRow = row.getValue(scaggpField)
    zoneRow = row.getValue(zoneField)

    if citynameRow != cityName:
        cityNameNull = cityNameNull + 1
    if citygpRow == "":
        citygpNull = citygpNull + 1
    if scaggpRow == "":
        scaggpNull = scaggpNull + 1
    if zoneRow == "":
        zoneNull = zoneNull + 1
    parcelNum = parcelNum + 1
```

```python
cursor = arcpy.SearchCursor(fc)
for row in cursor:
    cityRow = row.getValue(cityField)
    gpRow = row.getValue(gpField)
    uidRow = row.getValue(uidField)

    if cityValue == cityRow:
        if gpRow not in correctGP:
            incorrectGP = incorrectGP + 1
            parcelNum = parcelNum + 1
            print cityValue + " " + uidRow + " ": " + gpRow
        else:
            parcelNum = parcelNum + 1
```

- **CalculateField_management**

```python
if luRow == "1113":
    luRow_before = luRow
    arcpy.CalculateField_management (fc, luField, "1150")
```
CONCLUSIONS
Benefits of Python Scripting for Data Management and QC Workflow

- Time and labor efficiency in managing and reviewing numerous and sizable datasets
- High data consistency and reliability
  - Consistent chain of data management and review process
  - Conformity with standardized data format
- Effective file and directory operation
  - `os.makedirs`, `shutil.copy`, etc.
Thank you!

Jung H. Seo
Southern California Association of Governments
seo@scag.ca.gov