Introduction to SAS DataFlux software and its use at Texas Parks and Wildlife

OFWIM 2013 Annual Conference
October 03, 2013

Jimmy White
DataFlux Methodology

1. Data discovery or data auditing
DataFlux Methodology

1. Data discovery or data auditing
   - Structure discovery
   - Data discovery
   - Relationship discovery
<table>
<thead>
<tr>
<th>Metric Name</th>
<th>Metric Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Type</td>
<td>VARCHAR</td>
</tr>
<tr>
<td>Unique Count</td>
<td>3276</td>
</tr>
<tr>
<td>Pattern Count</td>
<td>5</td>
</tr>
<tr>
<td>Minimum Value</td>
<td>1</td>
</tr>
<tr>
<td>Maximum Value</td>
<td>99999</td>
</tr>
<tr>
<td>Maximum Length</td>
<td>5</td>
</tr>
<tr>
<td>Null Count</td>
<td>0</td>
</tr>
<tr>
<td>Blank Count</td>
<td>0</td>
</tr>
<tr>
<td>Actual Type</td>
<td>Integer</td>
</tr>
<tr>
<td>Data Length</td>
<td>9 chars</td>
</tr>
</tbody>
</table>
DataFlux Methodology

1. Data discovery or data auditing
   - Structure discovery
   - Data discovery
   - Relationship discovery
DataFlux Methodology

Data Quality
DataFlux Methodology

Data Quality

• Parsing
• Standardization
• Matching
DataFlux Methodology

Standardization

East Main Street

E Main St

(512) 389-2656

5123892656
DataFlux Methodology

- Data Quality
  - Parsing
  - Standardization
  - Matching
DataFlux Methodology

Matching

75 Columbus Avenue, Austin, Texas

75 Columbus Ave, Austin, TX
DataFlux Methodology

Data integration
DataFlux Methodology

- Joining
- Merging
- Householding

3 Data integration
Enriching data
DataFlux Methodology

Enriching data

- Address Verification
- Phone Validation
- Geocoding
DataFlux Methodology

Data Monitoring
Data Monitoring

- Auditing
- Alerts
dfPower Architect
Open data set

Add Gender field

Add Match code on last name

Add Match code on first name

Add Match code on TDL and or SSN

Group rows based on match codes

Create output file
Open data set

Determine data set

Standardize address fields

Verify address values

Add Geocode fields

Create valid field names for output data set

Create output file
Open data set

Determine data set

Standardize address fields

Verify address values

Add Geocode fields

Create valid field names for output data set

Create output file
Open data set

Determine data set

Standardize address fields

Verify address values

Create valid field names for output data set

Create output file
Questions ?