INTRODUCTION TO GEOPROCESSING
CONFLATION TOOLS AND WORKFLOWS

Dan Lee
dlee@esri.com
Agenda

What is Conflation?

Geoprocessing Conflation Tools
- Demo 1 – Basic scenario

Conflation Workflows
- Demo 2 – Real world scenario

Conclusions and Future Work
What is Conflation?
When using multi-source spatial data together

Common obstacles in analysis and mapping:

- Spatial and attribute inconsistency caused by differences in data collection and modeling
- High cost to fix the problems

Overlapping datasets

Adjacent datasets
Conflation reconciles multi-source datasets and optimizes data quality and usability

Conflation is the process of:

- Identifying corresponding features (known as feature matching)
- Making spatial adjustment and attribute transfer
- Ultimately, combining matched and unmatched features into one unified dataset with the optimal accuracy, completeness, consistency, and integrity

Long-term benefits:

- No longer living with various imperfect datasets
- More confidence in reliable analysis and high quality mapping

What’s the way to get there?
Our initial focuses

Develop highly automated tools in Geoprocessing framework:

- Starting with linear features (roads, parcel lines, etc.)
- Aiming at high feature matching accuracy (not promising 100%)
- Providing information to facilitate post-processing

Build practical workflows
Challenges in feature matching - the key to conflation
Feature matching (FM) for overlapping datasets
Based on proximity, topology, pattern, and similarity analysis, as well as attributes information

1:1 and 1:m matches

<table>
<thead>
<tr>
<th>OBJECTID *</th>
<th>SRC_FID</th>
<th>TGT_FID</th>
<th>FM_GRP</th>
<th>FM_MIN</th>
<th>FM_CONF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
</tr>
</tbody>
</table>

m:1 and m:n matches

<table>
<thead>
<tr>
<th>OBJECTID *</th>
<th>SRC_FID</th>
<th>TGT_FID</th>
<th>FM_GRP</th>
<th>FM_MIN</th>
<th>FM_CONF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2:1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1:1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1:1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2:2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2:2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2:2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>-1</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>4</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
FM-based tool #1 - Detect Feature Changes (DFC)

Finding feature differences

Update features vs. base features

Output CHANGE_TYPE

- Spatial (S) change
- Attribute (A) change
- Spatial and attribute (SA) change
- No change (NC)
- New update feature (N)
- To-Delete base feature (D)
FM-based tool #2 – Transfer Attributes (TA)

From source features to target features

- Transfer fields (e.g. ROAD_NAME)
- Target features are modified
FM-based tool #3 – Generate Rubbersheet Links (GRL)

Rubbersheeting moves source locations towards target locations based on established links

Generate Rubbersheet Links (GRL)
  - From source features to target features

Followed by Rubbersheet Features (RF)
  - Adjusting input features
Edgematching (EM) for adjacent datasets

Based on proximity, topology, and continuity analysis, as well as attributes information

Generate Edgematch Links (GEL)
- From source features to adjacent features

Followed by Edgematch Features (EF)
- Connects features guided by the established links

Recommend:
Conflation: Edgematching tools and workflows
12:30 – 1:15pm, Thur.
Demo Theater – Analysis & Geoprocessing, Hall B
Demo 1: Basic scenario

Unification of simple overlapping datasets
Unification of overlapping datasets

A popular scenario and requirements:

- To unify the two datasets into one with combined spatial and attribute information
Input streets

Update features with new streets and attributes

Base features with spatial accuracy and attributes

Together
This reflects the conflation strategy. With the simple and highly similar demo data, the process produces 100% accurate result.
Results

Attributes transferred

Changes detected

Rubbersheeting links generated

New features adjusted and added to base

Final Streets

<table>
<thead>
<tr>
<th>OBJECTID</th>
<th>Shape Type</th>
<th>NAME</th>
<th>Priority</th>
<th>Shape_Length</th>
<th>UniqueID</th>
<th>CHANGE_TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Polyline</td>
<td>W OLIVE</td>
<td>M</td>
<td>293.57725</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>109</td>
<td>Polyline</td>
<td>W OLIVE</td>
<td>M</td>
<td>109.071651</td>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td>129</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>L</td>
<td>51.455285</td>
<td>&lt;Null&gt;</td>
<td>D</td>
</tr>
<tr>
<td>200</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>M</td>
<td>112.706047</td>
<td>24</td>
<td>N</td>
</tr>
<tr>
<td>201</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>M</td>
<td>69.241267</td>
<td>12</td>
<td>N</td>
</tr>
<tr>
<td>202</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>M</td>
<td>99.29832</td>
<td>13</td>
<td>N</td>
</tr>
<tr>
<td>203</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>M</td>
<td>18.383534</td>
<td>14</td>
<td>N</td>
</tr>
<tr>
<td>204</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>M</td>
<td>58.252814</td>
<td>15</td>
<td>N</td>
</tr>
<tr>
<td>205</td>
<td>Polyline</td>
<td>&lt;Null&gt;</td>
<td>M</td>
<td>37.23</td>
<td>15</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>Polyline</td>
<td>LAUREL</td>
<td>L</td>
<td>384.915164</td>
<td>27</td>
<td>NC</td>
</tr>
<tr>
<td>19</td>
<td>Polyline</td>
<td>W FERN</td>
<td>L</td>
<td>340.788763</td>
<td>40</td>
<td>NC</td>
</tr>
<tr>
<td>66</td>
<td>Polyline</td>
<td>MILLS</td>
<td>L</td>
<td>90.060797</td>
<td>87</td>
<td>NC</td>
</tr>
<tr>
<td>67</td>
<td>Polyline</td>
<td>KATHL</td>
<td>L</td>
<td>173.068809</td>
<td>88</td>
<td>NC</td>
</tr>
<tr>
<td>69</td>
<td>Polyline</td>
<td>MILLS</td>
<td>L</td>
<td>91.095418</td>
<td>90</td>
<td>NC</td>
</tr>
</tbody>
</table>
Conflation Workflows
Three components in conflation workflows

- **Preprocessing**
  - In same projection
  - Data validation
  - Selection of relevant features

- **Conflation and evaluation**
  - Conflation tools
  - Workflow tools

- **Postprocessing**
  - Queued review
  - Interactive editing
Supplemental tools and guidelines for download

http://angp.maps.arcgis.com/home/item.html?id=36961cde1b074f1f944758f6abec87cc

You can also search by “conflation” at arcgis.com to find the download.

GP Conflation_Workflow_Guidelines

This item contains conflation workflow guidelines and supplemental tools.

Geoprocessing Sample by 1812
Last Modified: July 14, 2015

(0 ratings, 0 downloads)

Sign in to rate this item.

Facebook  Twitter
Demo 2: Real world scenario

Unification of complex overlapping datasets
Data overview

Two road datasets (northeast of Meigs County, OH):
- LocalNE – 1085 features
- StateNE – 1013 features

Both datasets:
- Have common and uncommon features and attributes
- Are well preprocessed
Breakdown of Demo 1 workflow into sub-workflows

Same goal and same strategy as Demo 1

Let's get started …
Step 1a of the workflow with evaluation
DFC result and potential match errors
QA potential match errors

Total 16 CFM_GRP were flagged; 11 had match issues due to data complexity and dissimilarity; 5 were ignorable

Match issue due to data complexity

Match issue ignorable
QA DFC result – CHANGE_TYPE D and N

\((\text{CHANGE_TYPE} = 'N') \text{ OR } (\text{CHANGE_TYPE} = 'D')) \text{ AND} (\text{NEAR_DIST} > 0) \text{ AND } (\text{NEAR_DIST} < 10)\)

Inspect records with high potential for errors:

- 35 reviewed
- 11 wrong Ns or Ds flagged
Feature matching accuracy estimates

Matched groups:
- Total: 896 groups
- Correct: 885 groups
- Incorrect: 11 groups
  \[\text{Accuracy} = \frac{885}{896} = 98.77\%\]

Unmatched:
- Total: 240 (155 Ns + 85 Ds)
- Correct: 229 (151 Ns + 78 Ds)
- Incorrect: 11 (4 Ns + 7 Ds)
  \[\text{Accuracy} = \frac{229}{240} = 95.42\%\]
  \[(biased by the total count)\]

Overall feature matching accuracy (average of matched and unmatched) \[97.09\%\]

Ready to join with inputs to tag Ns and Ds ...
Extract matched features for GRL and TA processes

LocalNE: 934 non-N out of 1085
StateNE: 935 non-D out of 1013

Ready for GRL process...
Step 2 of the workflow with evaluation and QA
GRL result
Generated total 26198 regular links and 10227 identity links
GRL evaluation results – Intersecting links

54 locations of intersecting links
GRL evaluation results – linking different vertex types

(qaNotes = 'src_tgt_VxType_diff') AND (tgtVxType >=2) OR (srcVxType >=2 ) AND NEAR_DIST = -1

79 of flagged links were more important
GRL evaluation results – locations of missing links

22 of the 595 source locations of missing links were on nodes; all others are on in-line vertices. 20 ORIG_FID of frequency >5 locations were reviewed and confirmed non-critical.
QA regular links - summary

Total 241 (0.92%) of 26198 links were reviewed:
- 44 were modified
- 86 were to be removed
- 111 were ok

42 missing link locations were reviewed:
- 14 links were added
- Links at other locations were not critical

(qaNotes = 'src_tgt_VxType_diff') AND( (tgtVxType >=2) OR( srcVxType >=2 ))

Ready for rubbersheeting …
26126 regular links were selected by (REV_FLAG <> 'Delete') OR (REV_FLAG IS NULL) to participate.
Rubbersheeting result
GRL result after rubbersheeting

Many regular links became identify links
How good is the rubbersheeting result?

Three indicators showing spatial improvement

**Less spatial differences**

<table>
<thead>
<tr>
<th></th>
<th>Before RF</th>
<th>After RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular links</td>
<td>26126</td>
<td>412</td>
</tr>
<tr>
<td>Identity links</td>
<td>10227</td>
<td>15456</td>
</tr>
</tbody>
</table>

**Improved location alignment**

(Not on the same scale due to the big difference in values)
QA #3 – Check rubbersheeting result

Source (original) and target

Target features

Source adjusted with N features highlighted

Ready to do TA …
Transfer attributes from adjusted source to target
Excluding Ns from adjusted source; excluding Ds from joined target
Attribute transfer result
QA #4 – Check attribute transfer result

NEAR_DIST >=0: no-transfer features found nearby source features for potentially missed matches

32 records were reviewed:

- 18 were edited with UC2014_ID values
Select adjusted N features; append them to target

(CHANGE_TYPE = 'N') AND(( REV_FLAG <> 'wrongN') OR REV_FLAG IS NULL)
Appended N features in final result
Unification of overlapping datasets completed!

### Automated processing

<table>
<thead>
<tr>
<th>Step</th>
<th>Processing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 (a, b)</td>
<td>1 min 3 sec</td>
</tr>
<tr>
<td>Step 2</td>
<td>1 min 14 sec</td>
</tr>
<tr>
<td>Step 3</td>
<td>1 min</td>
</tr>
<tr>
<td>Steps 4, 5</td>
<td>18 sec</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3 min 35 sec</strong></td>
</tr>
</tbody>
</table>

### Interactive processing (not counting final review)

<table>
<thead>
<tr>
<th>QA #1 (CFM_GRP and DN)</th>
<th>QA #2 (links)</th>
<th>QA #3</th>
<th>QA #4 (attribute transfer)</th>
<th>QA Total</th>
<th>Time (2-3 review counts per minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Count</td>
<td>51</td>
<td>283</td>
<td>x</td>
<td>32</td>
<td>366</td>
</tr>
<tr>
<td>(locations or feature groups)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit Count</td>
<td>46</td>
<td>255</td>
<td>x</td>
<td>18</td>
<td>319</td>
</tr>
<tr>
<td>(field values)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- QA #1: CFM_GRP and DN
- QA #2: Links
- QA #3: Attribute transfer
- QA #4: Additional QA

Review Count: ~ 2-3 hrs.
Conclusions and Future Work

Thanks to:

- Department of Public Works (DPW), Los Angeles County, USA.
- Institut Cartogràfic i Geològic de Catalunya (ICGC), Barcelona, Spain.
- Ohio State Department of Transportation, USA.
- National Institute for Water and Atmospheric Research (NIWA) and Land Information New Zealand (LINZ) - Crown Copyright Reserved.
- Resource Management Service, LLC, Birmingham, AL, USA.
- All others who supported us along the way.
Conflation can be done more efficiently now

It takes a workflow:

- Use the best practice in preprocessing.
- Highly accurate results and rich information are produced automatically.
- Small amount of interactive review and editing are necessary; time is worth-spending.
Consider conflation a higher priority

Study the tools and workflows; understand the results
- Start with small test areas

Customize the workflows for your organizations
- Improve data quality and usability
- Bring new life and value to your data

Work with broader communities
- Data sharing and collaboration
- Seamless analysis and mapping

Please send us your feedbacks 😊
Our future work

New tools and enhancements
- New option in DFC tool: Compare line directions
- New Gp tools: Transform Features, Align Features
- Better feature matching

Formalization of workflows
- Common scenarios oriented
- Integrated review and editing
- Other feature types
- Contextual conflation (spatially related features)

Please send us your use cases and requirements 😊
Recent papers

- Lee D (2015) Using Conflation for Keeping Data Harmonized and Up-to-date, to be presented at the ICA-ISPRS Workshop on Generalisation and Multiple Representation, 2015, Rio de Janeiro, Brazil
Please fill out the session survey in your mobile app

- Select Introduction to Geoprocessing Conflation tools and Workflows in the Mobile App - use Search Feature to find this title
- Click “Technical Workshop Survey”
- Answer a few short questions and enter your comments

Thank you for attending! 😊

Any questions, comments ...?