Working towards Supporting Multi-scale Databases Updating

Dan Lee (dlee@esri.com)
Mark Wigley (m.wigley@esri.ch)

Auto Carto 2020, Nov. 18 -20 → WhereNext

Agenda

Introduction

- Demands and challenges

Roles of conflation in multi-scale data integration and updating

- Build master databases and establish linkages across scales
- Detect and propagate changes

Key conflation tools and user stories

- Automatic and interactive processes
- Transfer Attributes workflow

Conflation workflows

- Typical processes
- Supplemental workflow tools

Work in progress

- Change detection of point and polygon features (prototype)
- Establishment of linkages between building features at different scales (prototype)

Future work

Introduction

Demands from government organizations and national mapping agencies (NMAs)

To improve master databases by:

- Adopting new data for better spatial accuracy
- Adding attributes from other sources

To incorporate changes:

- Detecting changes
- Updating affected features

To propagate changes across scales:

- Establishing linkages for multiscale databases
- Updating affected areas

Challenges with inconsistent multi-source data

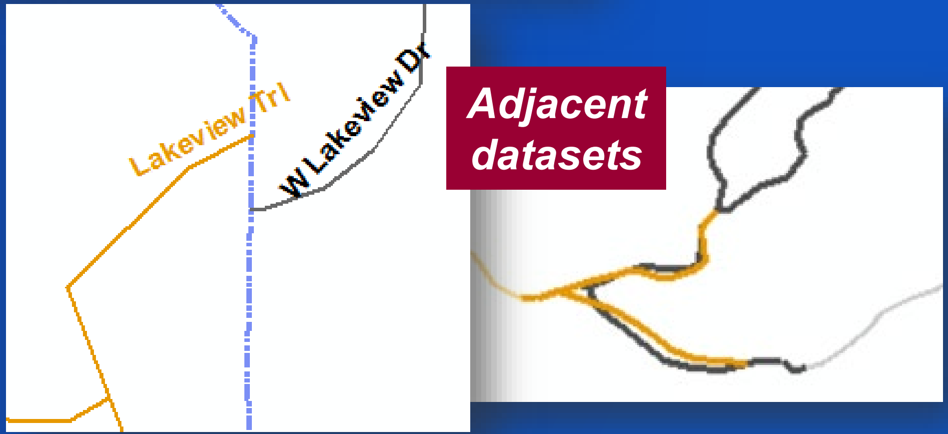
For overlapping datasets:

- Find spatial and attribute changes through feature matching
- Reconcile the differences

For adjacent datasets:

- Resolve disconnections and conflicts in bordering areas

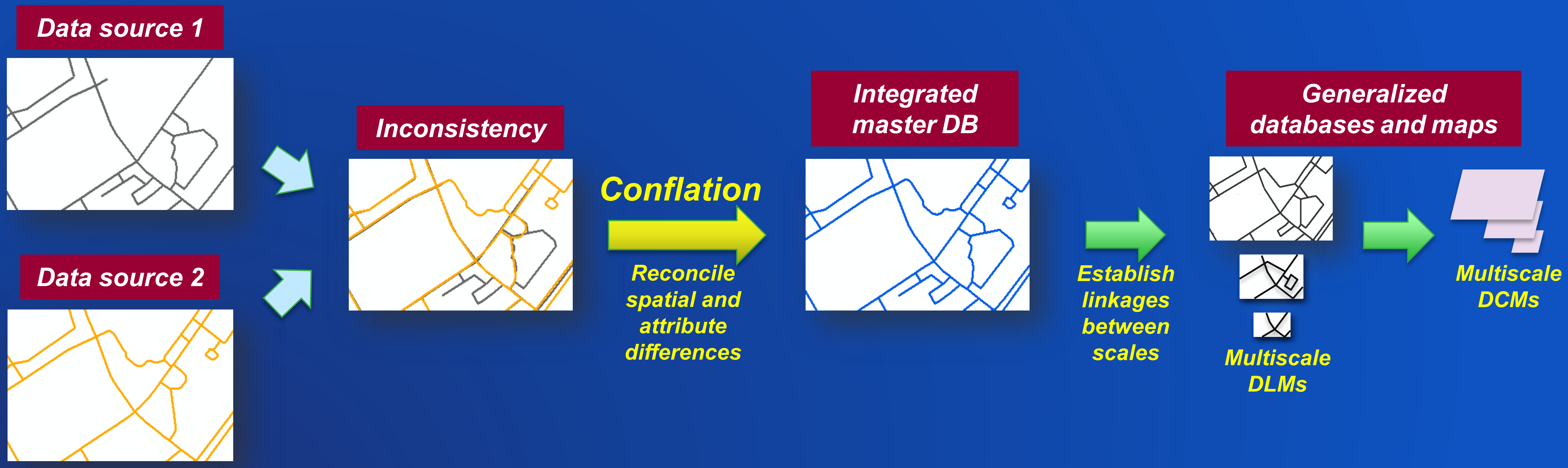
Conflation is the process to unify and enhance geographic data for reliable analysis and quality mapping



Roles of conflation in multi-scale data integration and updating

Conflation in building master databases – the basis for multiscale DLMs and DCMs

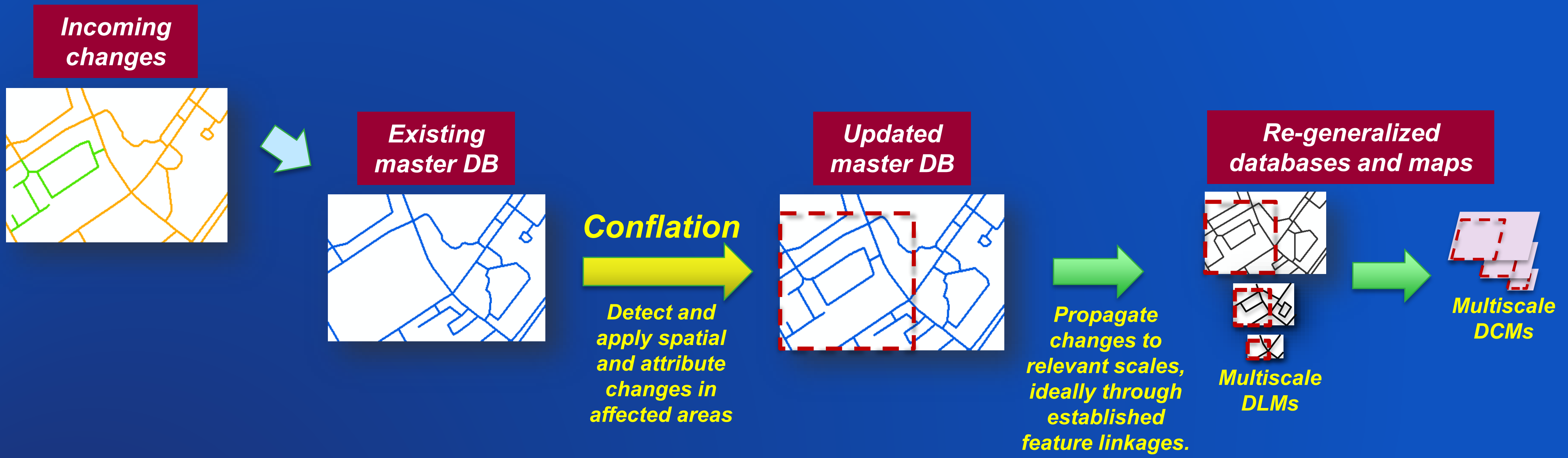
DLM = digital landscape model
DCM = digital cartographic model



Example NMA effort:

Rebuilding the cadastral map of The Netherlands: the artificial intelligence solution
https://fig.net/resources/proceedings/fig_proceedings/fig2020/papers/ts08h/TS08H_franken_florijn_et_al_10523.pdf

Conflation in maintaining and updating multiscale data for analysis and mapping



About changes and linkages (see details on establishing linkages later):

- Changes are usually incorporated in master databases; they could happen at derived scales as well.
- Linkages can be built between different levels of scales.
- In affected areas, features identified via established linkages will be re-generalized, therefore updated.
- When necessary, linkages can be updated after generalization.

Key conflation tools and user stories

Conflation tools in geoprocessing

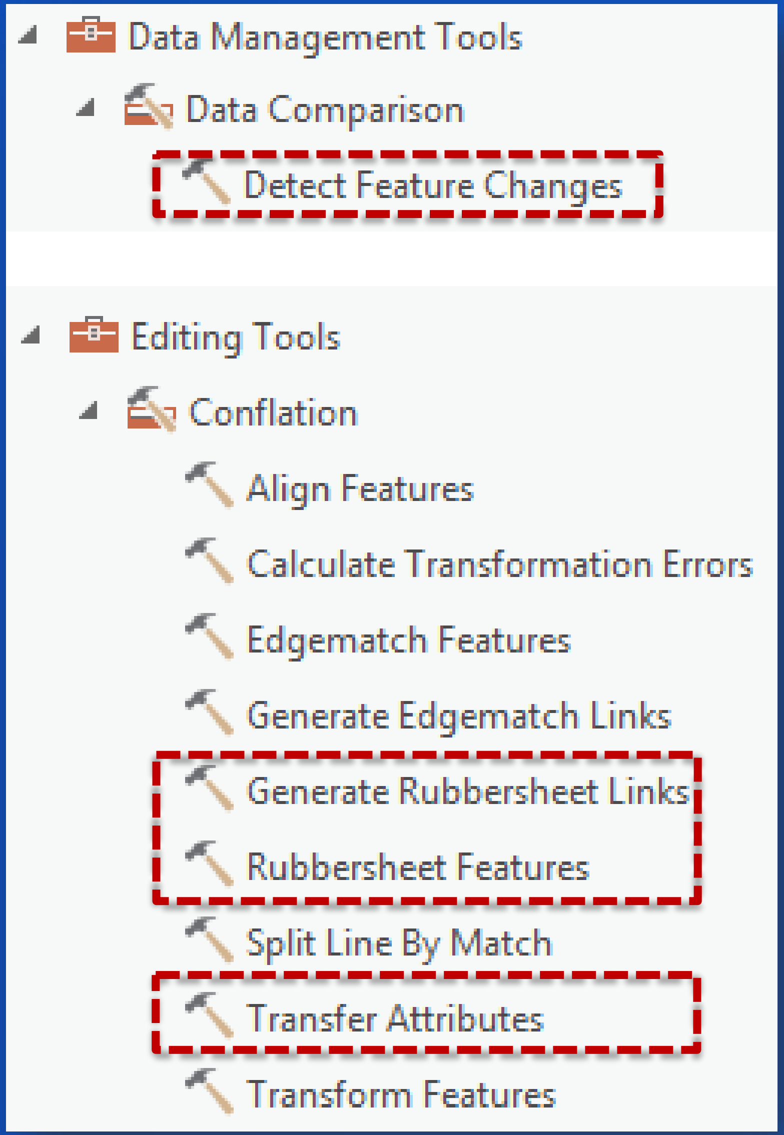
It's a challenge to develop automated conflation tools due to the complexity of the problems.

A set of tools have been released in Esri's software. These tools:

- Use in-house feature matching approaches
- Work well for common use cases
- Aim at high matching accuracy (not promising 100%)
- Provide information to facilitate post-review and editing

The key tools (highlighted) help conflate overlapping datasets (details to come next).

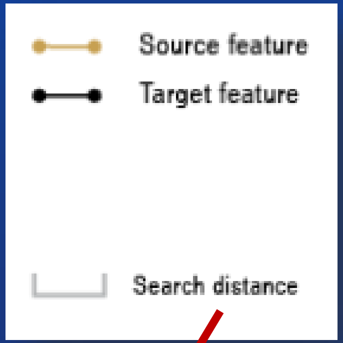
In ArcGIS Pro 2.6 and Desktop 10.8.1



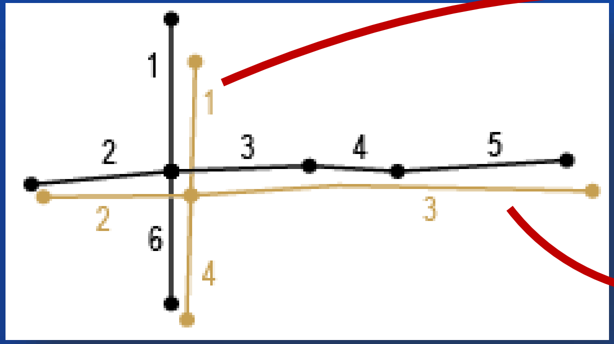
Feature matching (FM) for overlapping datasets

Based on proximity, topology, pattern, and similarity analysis, as well as attributes information

Output match table stores feature matching information

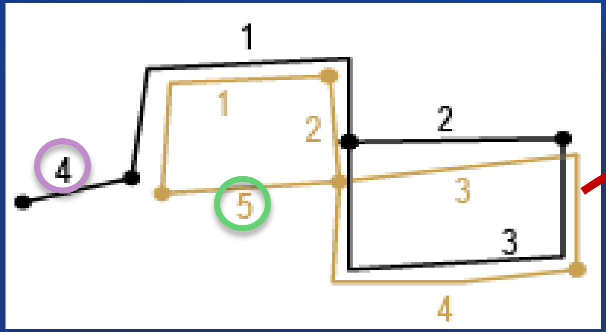


Search distance is used to find match candidates.



1:1 and 1:m matches

OBJECTID *	SRC_FID	TGT_FID	FM_GRP	FM_MN	FM_CONF
1	1	1	1	1:1	100
2	2	2	2	1:1	100
3	3	3	3	1:3	100
4	3	4	3	1:3	100
5	3	5	3	1:3	100
6	4	6	4	1:1	100



m:1 and m:n matches

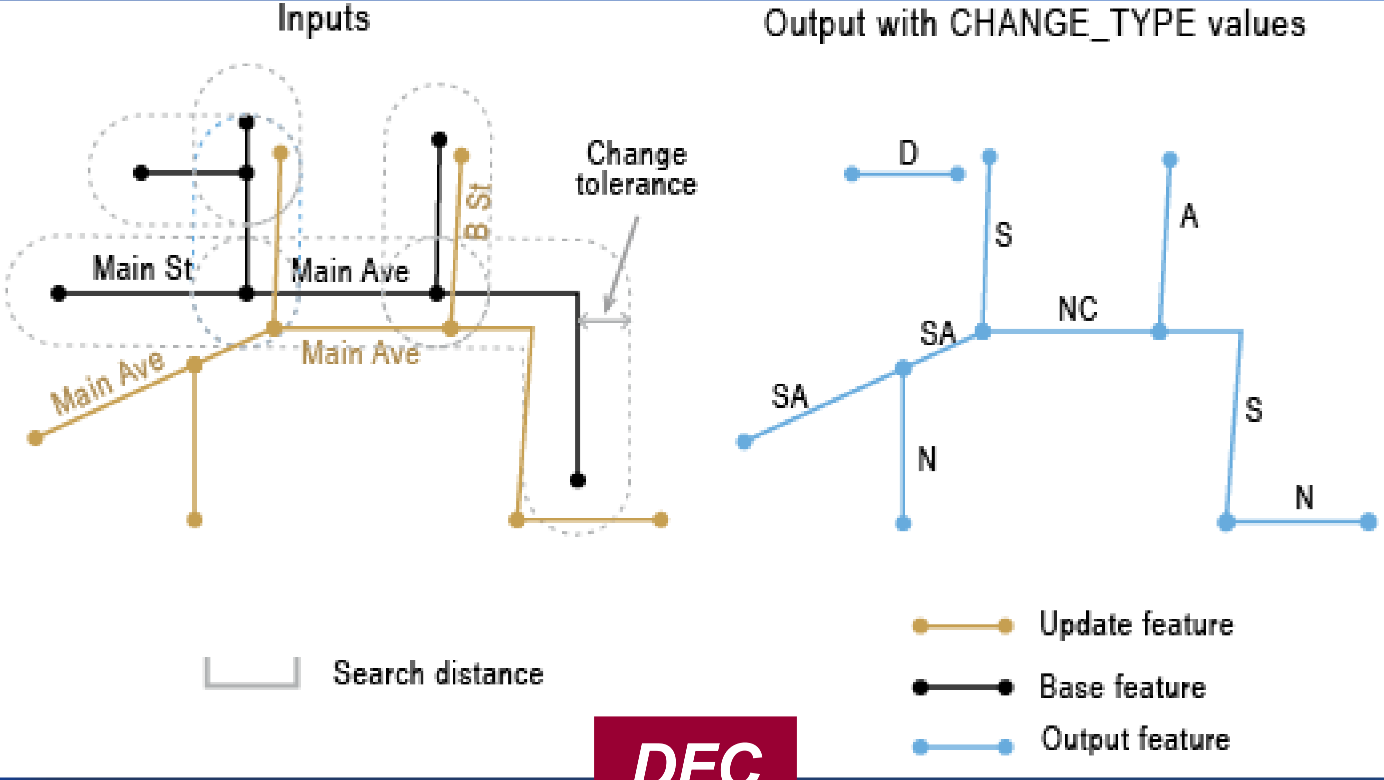
OBJECTID *	SRC_FID	TGT_FID	FM_GRP	FM_MN	FM_CONF
1	1	1	1	2:1	100
2	2	1	1	2:1	100
3	3	2	2	2:2	100
4	3	3	2	2:2	100
5	4	2	2	2:2	100
6	4	3	2	2:2	100
7	5	-1	-1	N/A	0
8	-1	4	-1	N/A	0

FM-based tool #1 - *Detect Feature Changes* (DFC)

Finds and reports feature differences

Output CHANGE_TYPE

- Spatial change (S)
- Attribute change (A)
- Spatial & attribute change (SA)
- Spatial and line direction change (S_LD)
- Spatial, attribute, and line direction change (SA_LD)
- No change (NC)
- Unmatched update feature (N)
- Unmatched base feature (D)



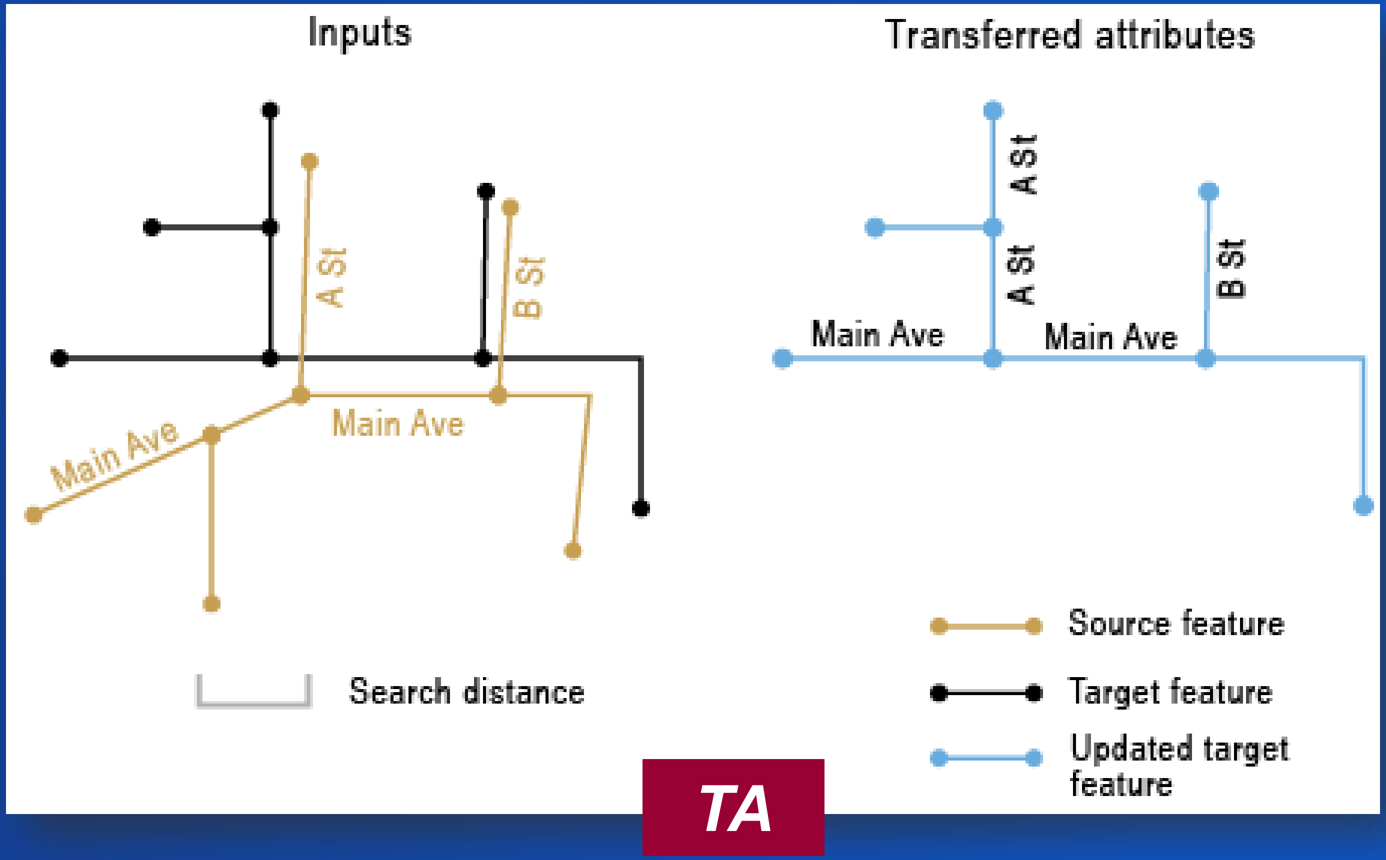
DFC

FM-based tool #2 – *Transfer Attributes* (TA)

Transfers attributes from source to target features

Target features are modified

- Transfer fields (e.g. ROAD_NAME, UniqueID) are added
- Optional transfer rules for m:n matches can be specified (see tool reference)



FM-based tool #3 – *Generate Rubbersheet Links* (GRL)

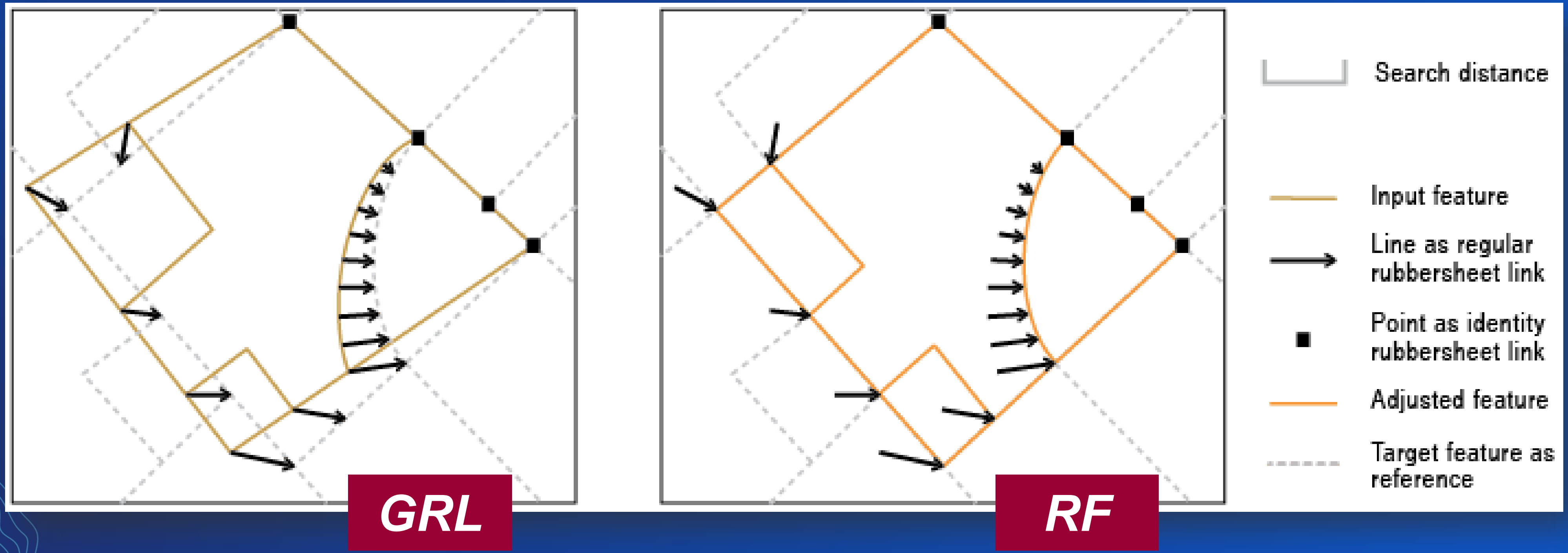
Derives links from source to corresponding target locations for rubbersheeting adjustment

Generate Rubbersheet Links (GRL)

- Regular links (lines) and identity links (points)

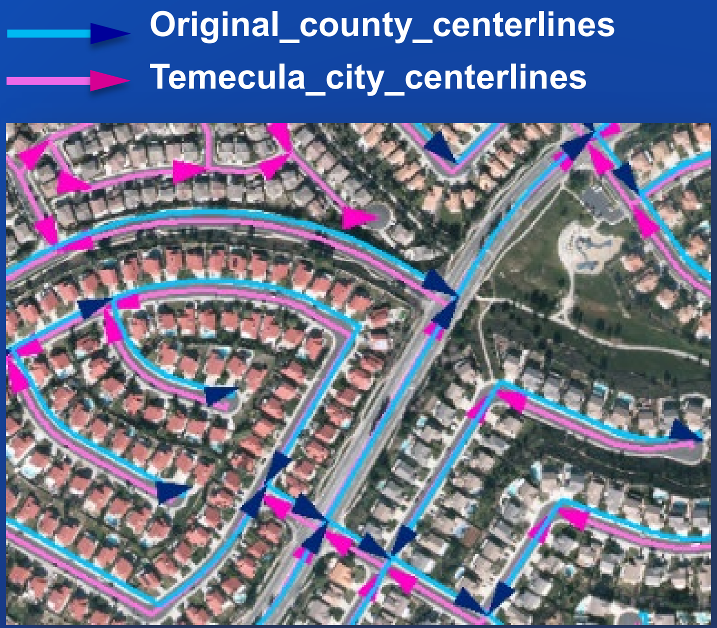
Followed by Rubbersheet Features (RF)

- Adjusting input features to target locations



User story 1: Enhancing county roads with spatially more accurate city roads

County centerline attributes and direction must be retained.



- Use *Detect Feature Changes* to find matching features and line direction differences
  - For 1:1 matches, flip city centerlines of opposite direction (Flip Line)
  - For m:n matches, merge/split city or county centerlines to get 1:1 matching segments, recalculate address ranges for county roads as needed, and flip city centerlines of opposite direction (tools + scripts)
- Transfer city centerline geometry to county centerlines (script)

~ 98%+ accuracy

Data/information source: RCTLMA (Riverside County Transportation and Land Management) CA, USA

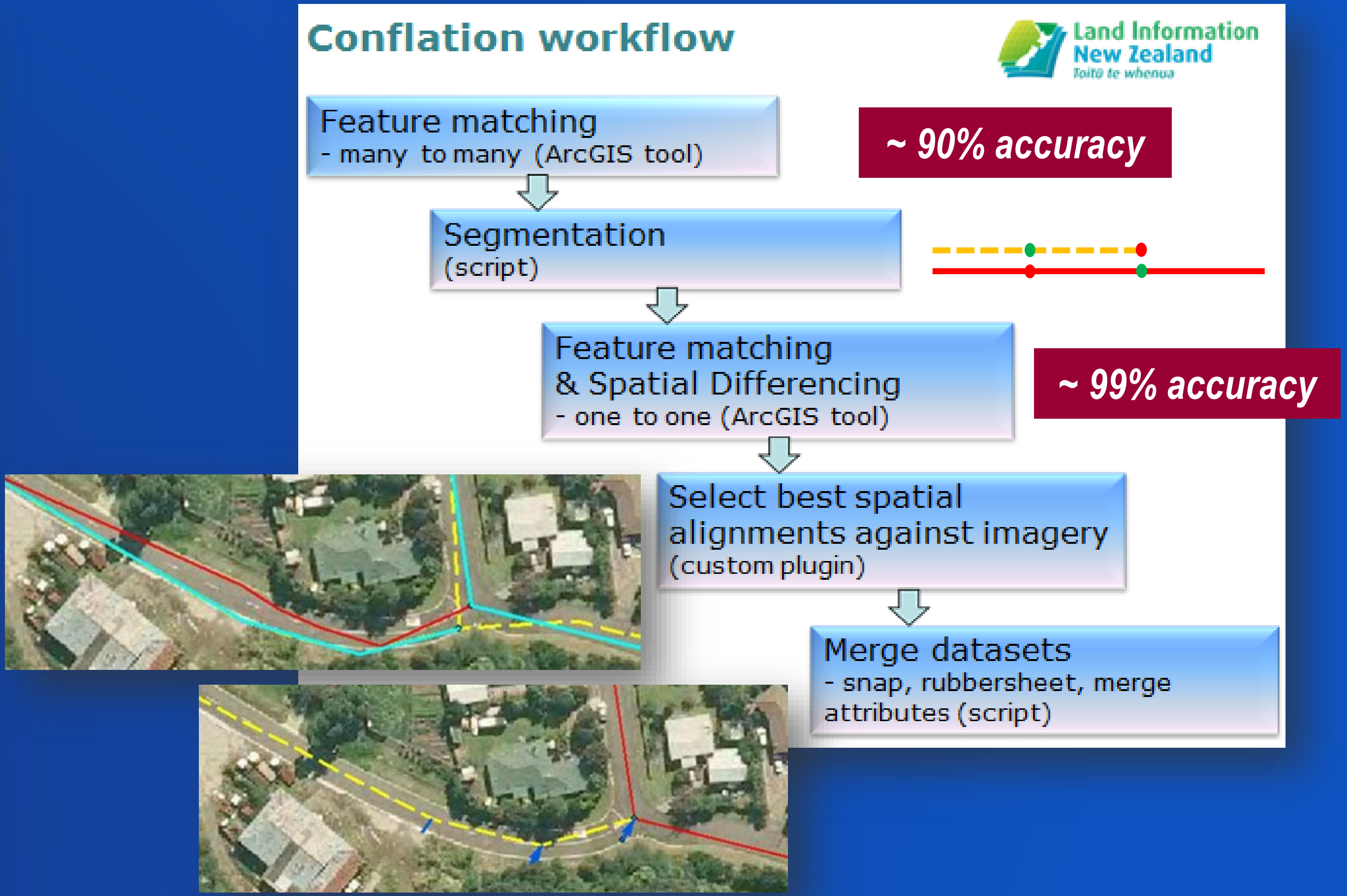
Acknowledgement: Thanks to Richard Fairhurst, for providing the information and screenshots.

User story 2: Combining electoral roads and topographic roads

There is no “most accurate” dataset.

Electoral roads

Topographic roads

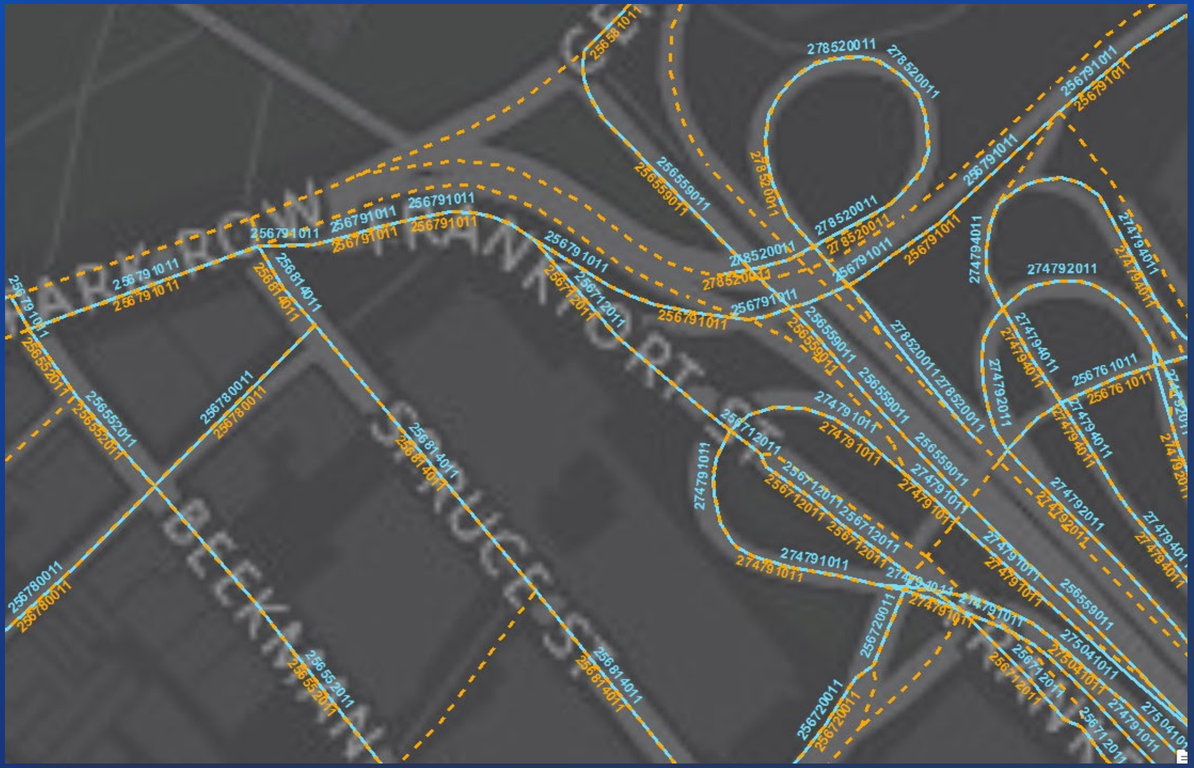


Information source: Land Information New Zealand (LINZ)

Acknowledgement: Thanks to Douglas Kwan, LINZ, for providing the information.

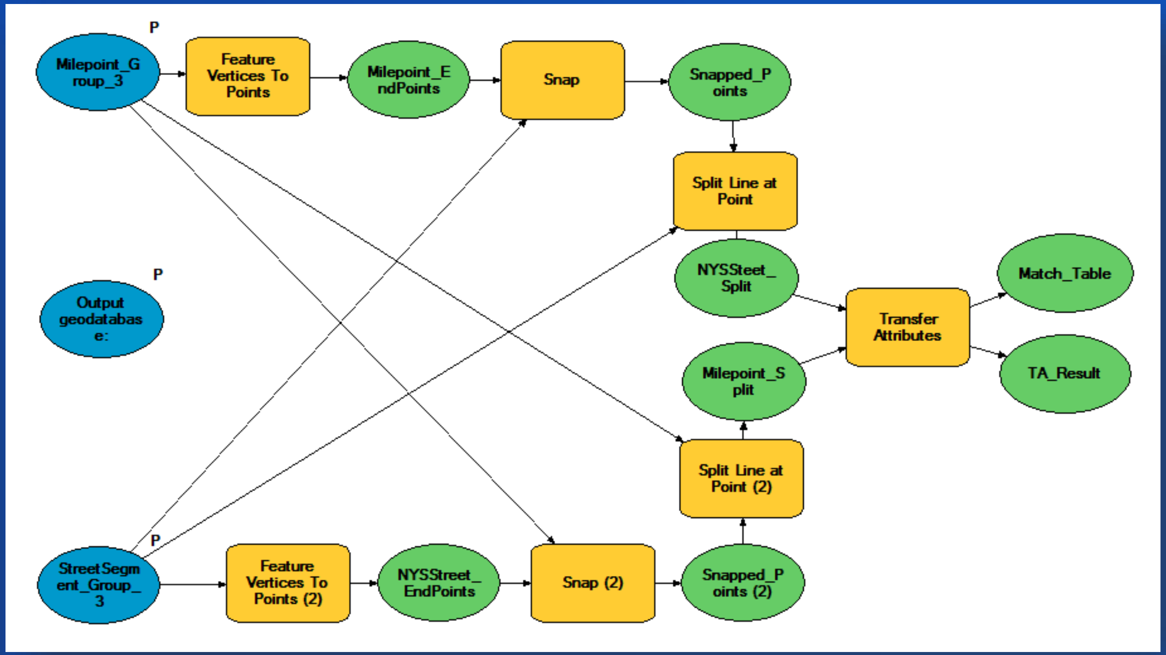
User story 3: Transferring attributes from State routes to Street segments

Segmentation for the datasets was different



State routes

Street Segments



- State Routes and Street segments were split by end points to provide a more similar segmentation between the two datasets.

~99.5% matching rate

Data/information source: NYSDOT, USA

Acknowledgement: Thanks to Kevin Hunt, for giving us the opportunity to work with him and share his data.

Conflation workflows

Typical processes

Pre-processing → **Conflation and evaluation** → **Review and edit**

Pre-processing

- Consistent projection
- Data validation
- Selection of relevant features

Conflation and evaluation

- Conflation tools
- Workflow tools

Review and edit

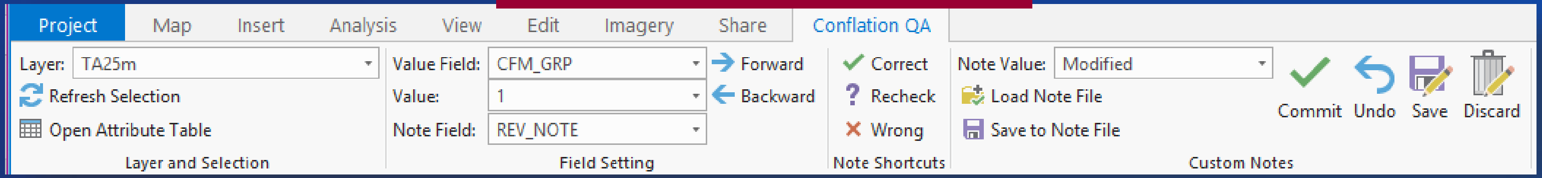
- Interactive review and edit

Supplemental Workflow Tools

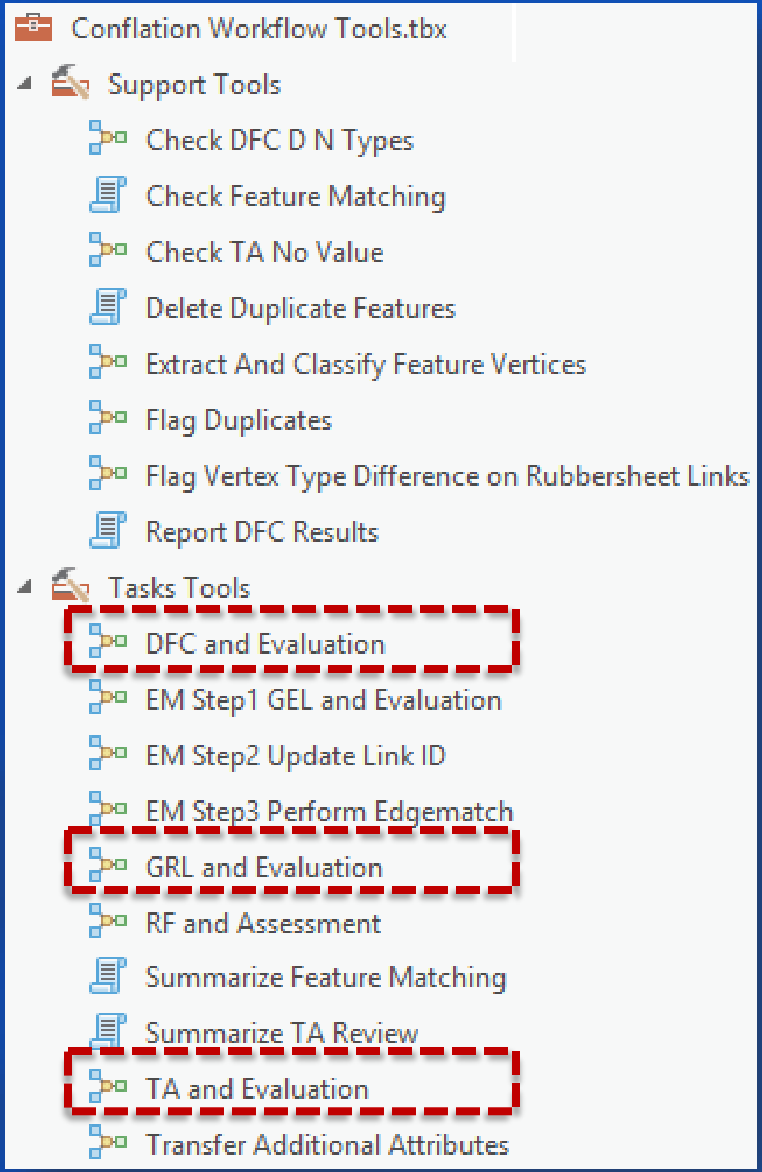
Focusing on popular tasks:

- **Automatic processes are done by task-specific tools**
- **Each tool involves a key System tool + Evaluation**
- **Potential issues are flagged in the outputs**
- **Review is done interactively using Conflation QA add-in tools**

SDK add-in for ArcGIS Pro

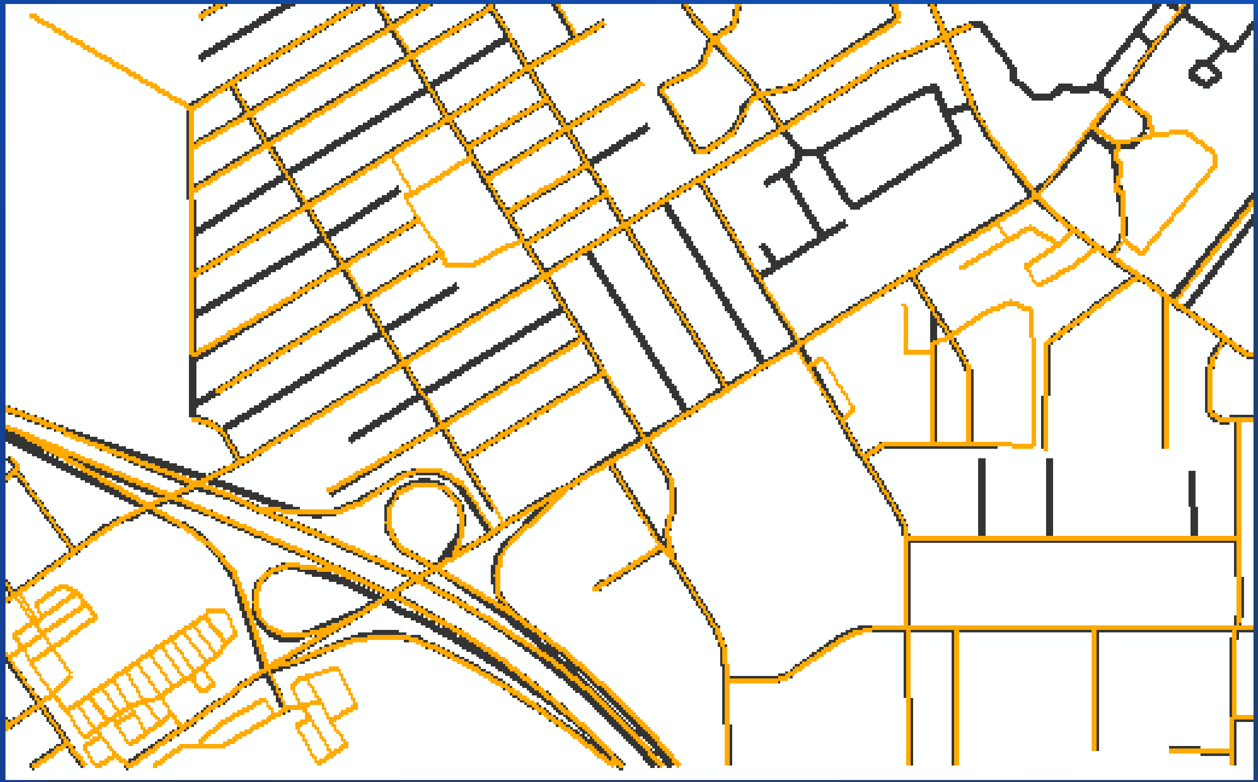


Python add-in for ArcGIS Desktop



Transfer Attributes (TA) workflow

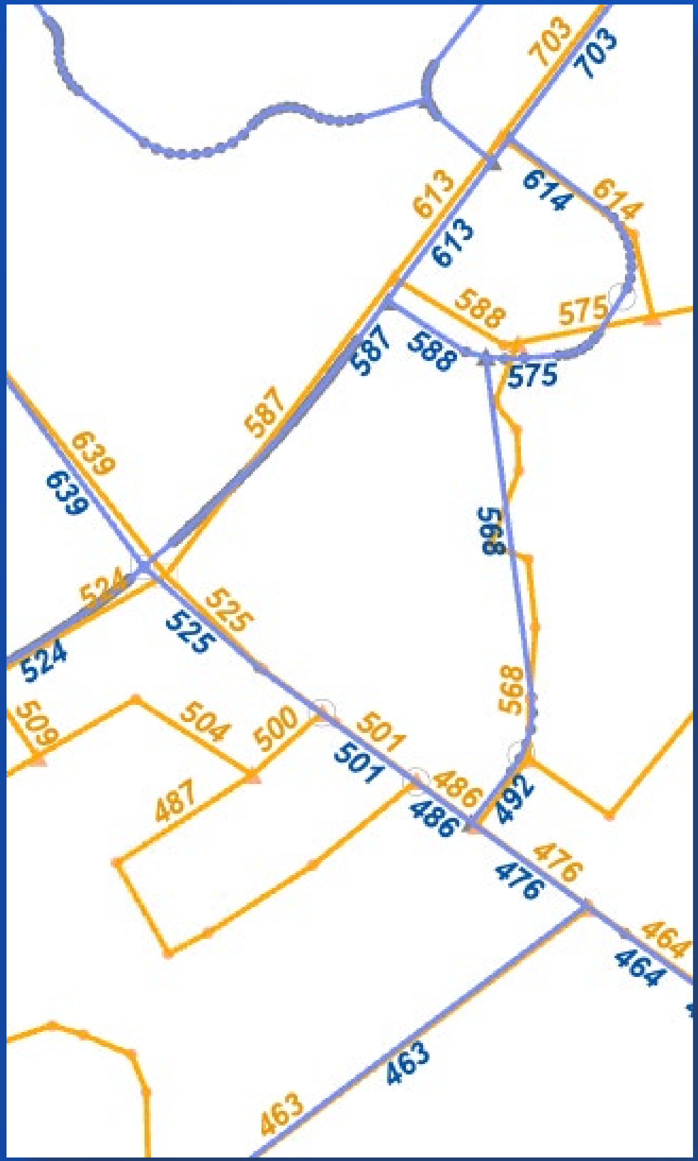
- **For transferring legacy attributes to more accurate geometry**
- **For transferring unique IDs from source to target features as linkages**



Overlapping road datasets (Esri demo data)

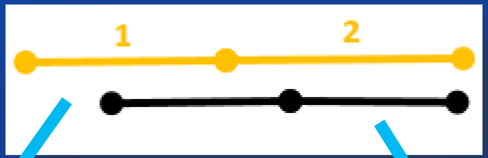
What are linkages?

- Common field values for corresponding source and target features or features at different scales
- Key to transferring additional attributes
- Basis for propagating changes
- Can be established using Transfer Attributes workflow



Establishing linkages for matching features

Two inputs



Source

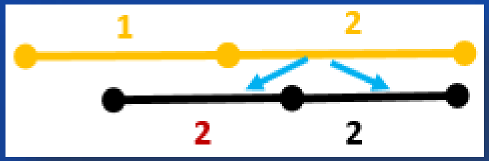
Existing fields…	TA_uID
	1
	2

Added field with calculated unique values

Target

Existing fields…

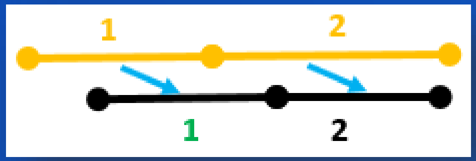
Automated TA result: showing a 1:2 match



Target received TA_uID

Existing fields…	TA_uID
	2
	2

Interactive review and edit favoring two 1:1 matches



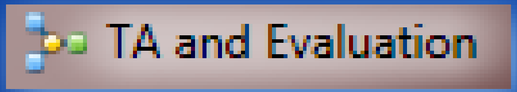
Target after interactive review and edit

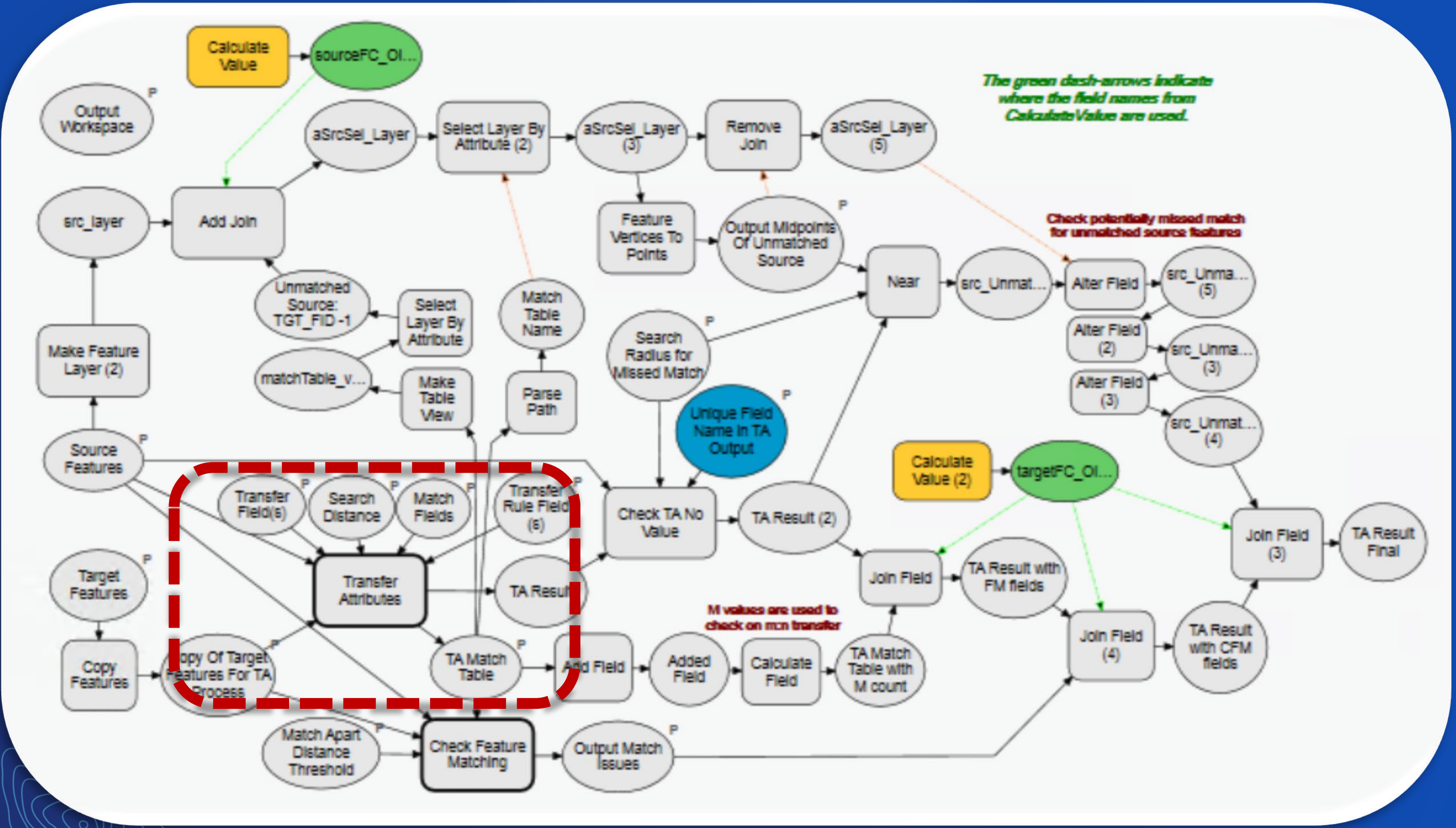
Existing fields…	TA_uID	rev_ TA_uID	final_ TA_uID
	2	1	1
	2		2

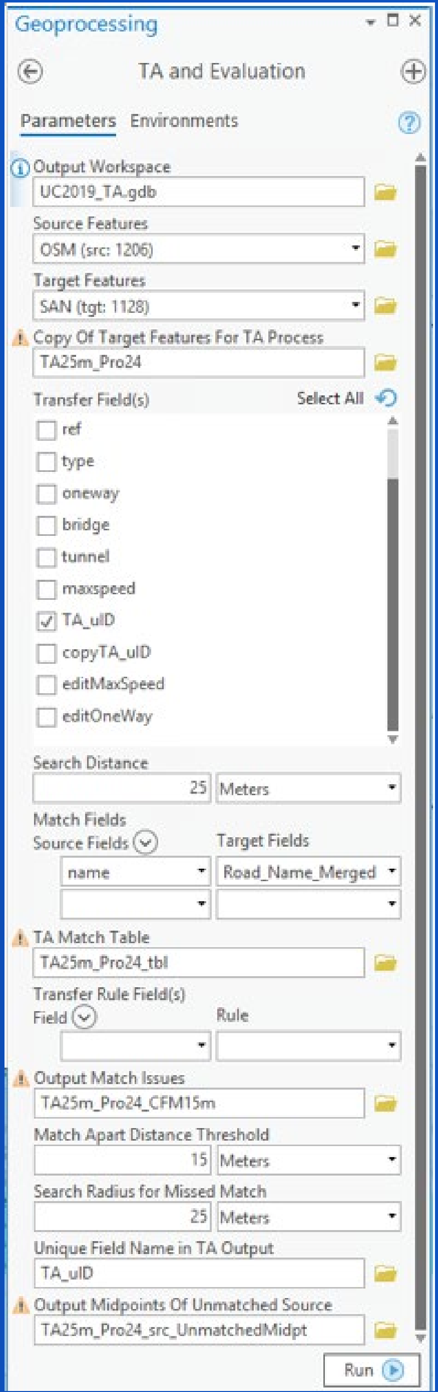
Linkage

Automated *TA and Evaluation*

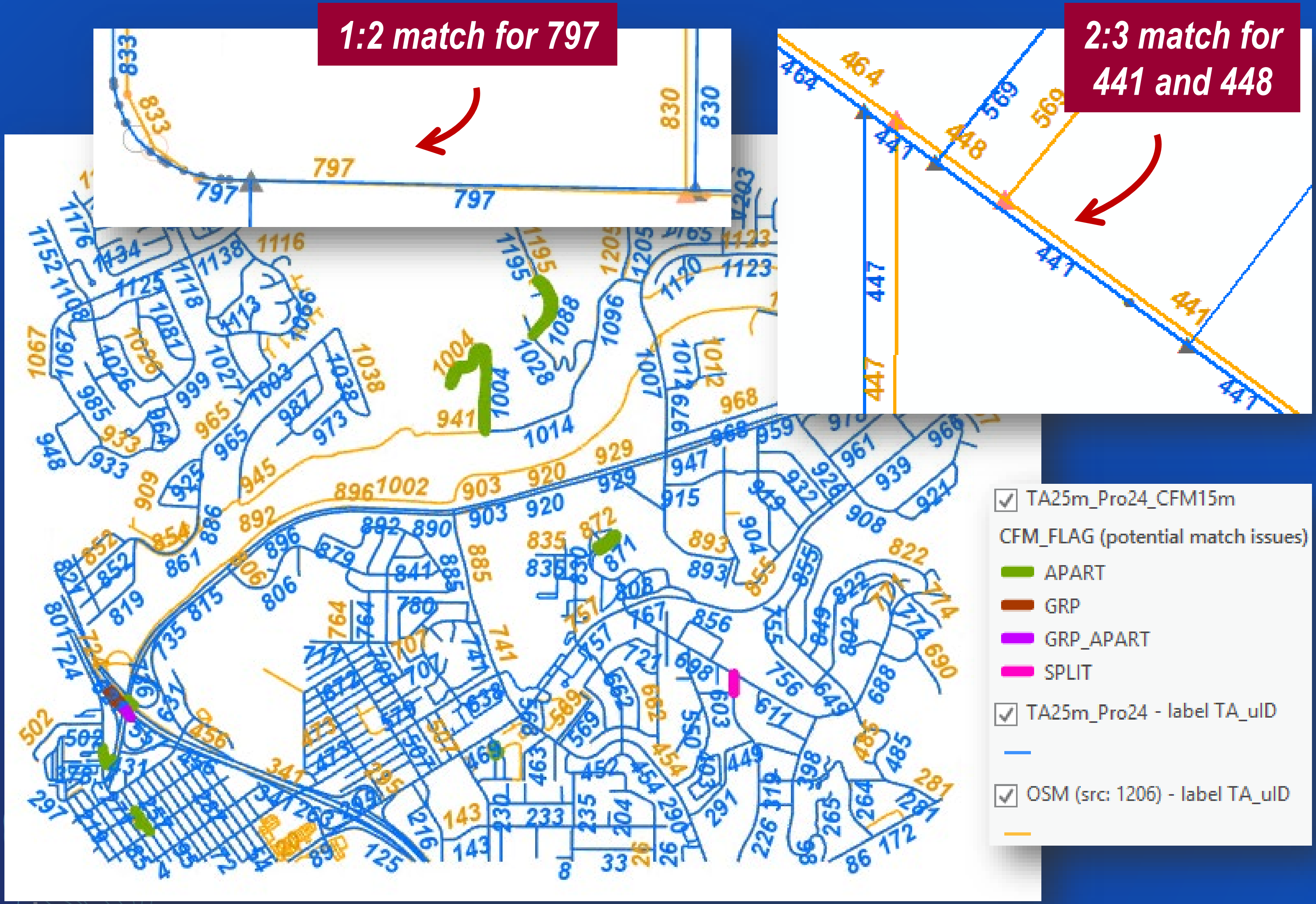
TA_uID (unique ID field) is transferred; potential issues are flagged for review







TA result and flags on potential issues



To review

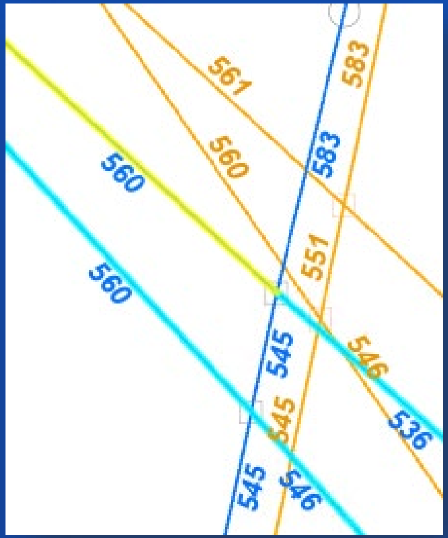
- **Potential wrong transfers**
- **Potential missed transfers**
- ***The m:n transfers flagged in field srcM_inMN (optional); see notes later***

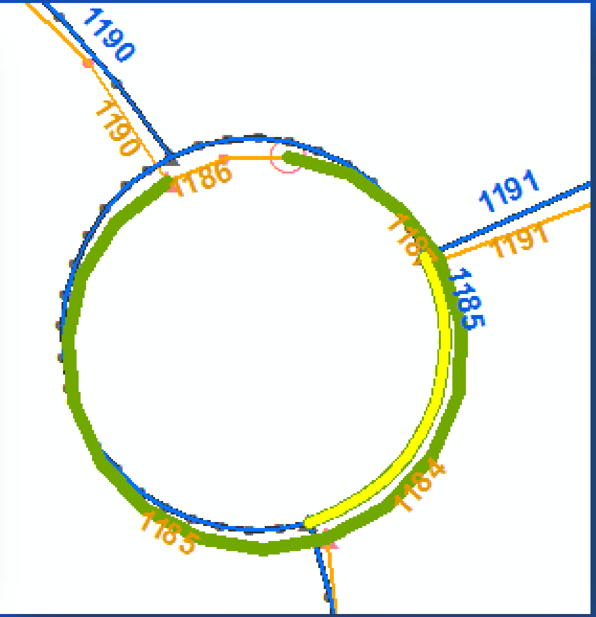
Review transfers with potential match issues

CFM_GRP >=0 (Check Feature Matching flags)

19 records were reviewed:

- **5 TA_uID values (unique IDs transferred from source features) were corrected**





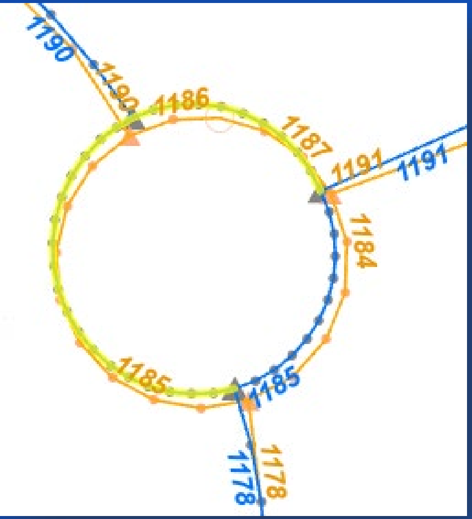
TA25m_Pro24

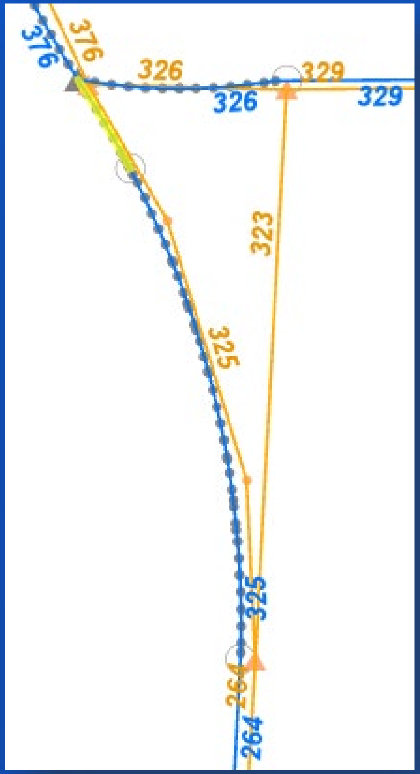
OBJECTID	Shape_Length	TA_uID	NEAR_FID	NEAR_DIST	FM_GRP	FM_MN	srcM_inMN	CFM_GRP	CFM_FLAG	CFM_DIST	srcORIGFID	srcNearFID	srcNearDist	REV_NOT	rev_TA_uID	final_TA_uID
147	24.825103	536	<Null>	<Null>	462	2:1	2	3	GRP_APART	21.839679	552	147	6.838031	Wrong	552	<Null>
402	90.700884	627	<Null>	<Null>	460	1:1	1	1	GRP	-1	<Null>	<Null>	<Null>	Wrong	561	<Null>
403	25.757369	560	<Null>	<Null>	461	1:1	1	2	GRP	-1	561	403	7.169561	Wrong	561	<Null>
515	90.268456	627	<Null>	<Null>	455	1:1	1	1	GRP	-1	<Null>	<Null>	<Null>	Wrong	560	<Null>
664	32.981175	1185	<Null>	<Null>	884	3:1	3	14	APART	17.71337	<Null>	<Null>	<Null>	Wrong	1184	<Null>
148	21.253369	536	<Null>	<Null>	458	1:2	1	3	GRP	-1	<Null>	<Null>	<Null>	Correct	<Null>	<Null>
149	10.856186	536	<Null>	<Null>	458	1:2	1	3	GRP	-1	<Null>	<Null>	<Null>	Correct	<Null>	<Null>
275	23.797976	585	<Null>	<Null>	234	2:1	2	5	APART	18.11008	<Null>	<Null>	<Null>	Correct	<Null>	<Null>

19 of 1,128 selected   Filters: 78 %

Review potential missed transfers

TA_uID IS NULL AND (NEAR_FID >0 OR srcNearFID >0)





48 records were reviewed:

- 7 transferred TA_uID values were incorrect

TA25m_Pro24

OBJECTID	Shape_Length	TA_uID	NEAR_FID	NEAR_DIST	FM_GRP	FM_MN	srcM_inMN	CFM_GRP	CFM_FLAG	CFM_DIST	srcORIGFID	srcNearFID	srcNearDist	REV_NOT	rev_TA_uID	final_TA_uID
142	42.646716	<Null>	599	16.108438	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	599	<Null>
401	40.03363	<Null>	731	9.641416	-1	N/A	-1	<Null>	<Null>	<Null>	731	401	9.630519	Wrong	731	<Null>
421	57.363602	<Null>	1185	0.214332	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	1185	<Null>
625	11.066227	<Null>	325	0.525794	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	325	<Null>
816	30.400859	<Null>	1187	1.463928	-1	N/A	-1	<Null>	<Null>	<Null>	1186	816	1.876831	Wrong	1187	<Null>
956	43.071085	<Null>	726	5.306041	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	726	<Null>
1018	0.069553	<Null>	850	0.02129	-1	N/A	-1	<Null>	<Null>	<Null>	850	1018	3.947928	Wrong	861	<Null>
1	10.550205	<Null>	415	23.914806	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Correct	<Null>	<Null>
58	120.644704	<Null>	-1	-1	-1	N/A	-1	<Null>	<Null>	<Null>	559	58	23.366778	Correct	<Null>	<Null>

48 of 1,128 selected   Filters: 78 %

Review of m:n transfers (optional)

- For an m:n match where m (number of source features) is greater than 1, by default the transfer will come from the longest source feature
- The SrcM_inMN field stores the m values. If necessary, these records can be reviewed.
- Parameter Transfer Rules can be set to control attribute transfers for m:n matches. See Transfer Attributes tool reference for details.
- Details on reviewing m to n transfers are omitted in this presentation, assuming the default is acceptable.

Attribute transfer accuracy estimates

- **Source feature count: 1206**
- **Target feature count: 1128**
- **Incorrect transfer count: 12**
- **Correct transfer count: 1116**

  Accuracy = 1116 / 1128 => 98.9%

- **Remaining no transfer features (TA_uID IS NULL AND REV_FLAG IS NULL): 80**

  All correct

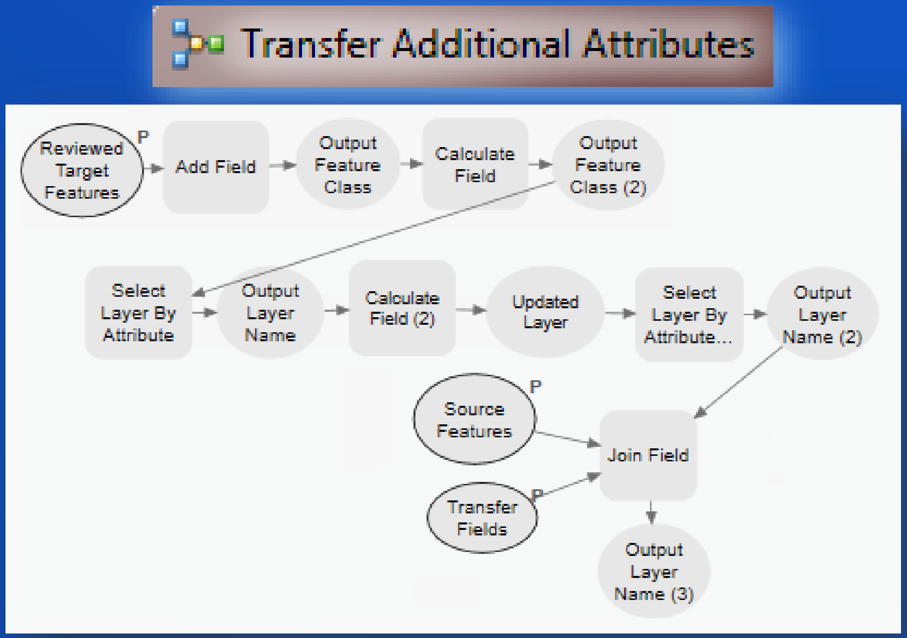
Post transfer if necessary

Merge TA result with edits

- Add a field final_TA_uID and Calculate final_TA_uID = TA_uID
- Select REV_NOTE = ‘Wrong’ and calculate final_TA_uID = rev_TA_uID

Transfer additional attributes via Join Field

- The common fields are: source TA_uID and target final_TA_uID



OBJECTID	TA_uID	NEAR_FID	NEAR_DIST	FM_GRP	FM_MN	srcM_inMN	CFM_GRP	CFM_FLAG	CFM_DIST	srcORIGFID	srcNearFID	srcNearDist	REV_NOTE	rev_TA_uID	final_TA_uID	name
816	<Null>	1187	1.463928	-1	N/A	-1	<Null>	<Null>	<Null>	1186	816	1.876831	Wrong	1187	1187	
421	<Null>	1185	0.214332	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	1185	1185	
664	1185	<Null>	<Null>	884	3:1	3	14	APART	17.71337	<Null>	<Null>	<Null>	Wrong	1184	1184	
1018	<Null>	850	0.02129	-1	N/A	-1	<Null>	<Null>	<Null>	850	1018	3.947928	Wrong	861	861	Capistrano Drive
401	<Null>	731	9.641416	-1	N/A	-1	<Null>	<Null>	<Null>	731	401	9.630519	Wrong	731	731	San Diego Freeway
956	<Null>	726	5.306041	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	726	726	San Diego Freeway
142	<Null>	599	16.108438	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	599	599	San Diego Freeway
402	627	<Null>	<Null>	460	1:1	1	1	GRP	-1	<Null>	<Null>	<Null>	Wrong	561	561	San Diego Freeway
403	560	<Null>	<Null>	461	1:1	1	2	GRP	-1	561	403	7.169561	Wrong	561	561	San Diego Freeway
515	627	<Null>	<Null>	455	1:1	1	1	GRP	-1	<Null>	<Null>	<Null>	Wrong	560	560	
147	536	<Null>	<Null>	462	2:1	2	3	GRP_APART	21.839679	552	147	6.838031	Wrong	552	552	San Diego Freeway
625	<Null>	325	0.525794	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	325	325	Mac Donald Street
1	<Null>	415	23.914806	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Correct	<Null>	<Null>	<Null>
2	543	<Null>	<Null>	8	1:1	1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	543	San Diego Freeway
3	534	<Null>	<Null>	7	1:2	1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	534	San Diego Freeway
4	534	<Null>	<Null>	7	1:2	1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	534	San Diego Freeway

12 of 1,128 selected   Filters: 76 %

Conflation is a necessity and goes a long way

Increased productivity

- Tedious and costly manual process is replaced by automatic + interactive workflow
- Data harmonization and updating can be done more efficiently

Enhanced data quality and usability

- Conflation brings more value to your data and products; time worth spending

Improved data sharing and collaboration

- Up-to-date and consistent data lead to effective communication and ensure reliable analysis and quality mapping

Work in progress

Prototype of change detection for points and polygons

Test data provided by swisstopo

- Point buildings from the years of 2015 and 2018,
  - Most features have common unique UUIDs
- Polygon buildings from the years of 2015 and 2018,
  - Most features have common unique UUIDs

Preliminary tools

- Detect Feature Changes for Points
- Detect Feature Changes for Polygons

Acknowledgements:

Cartography Department of swisstopo, Switzerland, provided to us the test data – point and polygon buildings.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Office of Topography swisstopo
Cartography

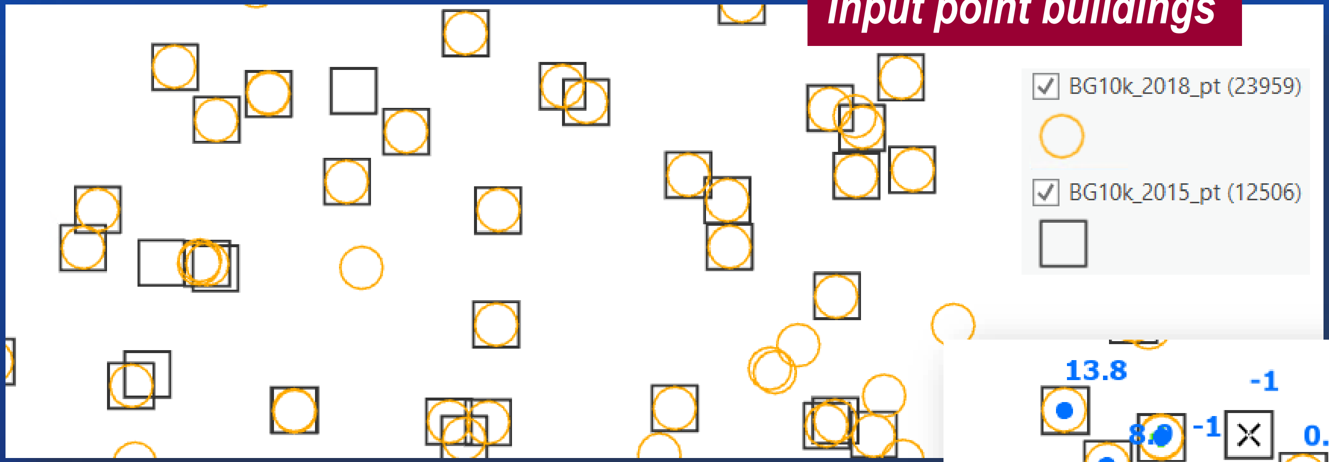
Yuxiao Zhao, graduate student at Ohio State University, USA, worked with us as a summer intern in 2019 and implemented python scripts for change detection on point and polygon buildings.

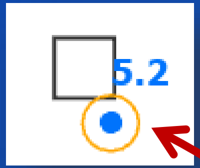
Detect feature changes for points (buildings)

In both datasets:

- Points with matching UUID values may or may not line up spatially
- Points without matching UUID values may line up spatially
- The remaining points don't match either spatially or by UUID.

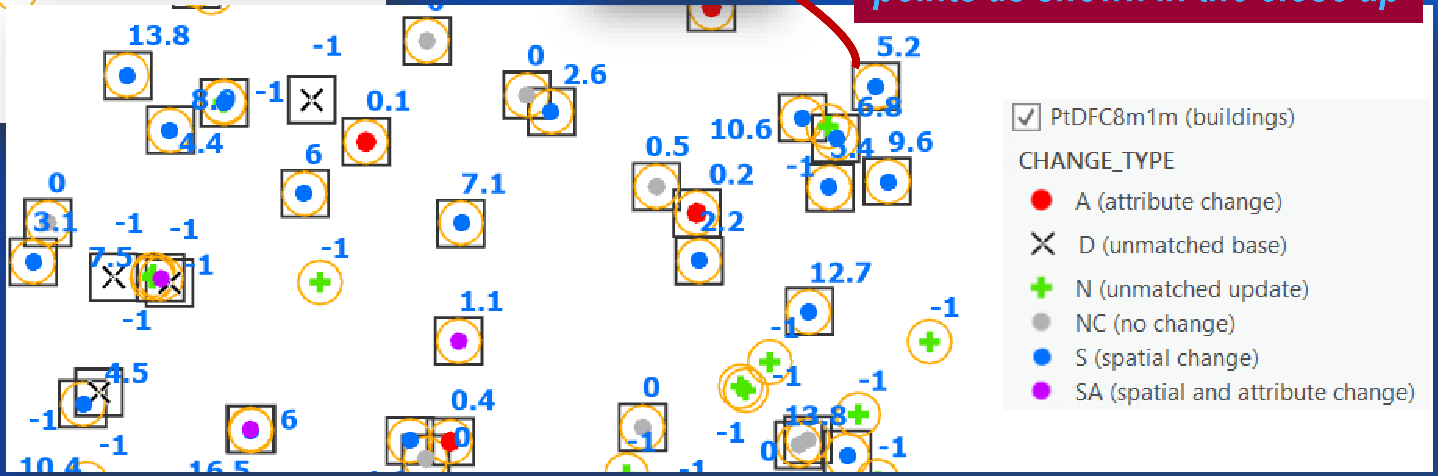
Input point buildings





Output Change Types

The blue labels show the distances between matched points as shown in the close up

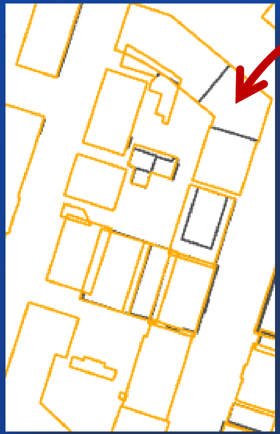


Detect feature changes for polygons (buildings)

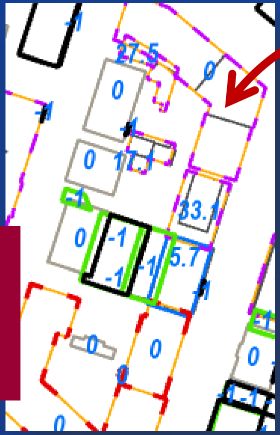
In both datasets:

- Polygons with matching UUID values may or may not line up spatially
- Polygons without matching UUID values may line up spatially
- The remaining polygons don't match either spatially or by UUID.
- Connected polygons may or may not line up.

Input polygon buildings (years of 2018 vs. 2015)



Blue labels are percentage values of areal difference

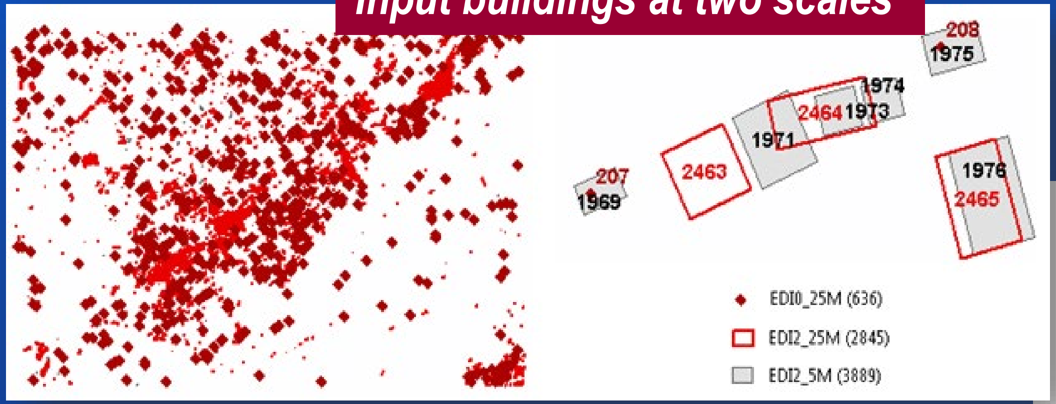


MFC stands for match field change

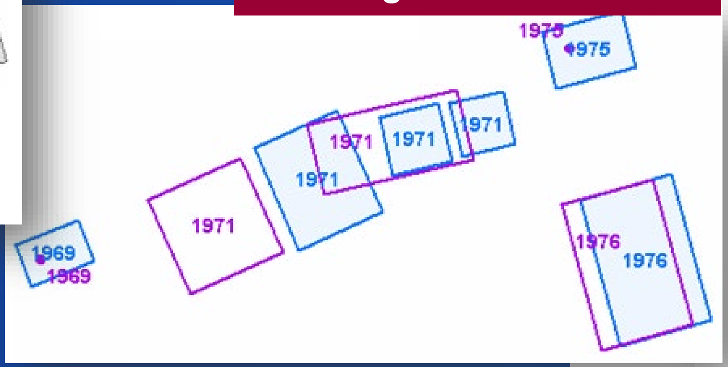
Output Change Types

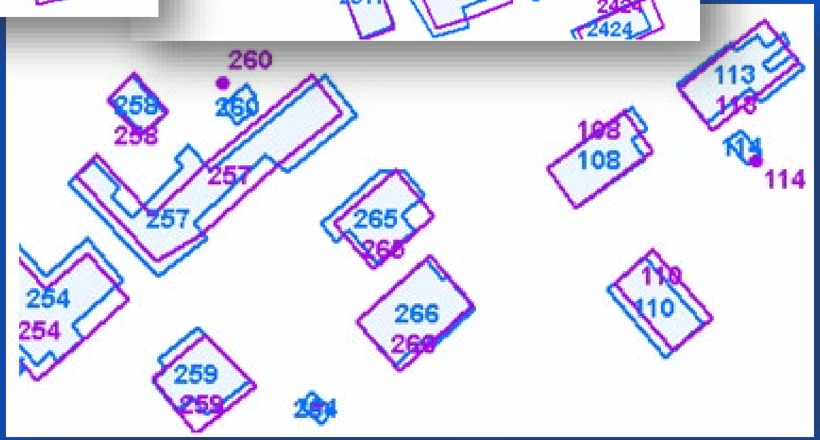
Prototype of linking features: ICGC 1:5k & 1:25k buildings

Input buildings at two scales



Close ups of resulting unique IDs linking buildings of two scales





Based on generalization principles – buildings can be:

- Simplified, typified, aggregated, displaced, or collapsed to points.
- Spatial constraints, such as minimum spacing, size, etc. are considered in searching for corresponding features.

Unique IDs as linkages are assigned to corresponding buildings

Reference:

Baella, B., Lee, D., Lleopart, A., Pla, M. (2014). ICGC MRDB for topographic data: first steps in the implementation, The 17th ICA Generalization Workshop, 2014, Vienna, Austria

Future work

Forward looking goals

New tools and enhancements

- Improve feature matching quality and efficiency
- Develop new tools to support more use cases
- Enhance outputs with more information to facilitate review process

Integrated conflation system

- Manage automated and interactive conflation tasks
- Incorporate other data sources (imagery, lidar, GPS)

Formalization of workflows

- Expand work in progress to support incremental multi-scale data updating
- Research on contextual conflation (considering spatially related features)

Thank you for your interest!

We look forward to your feedback and discussions.

esri | THE SCIENCE OF WHERE