Conflation Tools and Workflows: An Introduction

DAN LEE and SILVIA CASAS

ESRI USER CONFERENCE

Agenda

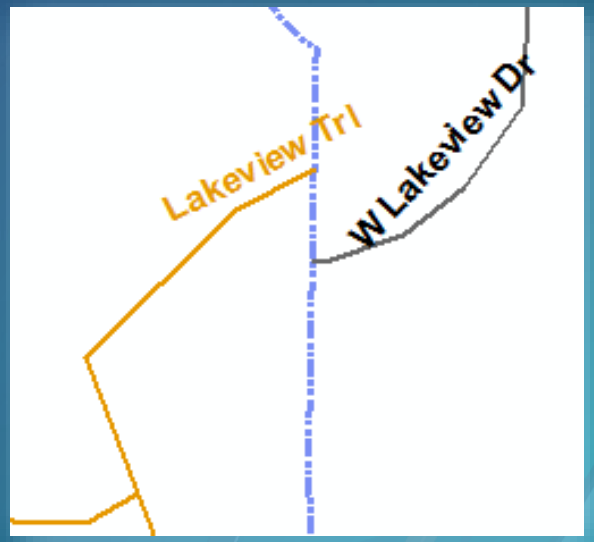
- Why Conflation?
- Geoprocessing Conflation Tools
  - User stories
- Conflation Workflows
  - Popular goals and workflow breakdowns
  - Transfer attributes workflow (demo)
- Conclusions and Future Work
- Q&A

Why Conflation?

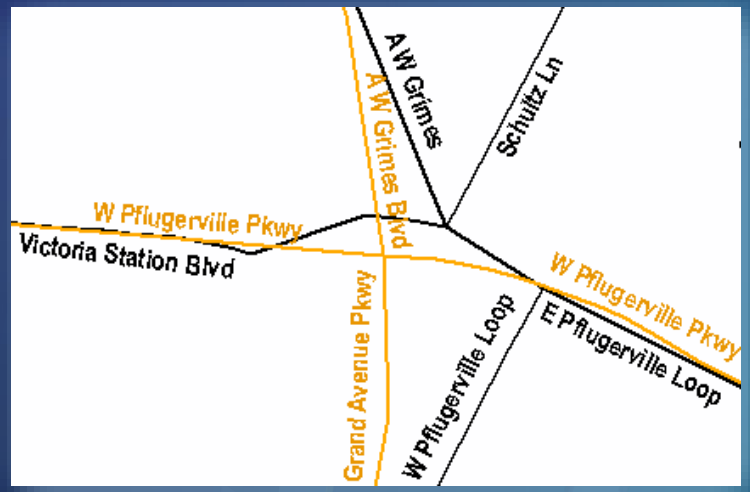
Common problems in analysis and mapping

- Spatial and attribute inconsistency in multi-source data, caused by differences in data collection and modeling
- It's challenging and costly to fix the problems

Adjacent datasets



Overlapping datasets



Conflation reconciles multi-source data and improves its quality and usability

Between overlapping datasets:

- Detect feature changes (differences) through feature matching
- Make spatial adjustment and attribute transfer

Between adjacent datasets:

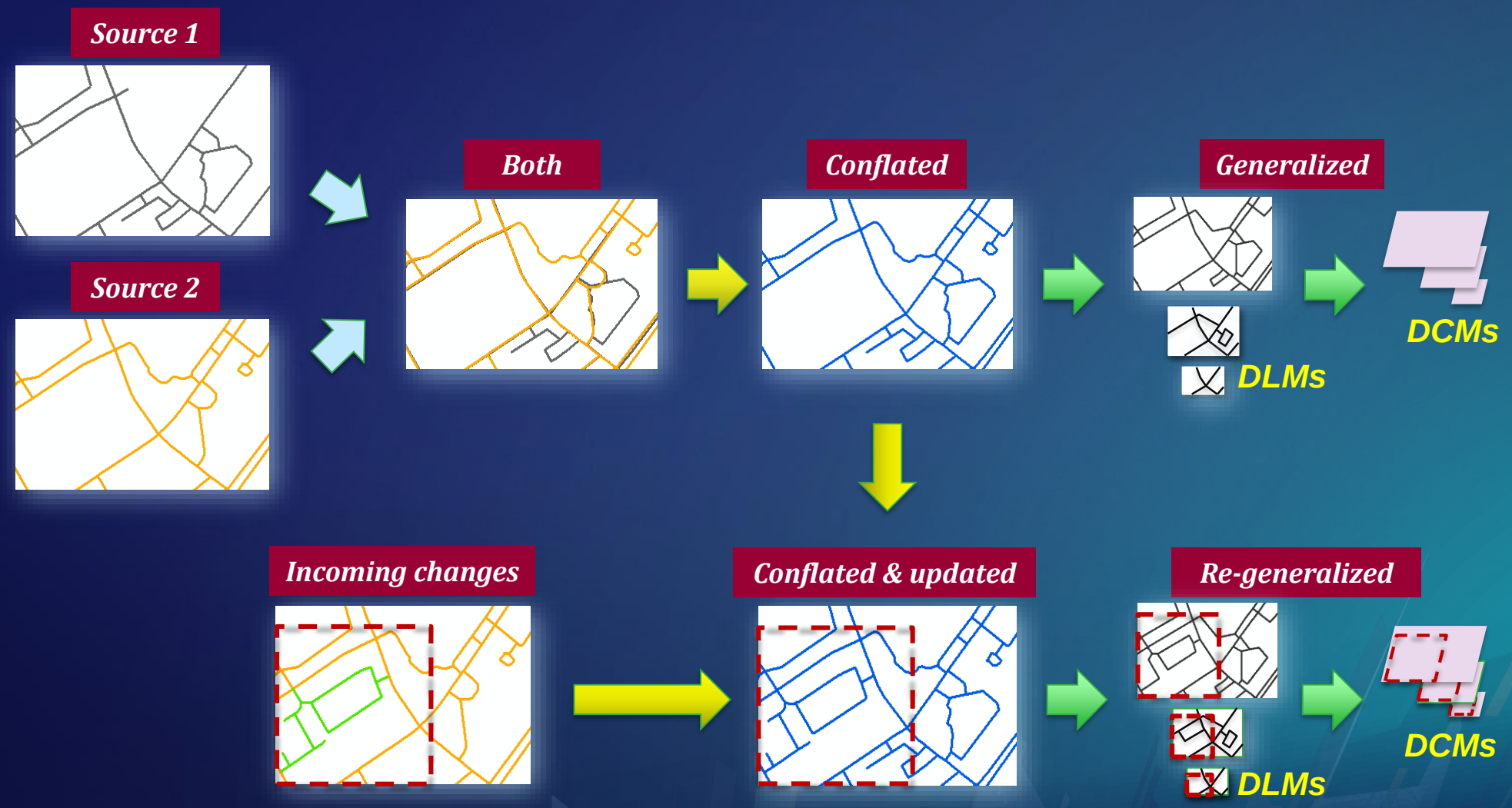
- Detect and resolve feature conflicts and disconnections through edge matching and alignment

Ultimate goals are to:

- Maintain an unified and seamless dataset – enriched and up-to-date
- No longer live with various imperfect datasets
- Rely on the data to perform analysis and quality mapping with confidence

Conflation in multi-scale data updating and mapping

DLM – digital landscape model; DCM – digital cartographic model



Geoprocessing Conflation Tools

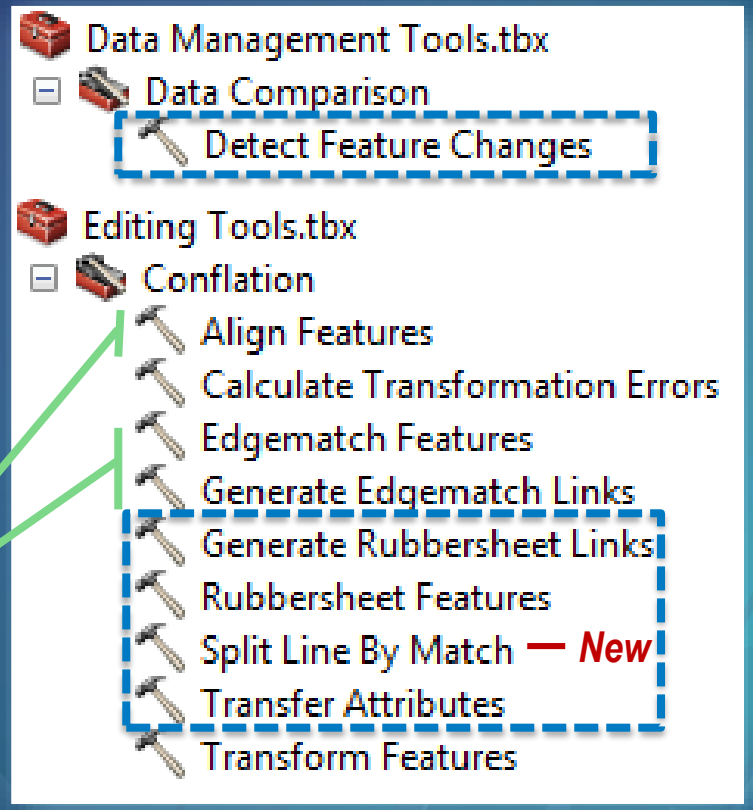
Highly automated geoprocessing tools

New and enhanced system tools

- Focusing on common use cases
- Aiming at high accuracy (not promising 100%)
- Providing information to facilitate post-inspection

Improved workflows

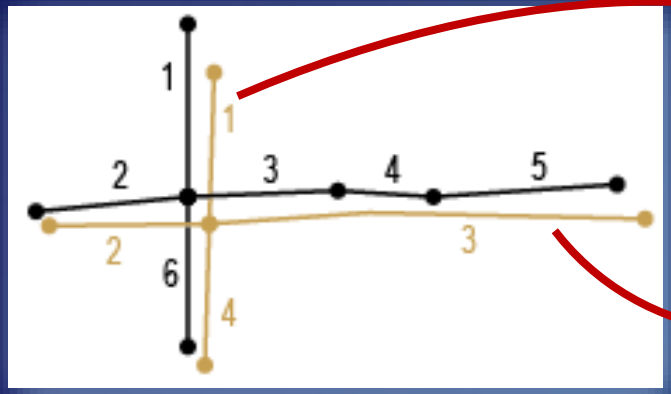
In ArcGIS 10.6.1 and Pro 2.2

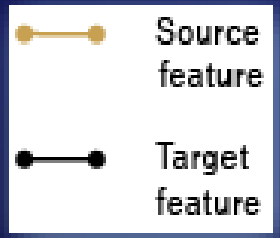


Conflation: Edgematching tools and workflows
11:15am – 12:00pm, Tuesday
Demo Theater 10 – Spatial Analysis

Feature matching (FM) for overlapping datasets

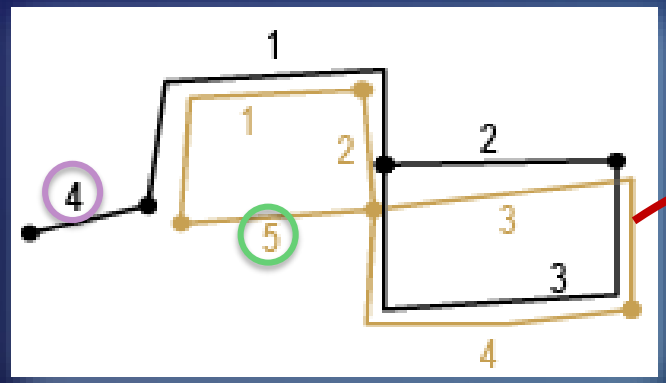
Based on proximity, topology, pattern, and similarity analysis, as well as attributes information





1:1 and 1:m matches

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

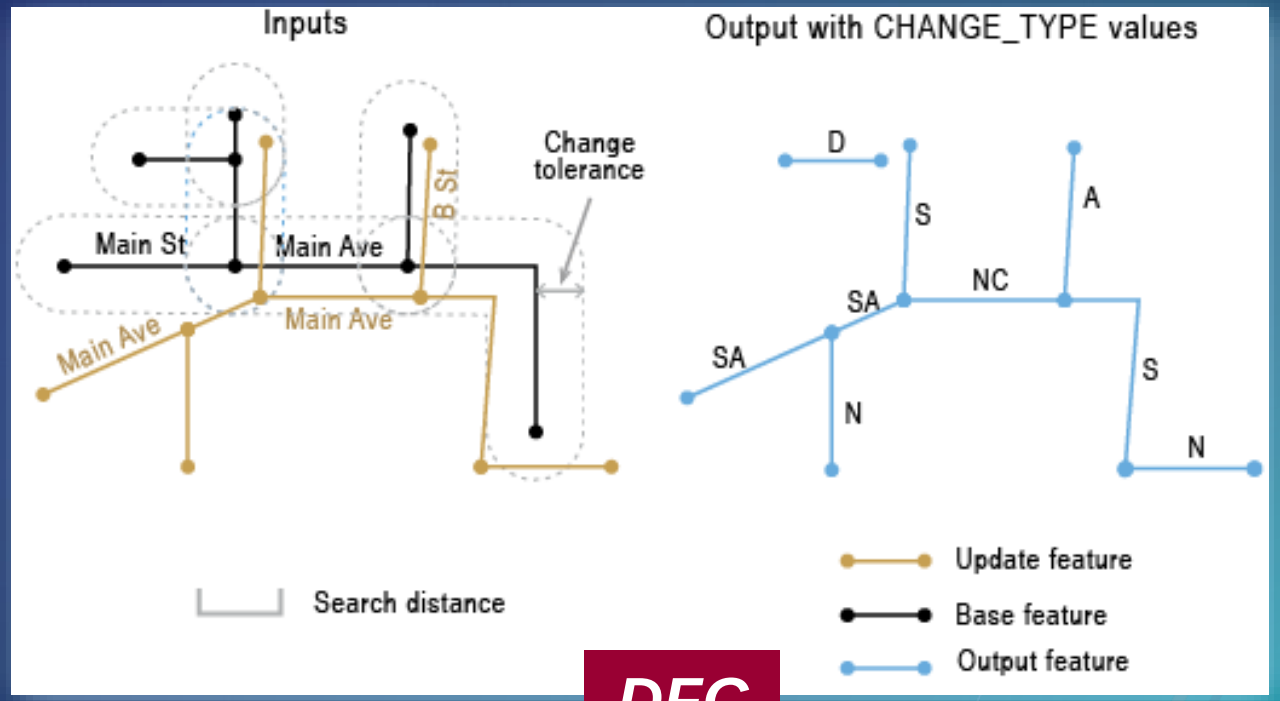
OBJECTID *	SRC_FID	TGT_FID	FM_GRP	FM_MN	FM_CONF
1	1	1	1	2:1	100
2	2	1	1	2:1	100
3	3	2	2	2:2	100
4	3	3	2	2:2	100
5	4	2	2	2:2	100
6	4	3	2	2:2	100
7	5	-1	-1	N/A	0
8	-1	4	-1	N/A	0

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

Finding 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)
- New update feature (N)
- To-be-deleted base feature (D)

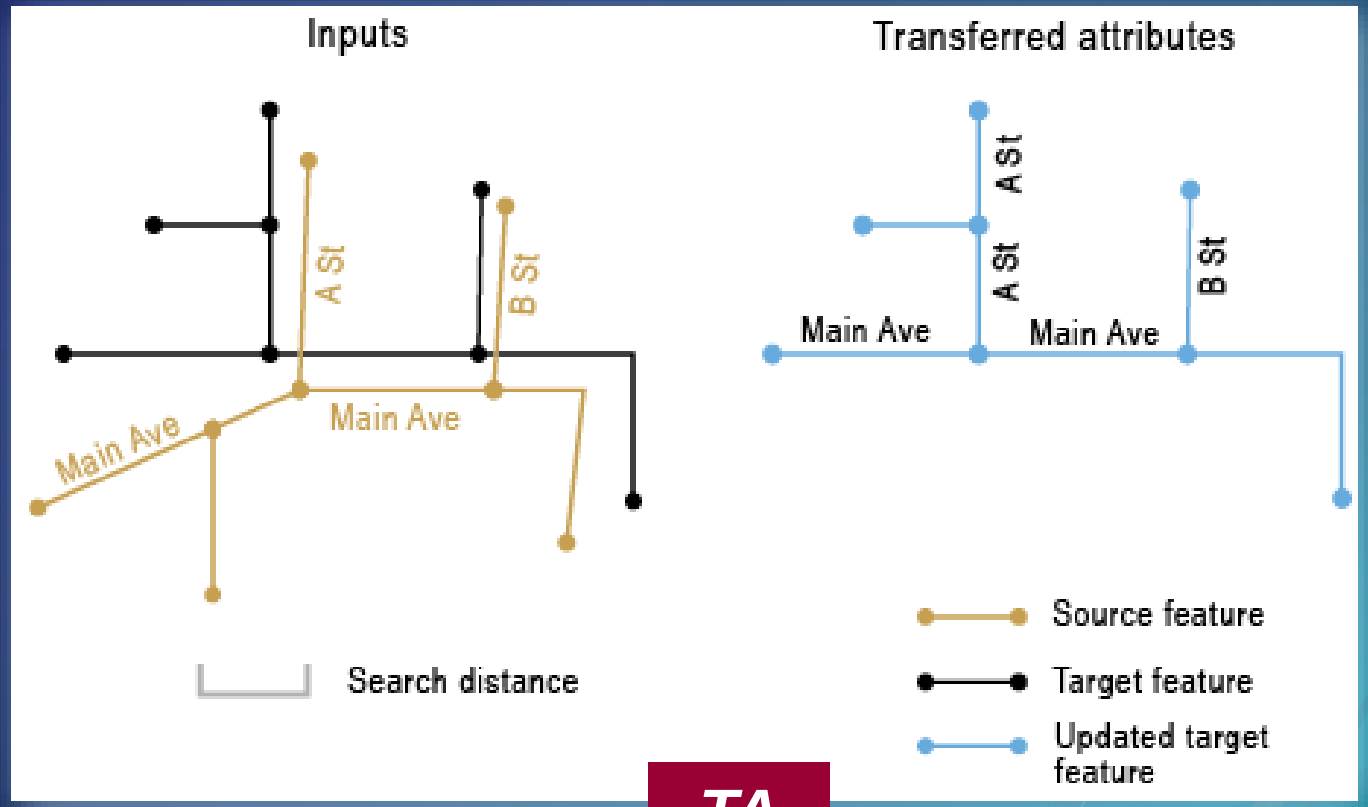


DFC

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

From source features to target features

- Transfer fields (e.g. ROAD_NAME, UniqueID)
- Target features are modified with the transferred fields

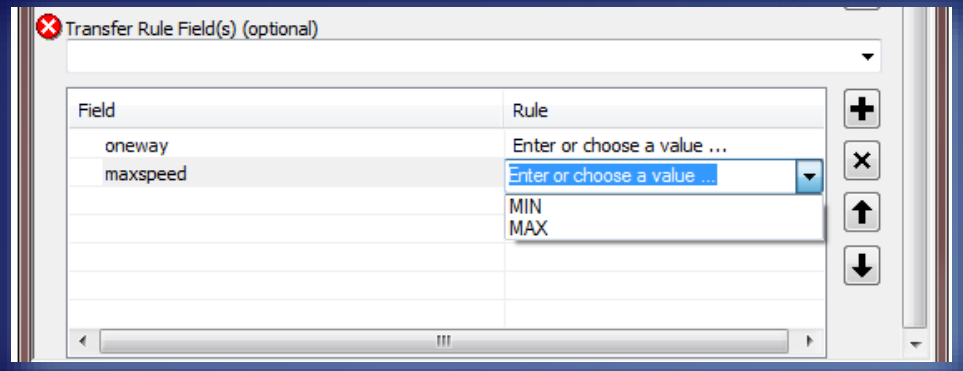


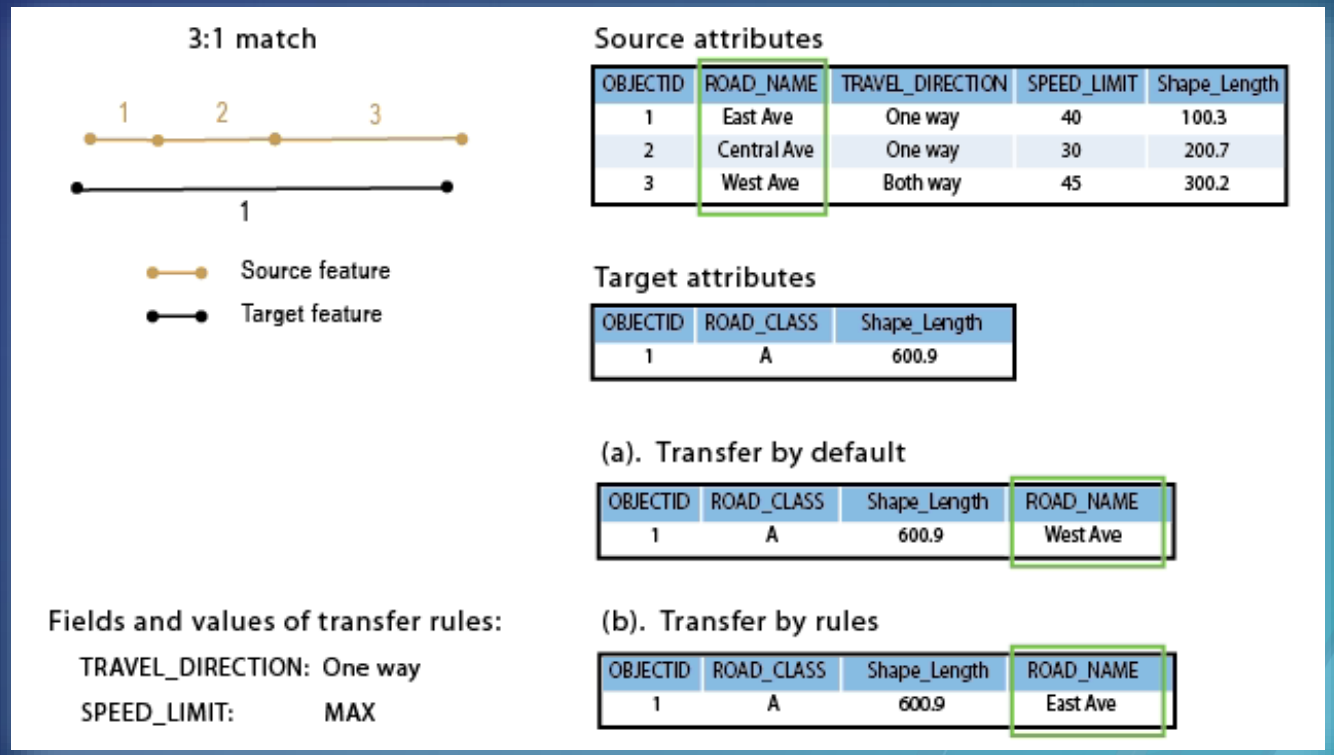
TA

What happens to m:n matches …

Transfer Rule Field(s) - for Transfer Attributes (TA)

To control m:n (m > 1) transfers based on source attribute information





Without rules (default):

- **Transfer from the longest source**

Ranked rules:

- **Field 1, rule value 1**
- **Field 2, rule value 2**
- **…**

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

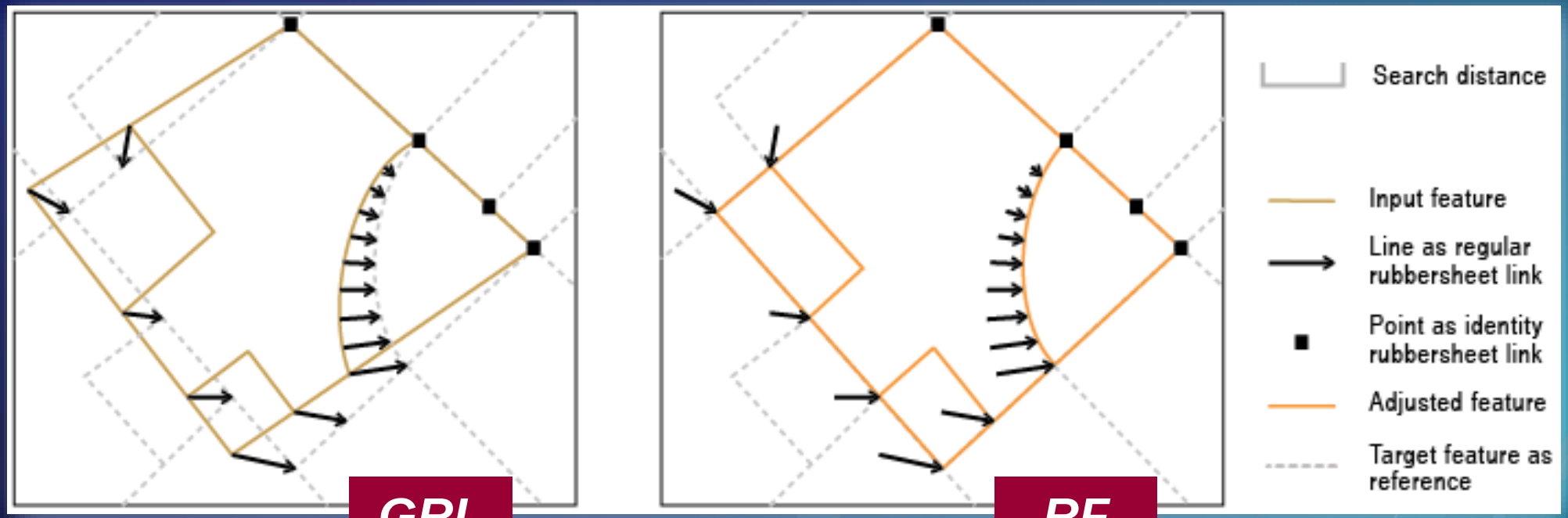
Rubbersheeting moves source locations towards target locations based on established links

Generate Rubbersheet Links (GRL)

- **From source features to target features**

Followed by Rubbersheet Features (RF)

- **Adjusting input features**



GRL ***RF***

Let's see how the tools have been used ...

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

County centerline attributes and direction must be retained.

Original_county_centerlines
Temecula_city_centerlines

Updated_county_centerlines
Original_county_centerlines

Updated_county_centerlines
Temecula_city_centerlines

- **Use DFC 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

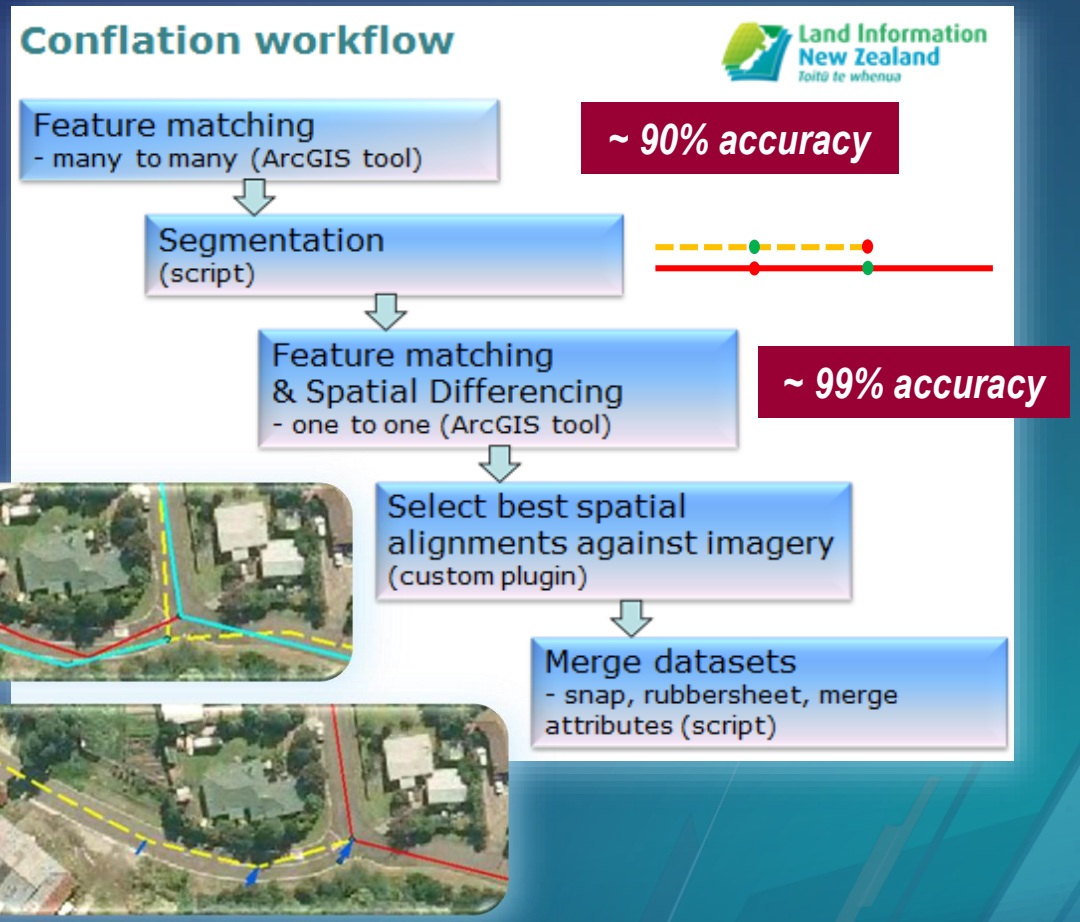
Data/information source: RCTLMA (Riverside County Transportation and Land Management) CA, USA
Acknowledgement: Thanks to Richard Fairhurst, for providing the information and screenshots.

User story 2: Combining electoral roads and topographic roads

There is no "most accurate" dataset.

Electoral roads

Topographic roads

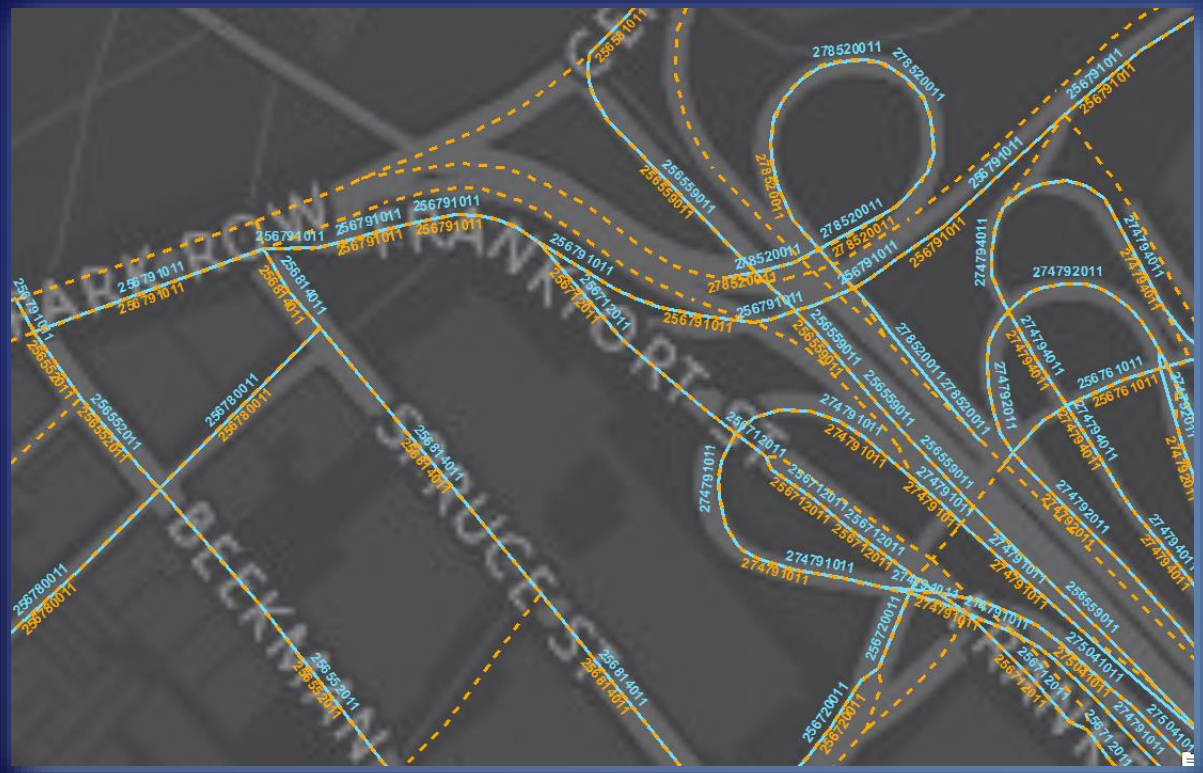


Information source: Land Information New Zealand (LINZ)

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

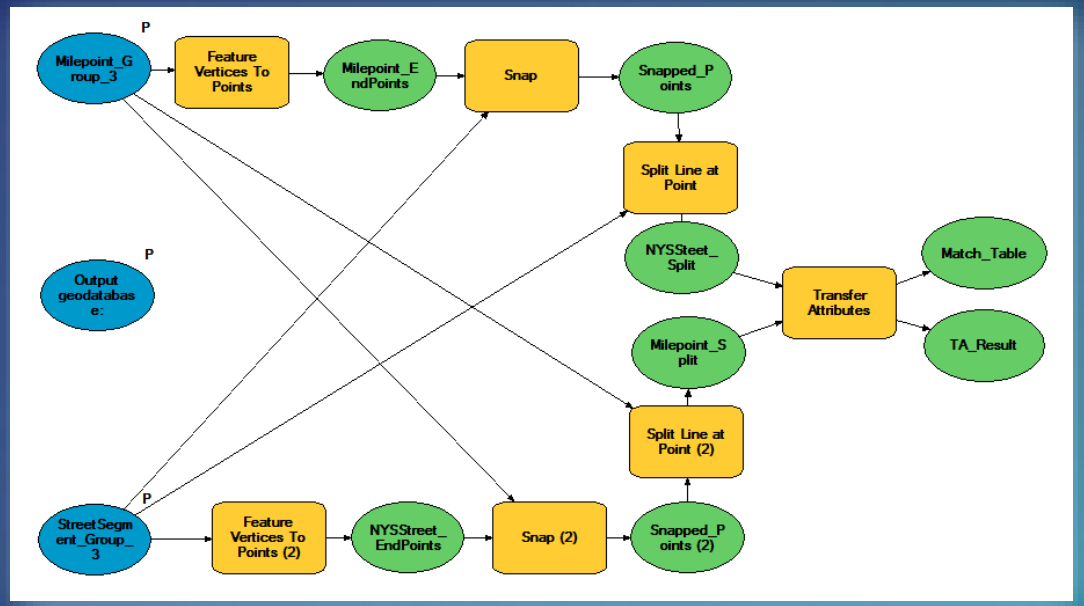
User story 3: Transferring attributes from State routes to Street segments

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

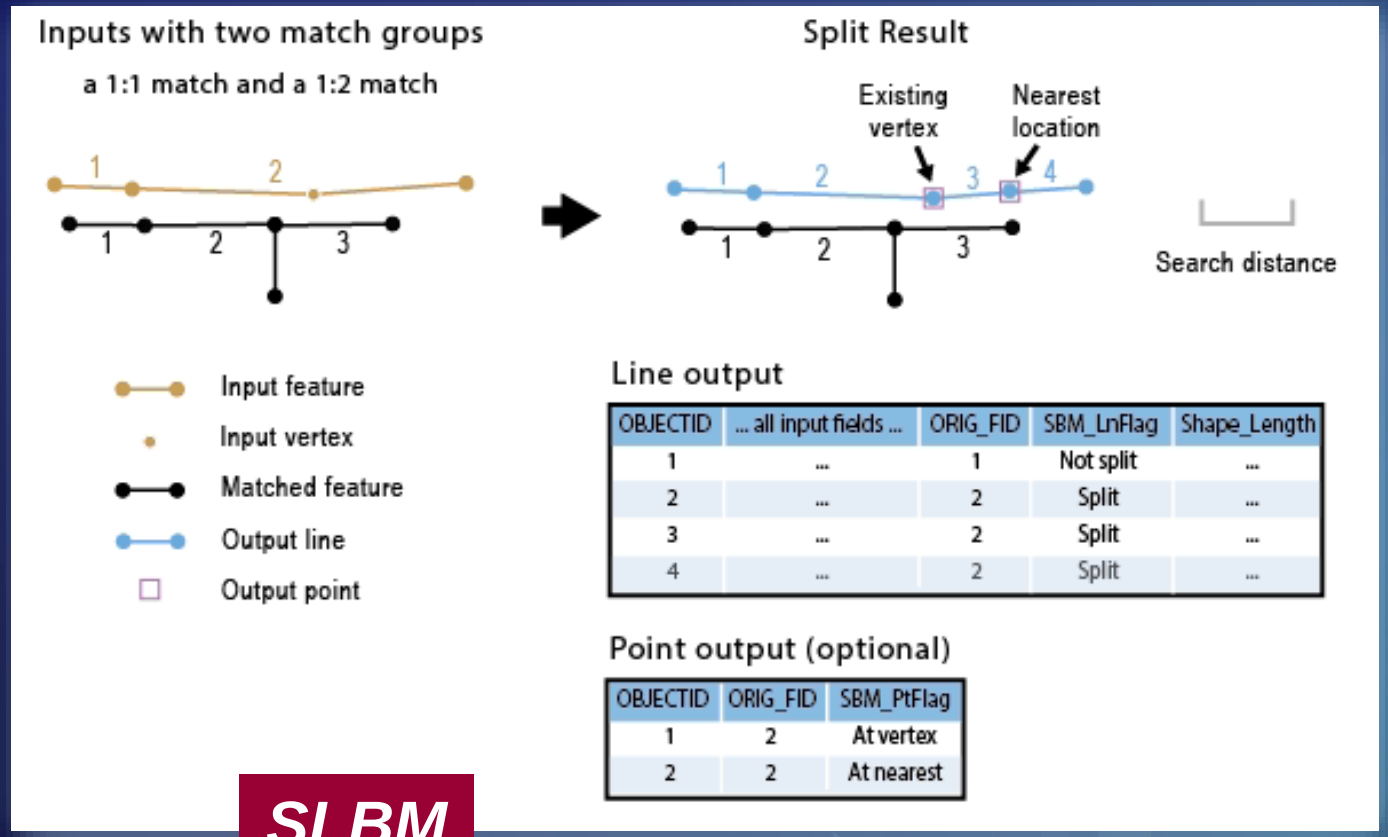
Data/information source: NYSDOT, USA

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

Split Line By Match (SLBM)

Splits input features based on matching relationships with matched features to obtain more 1:1 correspondences

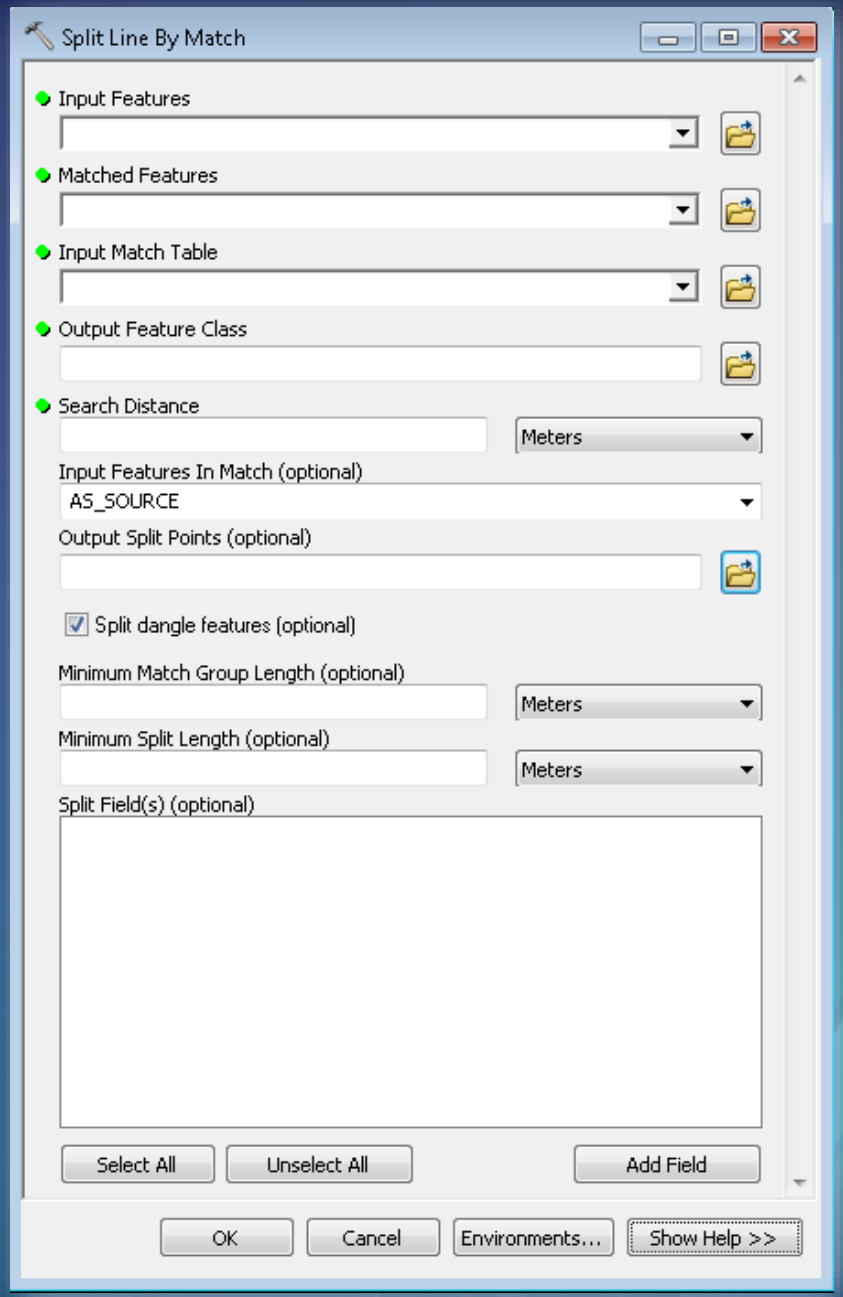
- **Pre-requires match table (e.g. generated by DFC)**
- **Multiple parameters**



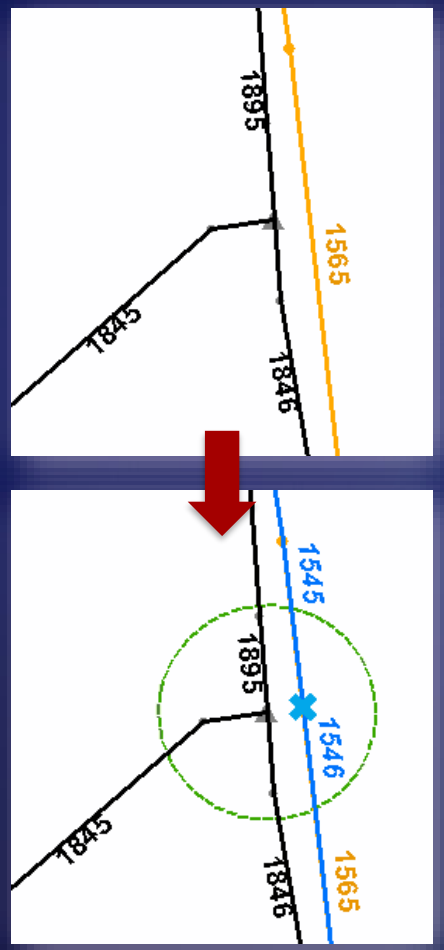
OBJECTID	... all input fields ...	ORIG_FID	SBM_LnFlag	Shape_Length
1	...	1	Not split	...
2	...	2	Split	...
3	...	2	Split	...
4	...	2	Split	...

OBJECTID	ORIG_FID	SBM_PtFlag
1	2	At vertex
2	2	At nearest

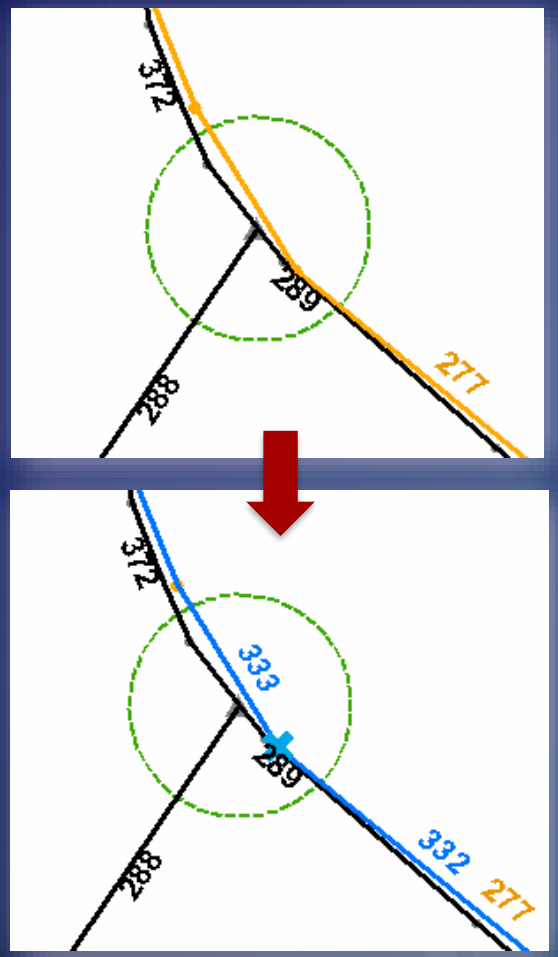
SLBM



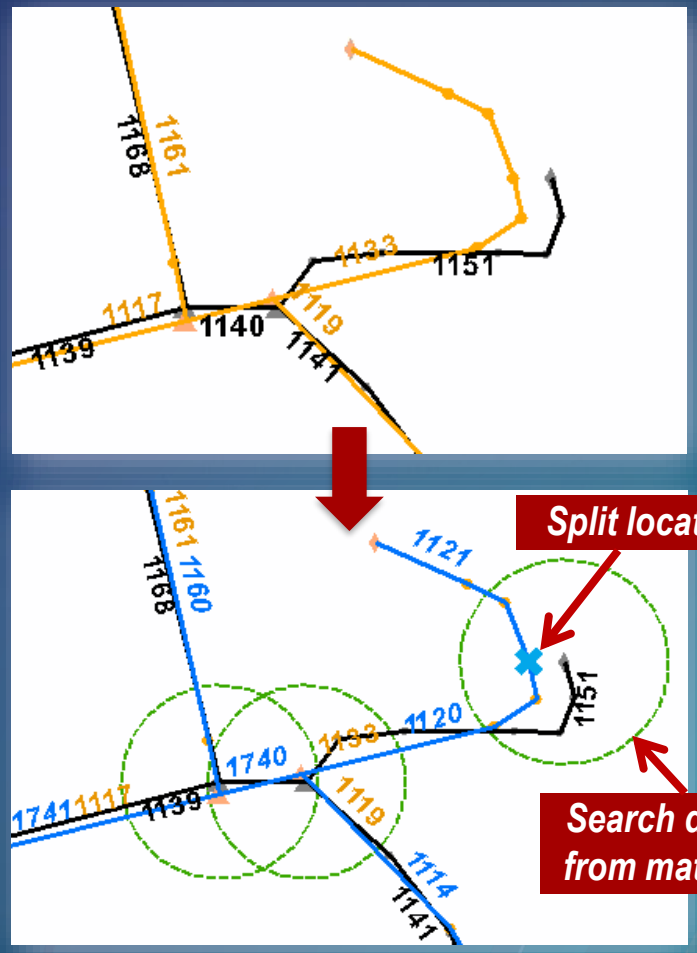
Example results of Split Line By Match



Split at the nearest location



Split at the only vertex

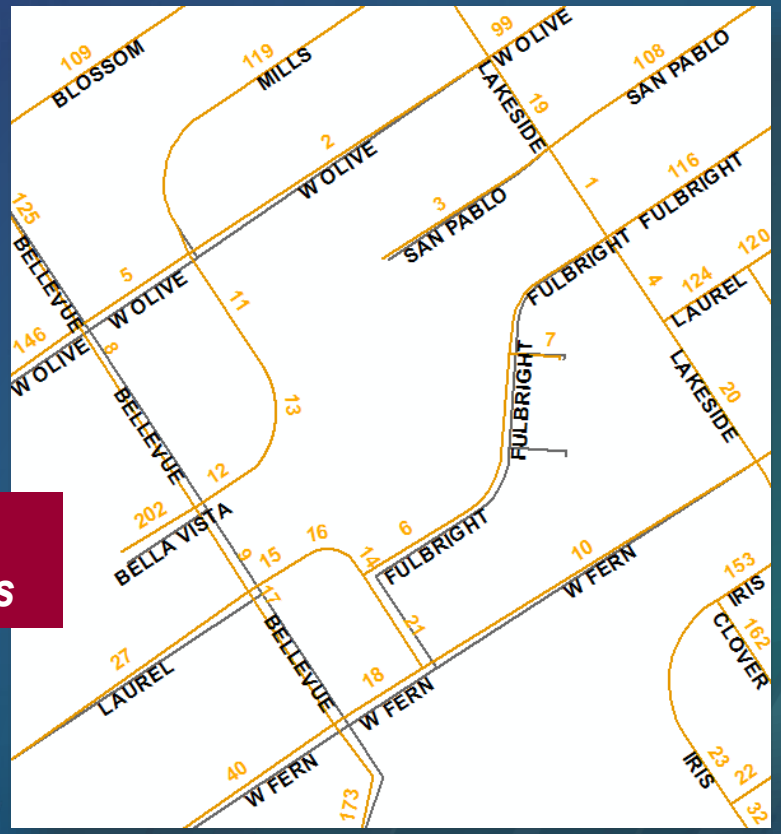


Split at the nearest vertex

Conflation Workflows

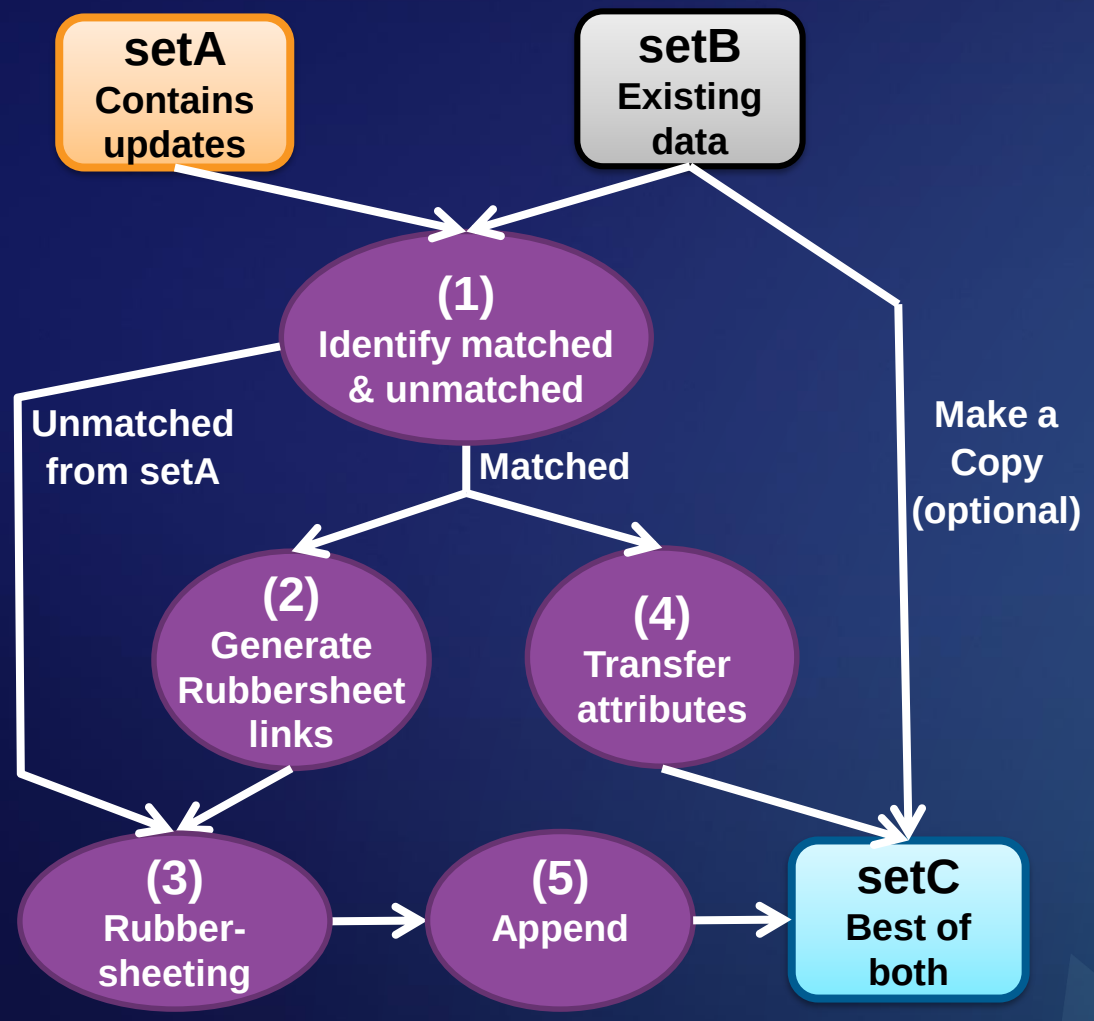
Unification of two overlapping datasets

- One of the popular goals



Two overlapping, inconsistent datasets

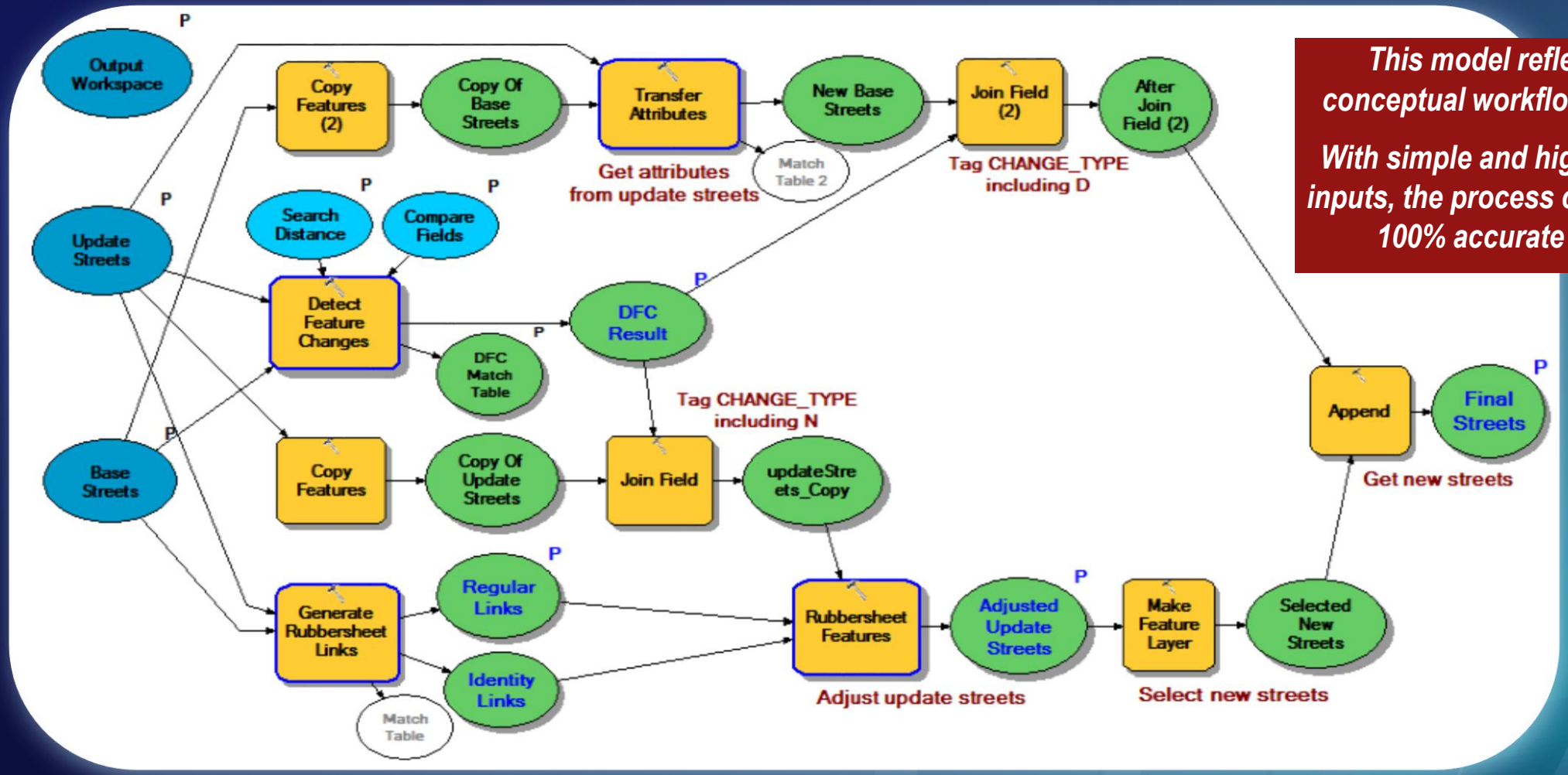
Conceptual workflow strategy



Key processes:

- Identify feature differences
- Make spatially adjustment
- Transfer attributes for matched features
- Add new features; delete old features if necessary

Conflate overlapping datasets

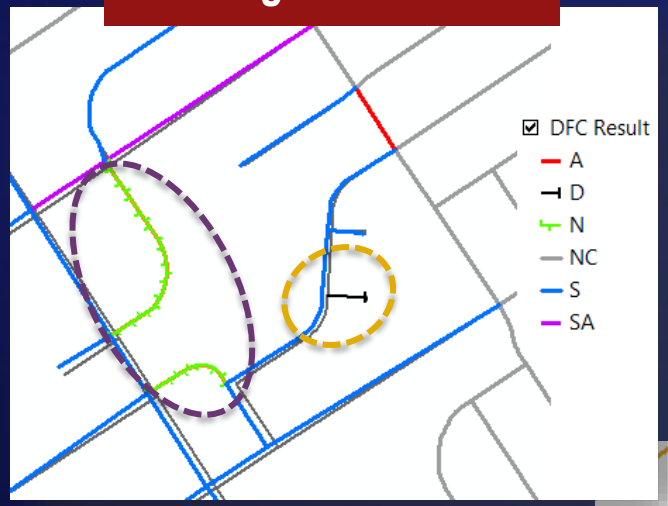


This model reflects the conceptual workflow strategy.

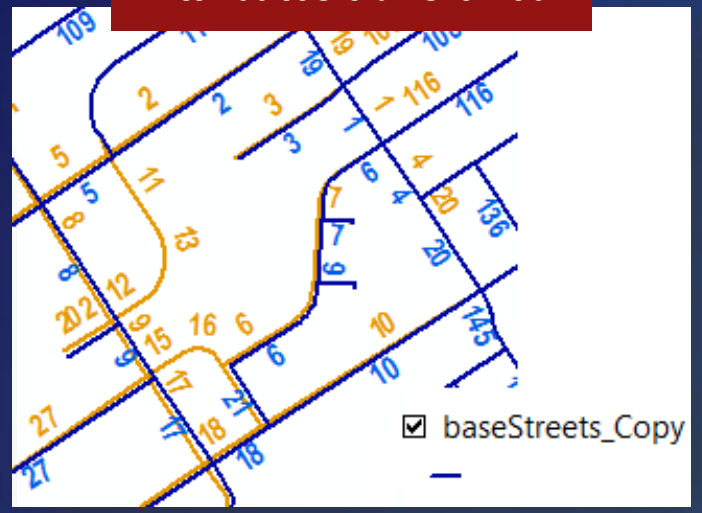
With simple and highly similar inputs, the process can produce 100% accurate result.

Perfect results of simple data

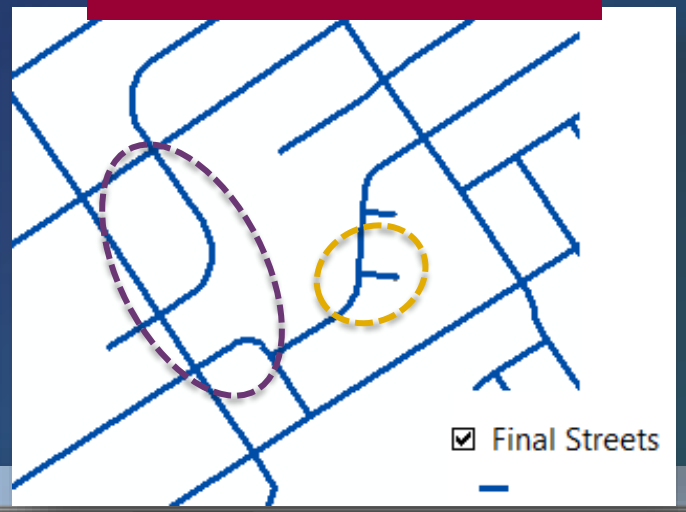
Changes detected



Attributes transferred



New features adjusted and added to base



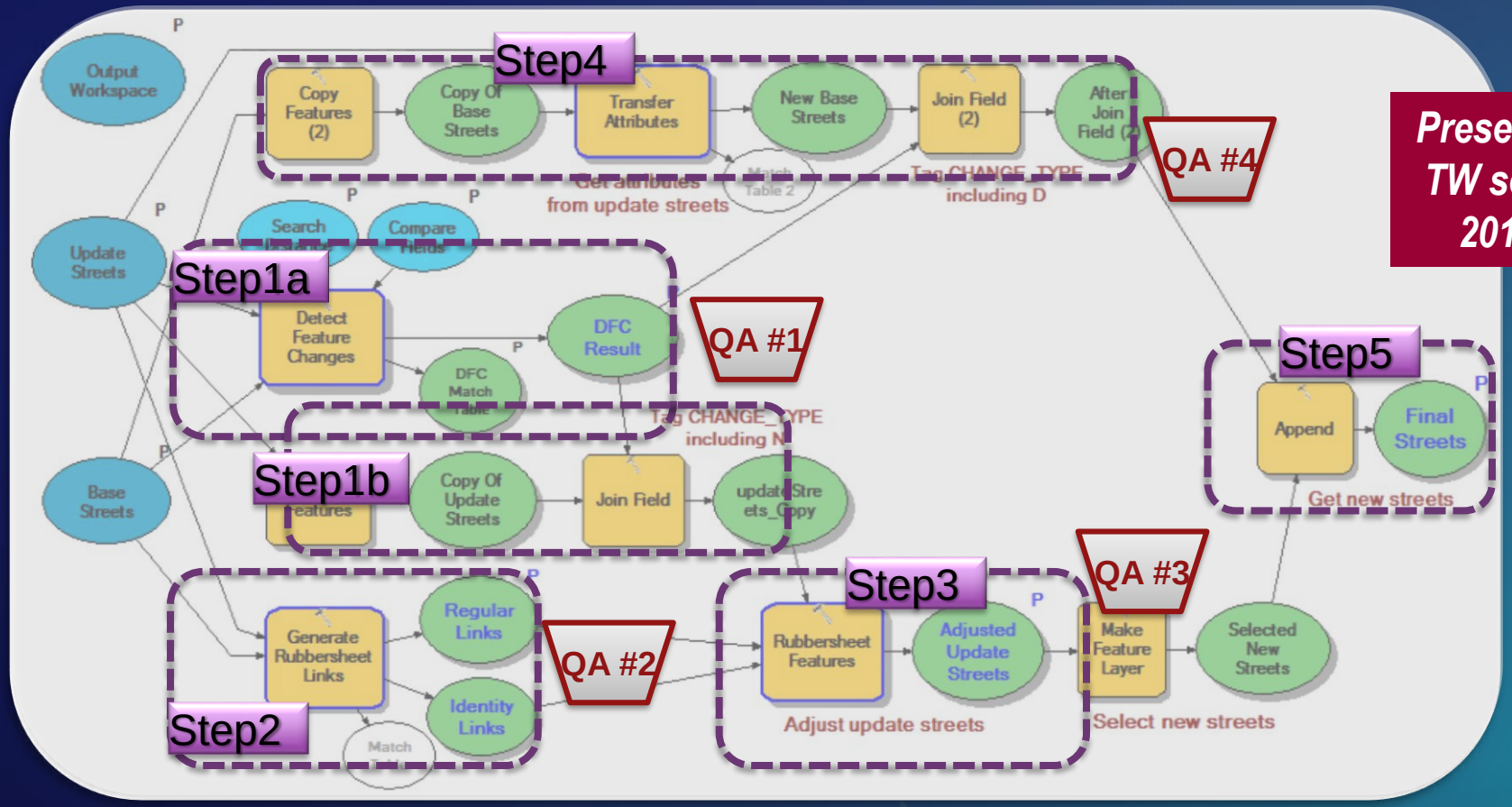
Final Streets

OBJECTID *	Shape *	NAME	Priority	Shape_Length	UniqueID	CHANGE_TYPE
100	Polyline	W OLIVE	M	293.577825	2	A
109	Polyline	W OLIVE	M	109.071651	5	A
128	Polyline	<Null>	L	51.455286	<Null>	D
200	Polyline	<Null>	M	112.706047	11	N
201	Polyline	<Null>	M	59.242967	12	N
202	Polyline	<Null>	M	90.29932	13	N
203	Polyline	<Null>	M	18.383534	14	N
204	Polyline	<Null>	M	58.252814	15	N
205	Polyline	<Null>	M	37.23	16	N
6	Polyline	LAUREL	L	384.916164	27	NC
19	Polyline	W FERN	M	340.788763	40	NC
66	Polyline	MILLS	L	90.060797	87	NC
67	Polyline	KATHL	L	173.068809	88	NC
69	Polyline	MILLS	L	91.095418	90	NC

Rubbersheet links generated

Of course the reality is more complicated …

Breakdown of conceptual workflow into sub-workflows



Presented at UC TW sessions in 2014 - 2016

http://proceedings.esri.com/library/userconf/proc16/tech-workshops.html

Google search by "Esri UC technical workshop presentations"

Conflation workflow in real world scenarios

Pre-processing → **Conflation and evaluation** → **Post-processing**

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

- Conflation tools
- Workflow tools

- Queued review
- Interactive editing

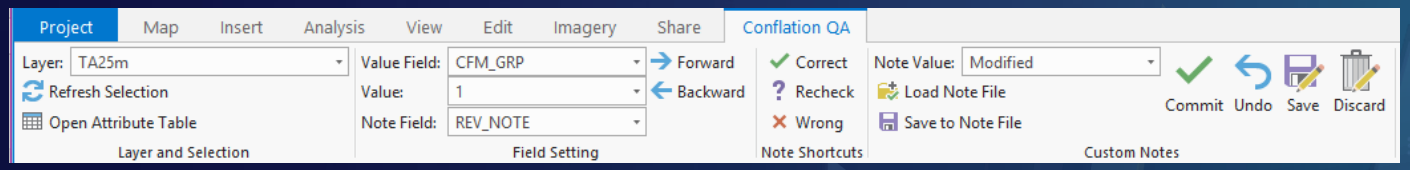
Supplemental Conflation Workflow Tools

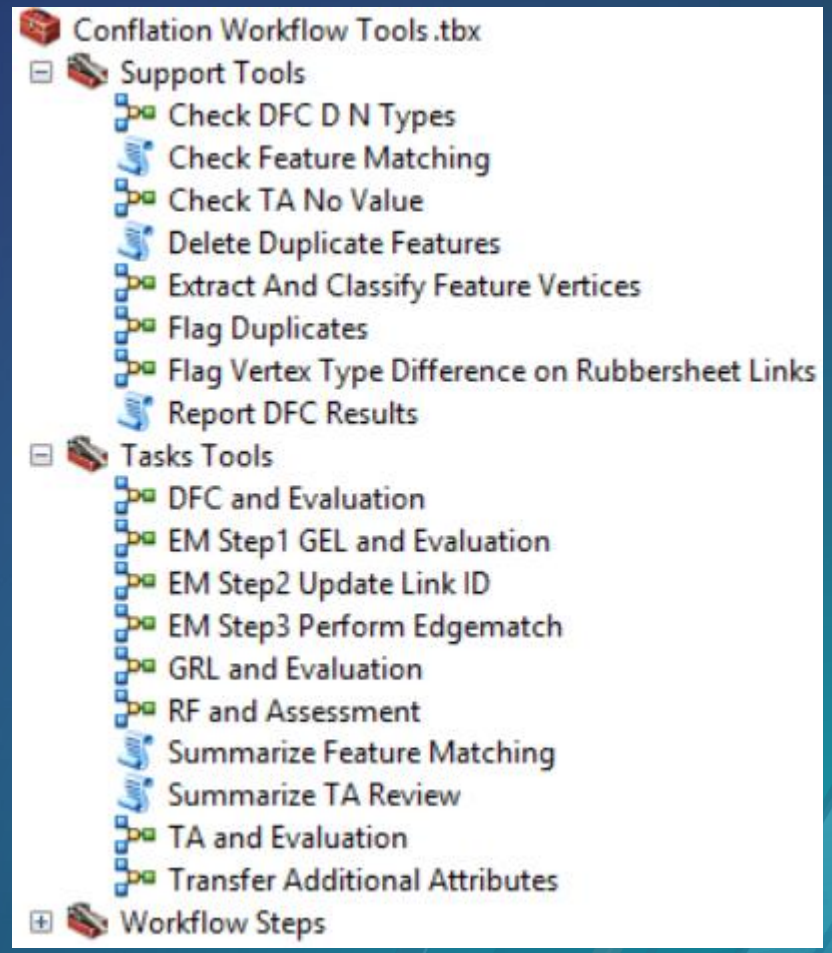
Popular tasks:

- Transfer attributes
- Spatial adjustment
- Detect feature changes
- Edge matching

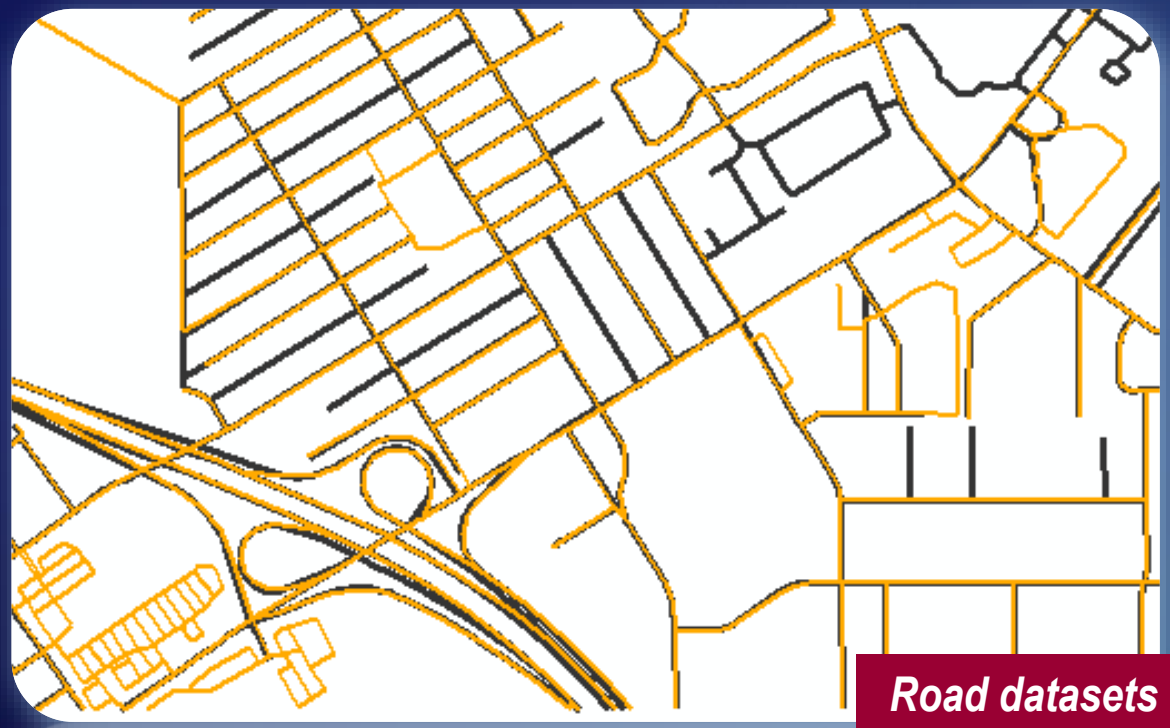
Focusing on tasks:

- Task specific tool set
- Enriched results to facilitate inspections
- Enhanced interactive tools (python add-in for ArcMap; SDK add-in for Pro)





This demo shows TA workflow in Pro 2.2 …

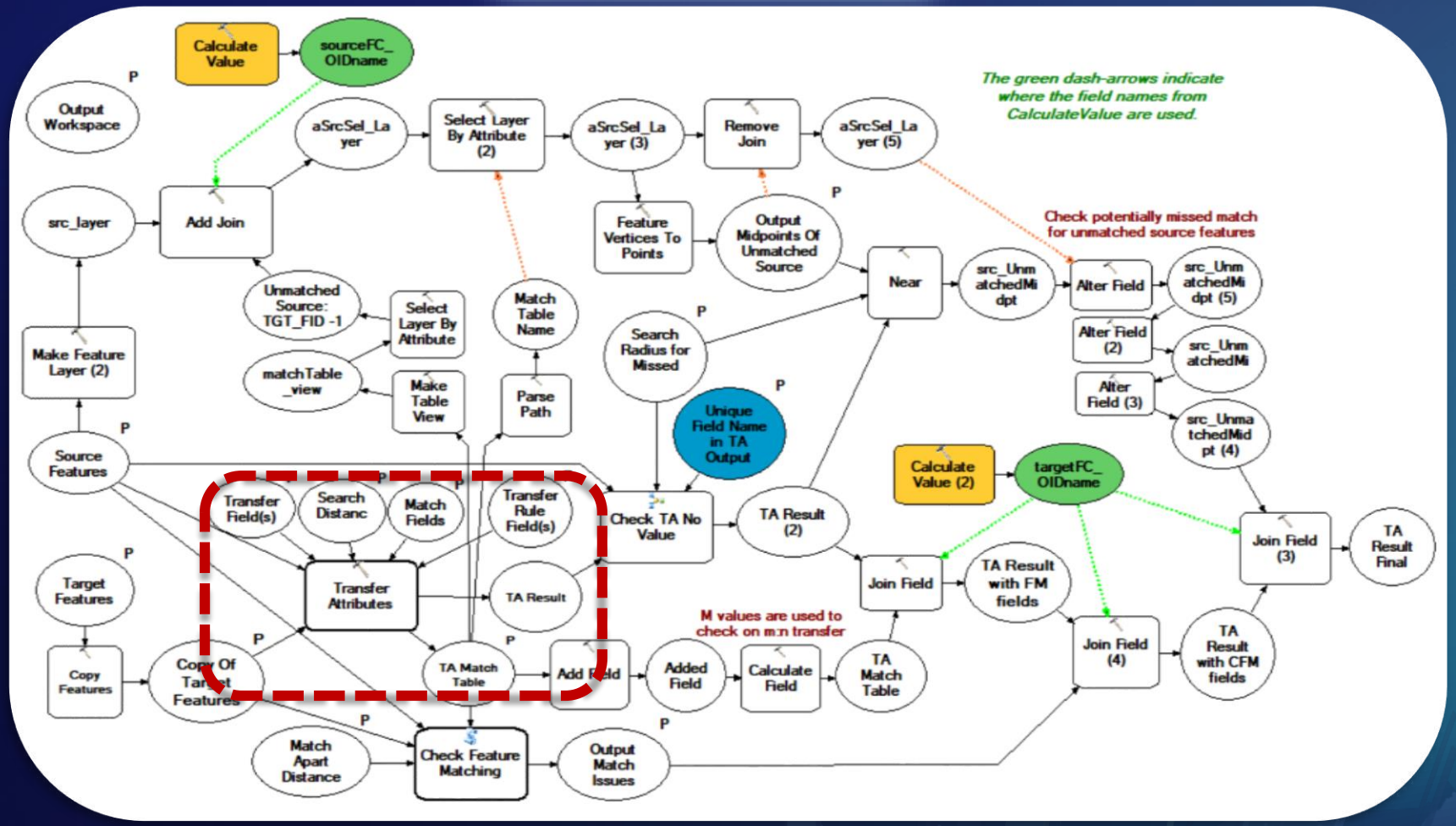


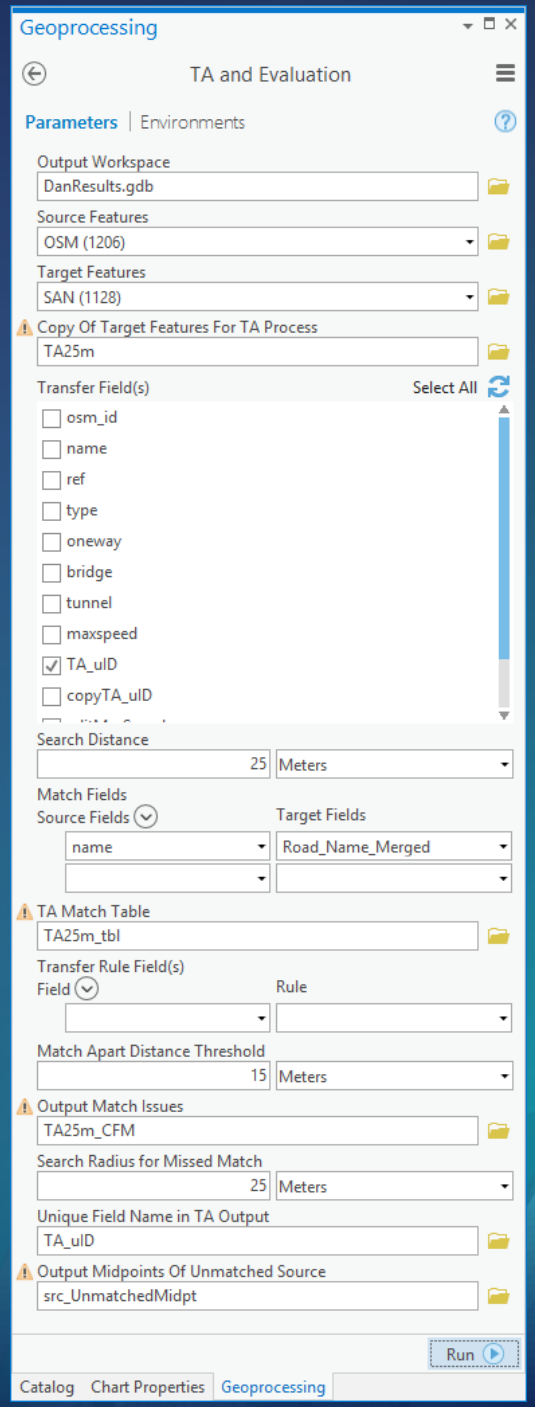
Road datasets from two sources

Transfer Attributes workflow (demo)

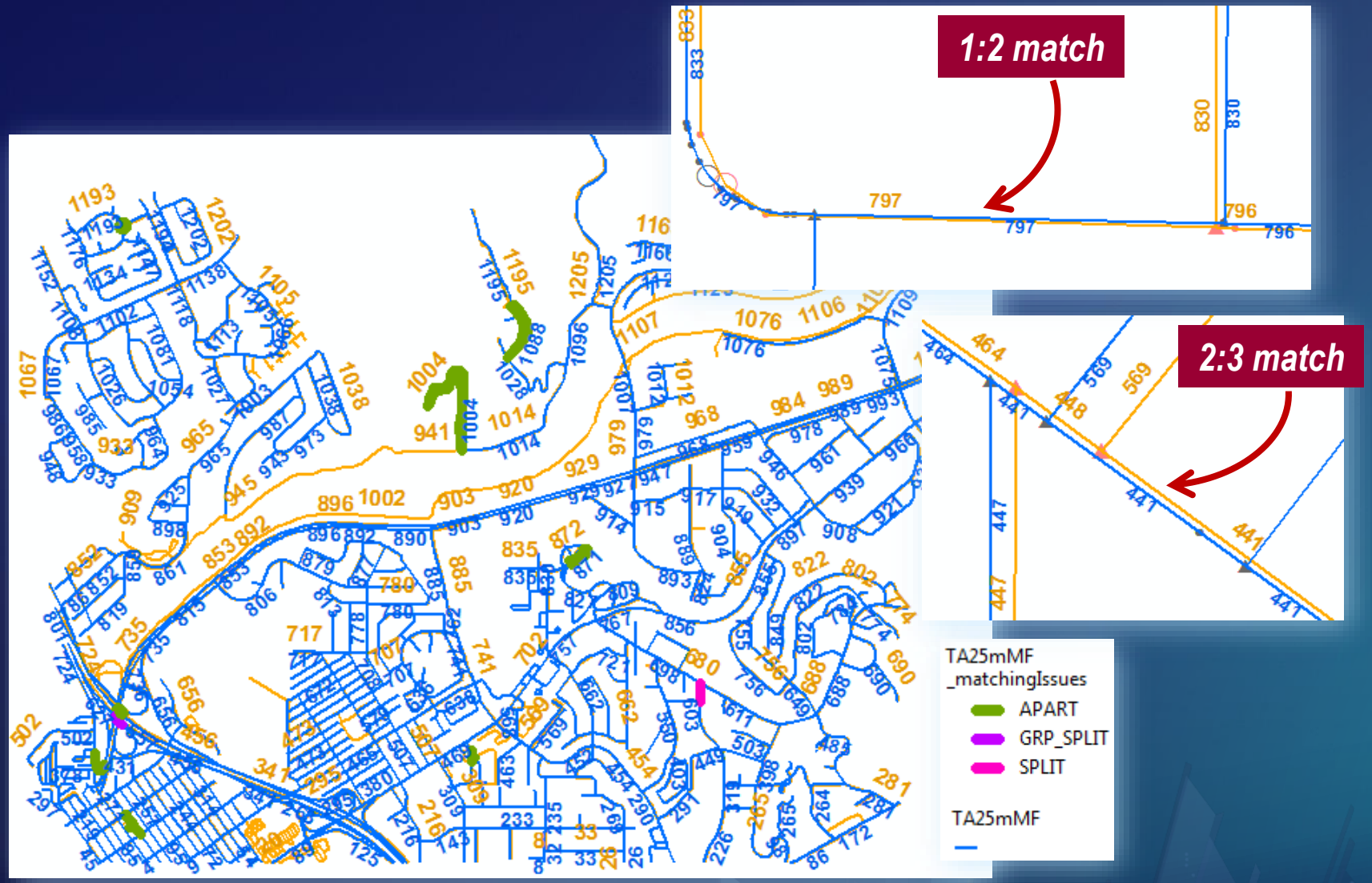
TA and Evaluation

Potential issues are flagged in a copy of the target





TA result and flags on potential issues



QA

- **Review potential wrong transfers**
- **Review potential missed transfers**
- ***Multi-source transfers (m:n) are noted in srcM_inMN***

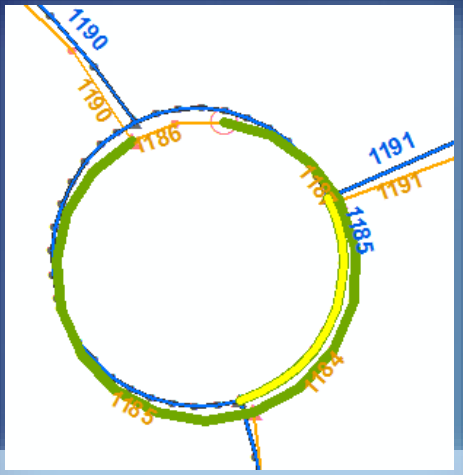
Demo of TA QA …

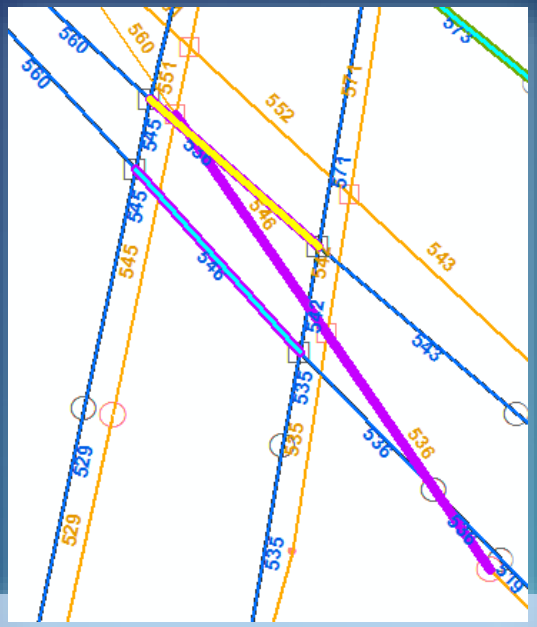
Review transfers with potential match issues

CFM_GRP >=0

13 records were reviewed:

- **2 TA_uID values were corrected**





TA25mMF

OBJECTID	Shap	TBMG	RD20NA	RD	RD20FU	Road_N	Shape_L	TA_uID	NEAR_FID	NEAR_DIS	FM_GRP	FM_MN	srcM_in	CFM_	CFM_FL	CFM_DI	srcORIGFID	srcNearFID	srcNearDis	REV_FLAG	rev_TAuID
664	Polyli	A3	PRIVATE	RD	PRIVAT	Private R	32.98117	1185	<Null>	<Null>	884	3:1	3	5	APART	17.7133	<Null>	<Null>	<Null>	Wrong	1184
147	Polyli	A6	I-5 SB		I-5 SB	Interstate	24.82510	536	<Null>	<Null>	462	2:1	2	3	GRP_SP	-1	552	147	6.838031	Wrong	552
275	Polyli	A6	I-5 NB		I-5 NB	Interstate	23.79797	585	<Null>	<Null>	234	2:1	2	1	APART	18.1100	<Null>	<Null>	<Null>	Correct	<Null>
400	Polyli	A6	I-5 NB		I-5 NB	Interstate	23.62526	573	<Null>	<Null>	233	1:1	1	6	APART	16.8328	<Null>	<Null>	<Null>	Correct	<Null>
458	Polyli	C4	BENET HIL	RD	BENET	Benet Hill	275.3343	1088	<Null>	<Null>	810	1:1	1	11	APART	19.3391	<Null>	<Null>	<Null>	Correct	<Null>
484	Polyli	B4	PRIVATE	RD	PRIVAT	Private R	744.7975	1004	<Null>	<Null>	777	1:1	1	10	APART	27.7146	1005	484	12.868726	Correct	<Null>
611	Polyli	D6	CROUCH	ST	CROUC	Crouch S	101.9074	643	<Null>	<Null>	527	2:1	2	4	SPLIT	-1	<Null>	<Null>	<Null>	Correct	<Null>
752	Polyli	A7	ALLEY		ALLEY	Alley	138.9000	208	<Null>	<Null>	238	1:1	1	7	APART	18.5957	<Null>	<Null>	<Null>	Recheck	<Null>
890	Polyli	C5	ECCLESIA	DR	ECCLES	Ecclesia	133.4632	871	<Null>	<Null>	667	1:1	1	9	APART	20.6245	<Null>	<Null>	<Null>	Recheck	<Null>
897	Polyli	C6	PRIVATE	RD	PRIVAT	Private R	58.41161	420	<Null>	<Null>	373	1:1	1	8	APART	15.5651	<Null>	<Null>	<Null>	Recheck	<Null>
961	Polyli	A6	I-5 SB OF	RA	I-5 SB O	Interstate	27.39497	546	<Null>	<Null>	457	1:1	1	3	GRP_SP	-1	<Null>	<Null>	<Null>	Correct	<Null>
970	Polyli	J7	SR-76 WB		SR-76	State Ro	148.1798	401	<Null>	<Null>	363	2:1	2	2	APART	20.8946	<Null>	<Null>	<Null>	Correct	<Null>
1061	Polyli	C3	BENET HIL	RD	BENET	Benet Hill	91.54481	1089	<Null>	<Null>	811	1:1	1	12	APART	16.9071	<Null>	<Null>	<Null>	Correct	<Null>

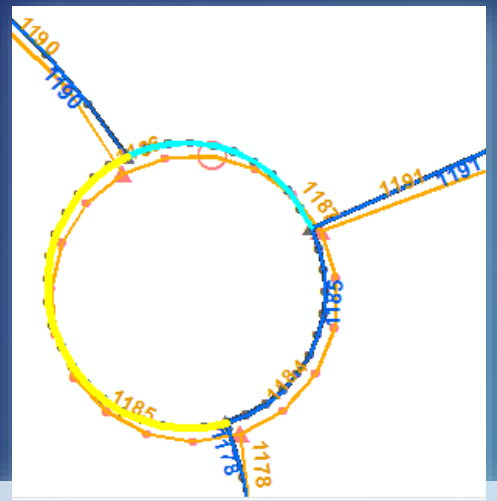
0 (13 out of 1128 Selected)

Review potential missed transfers

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

48 records were reviewed:

- **6 transferred TA_uID values were incorrect**



TA25mMF

OBJECTID	Shap	TBMG	RD20NA	RD	RD20FU	Road_N	Shape_L	TA_uID	NEAR_FID	NEAR_DIS	FM_GRP	FM_MN	srcM_in	CFM_	CFM_FL	CFM_DI	srcORIGFID	srcNearFID	srcNearDis	REV_FLAG	rev_TAuID
401	Polyli	J6	I-5 NB		I-5 NB	Interstate	40.03363	<Null>	731	9.641416	-1	N/A	-1	<Null>	<Null>	<Null>	731	401	9.630519	Wrong	731
419	Polyli	C5	ECCLESIA	DR	ECCLES	Ecclesia	21.15319	<Null>	871	0.288407	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Correct	<Null>
421	Polyli	J3	PRIVATE	RD	PRIVAT	Private R	57.36360	<Null>	1185	0.214332	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	1185
553	Polyli	B7	I-5	RA	I-5 RA	Interstate	35.14299	<Null>	165	1.88825	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Recheck	<Null>
554	Polyli	B7	I-5	RA	I-5 RA	<Null>	38.92028	<Null>	197	12.469526	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Recheck	<Null>
566	Polyli	B7	I-5	RA	I-5 RA	<Null>	24.05586	<Null>	213	4.873462	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Recheck	<Null>
593	Polyli	C6	PRIVATE	DY	PRIVAT	<Null>	29.27072	<Null>	613	12.225035	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Correct	<Null>
609	Polyli	B7	MCNEIL	ST	MCNEIL	Mcneil St	16.74027	<Null>	358	9.022339	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Correct	<Null>
625	Polyli	E7	MACDON	ST	MACDO	Macdona	11.06622	<Null>	325	0.525794	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	325
703	Polyli	B6	PRIVATE	RD	PRIVAT	Private R	29.63157	<Null>	541	12.521015	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Correct	<Null>

0 (31 out of 1128 Selected)

Attribute transfer accuracy estimates

- **Total target features: 1128**
- **Correct: 1120**
- **Incorrect: 8**

Accuracy = 1120/ 1128 => 99.29%

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

All correct

Post transfer if necessary

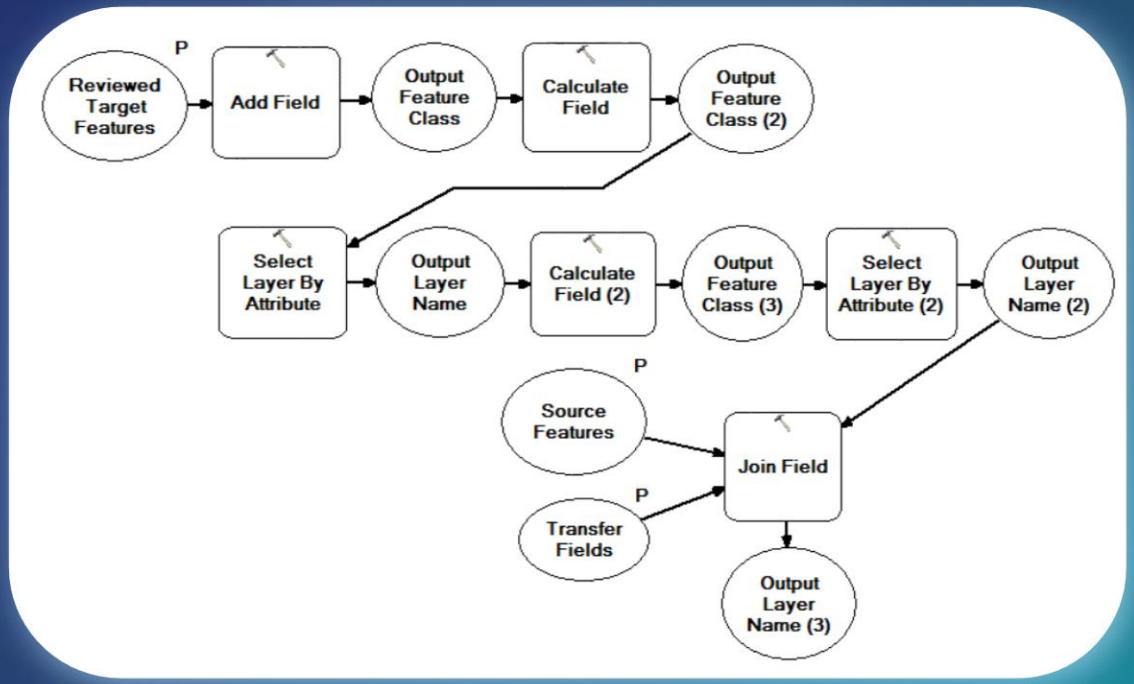
Based on updated final_TA_uID

Calculate final_TA_uID

- **Add a field final_TA_uID**
- **Calculate final_TA_uID = TA_uID**
- **Select rev_flag = ‘Wrong’**
- **Calculate final_TA_uID = rev_TA_uID**

Transfer additional attributes

- **Use Join Field on source TA_uID and target final_TA_uID to transfer additional attributes**



RD20FU	Road_	Shape_Len	TA_uID	NEAR_FID	NEAR_DIS	FM_GRP	FM_MN	srcM_in	CFM_	CFM_FL	CFM_DI	srcORIGFID	srcNearFID	srcNearDis	REV_FLAG	rev_TAuID	final_TAuID	oneway	bridge
PRIVAT	Private	30.400859	<Null>	1187	1.463928	-1	N/A	-1	<Null>	<Null>	<Null>	1186	816	1.876831	Wrong	1187	1187	1	0
PRIVAT	Private	57.363602	<Null>	1185	0.214332	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	1185	1185	1	0
PRIVAT	Private	32.981175	1185	<Null>	<Null>	884	3:1	3	5	APART	17.7133	<Null>	<Null>	<Null>	Wrong	1184	1184	1	0
CAPIST	Capistr	0.069553	<Null>	850	0.02129	-1	N/A	-1	<Null>	<Null>	<Null>	850	1018	3.947928	Wrong	861	861	1	0
I-5 NB	Intersta	40.03363	<Null>	731	9.641416	-1	N/A	-1	<Null>	<Null>	<Null>	731	401	9.630519	Wrong	731	731	1	1
I-5 SB	Intersta	24.825103	536	<Null>	<Null>	462	2:1	2	3	GRP_SP	-1	552	147	6.838031	Wrong	552	552	1	1
MACDO	Macdon	11.066227	<Null>	325	0.525794	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	Wrong	325	325	0	0
PRIVAT	<Null>	10.550205	<Null>	-1	-1	-1	N/A	-1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>
I-5 SB	Intersta	28.867189	543	<Null>	<Null>	8	1:1	1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	543	1	1
I-5 SB	Intersta	9.015739	534	<Null>	<Null>	7	1:2	1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	534	1	0
I-5 SB	Intersta	283.923776	534	<Null>	<Null>	7	1:2	1	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	534	1	0

Conclusions and Future Work

Thanks to:

- Department of Public Works (DPW), Los Angeles County, USA.
- Institut Cartogràfic i Geològic de Catalunya (ICGC), Barcelona, Spain.
- Kevin Hunt, New York State Department of Transportation, USA.
- Richard Fairhurst, Riverside County Transportation and Land Management) CA, USA RCTLMA,
- 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 pre-processing.**
- **Run automated tools to obtain highly accurate results and evaluation information.**
- **Interactively review and edit the results. The time is worth-investing.**

Consider conflation a higher priority

Study the tools and understand the results

- Start with small test areas

Customize the workflows for your scenarios

- Improve data quality and usability
- Bring new value and potential to your data

Work with broader communities

- Data sharing and collaboration
- Seamless analysis and mapping

Please send us your feedbacks and share your stories … ☺

Future work

New tools and enhancements

- Improve pattern recognition and feature matching
- Refine rubbersheet links
- Enrich system tool results
- New tools to support more use cases and other feature types

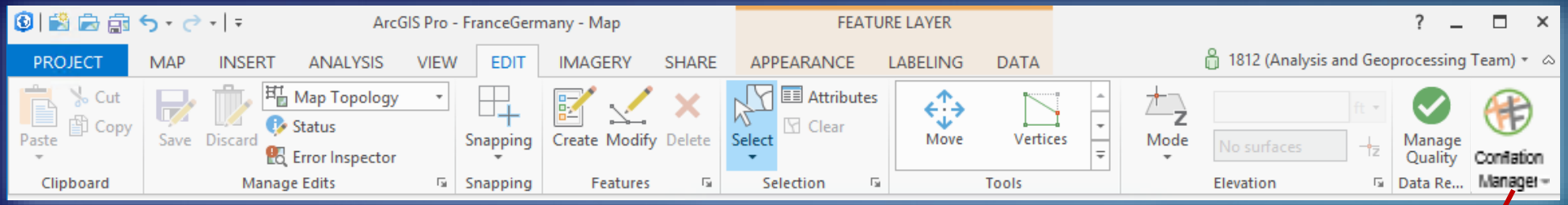
Integrated processing and inspection system

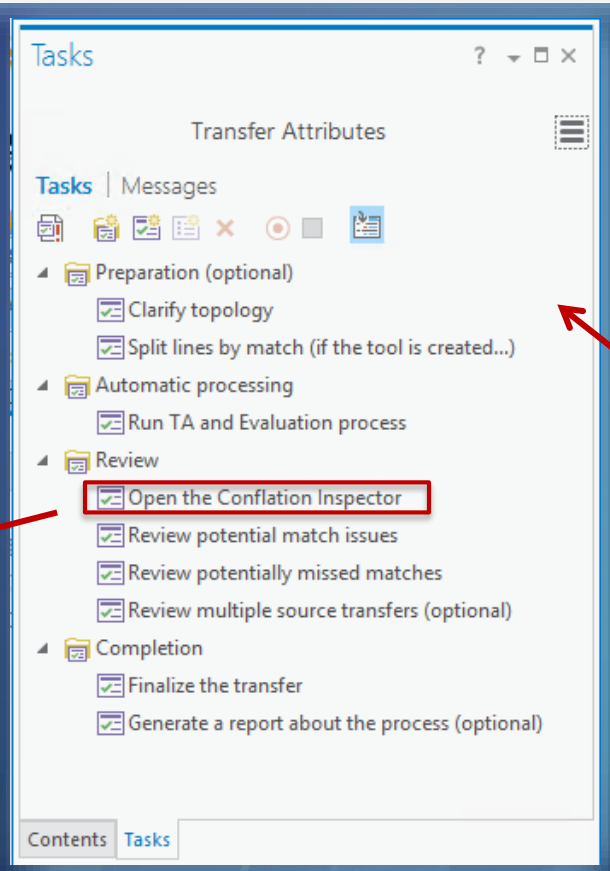
- Design of Conflation Manager is underway

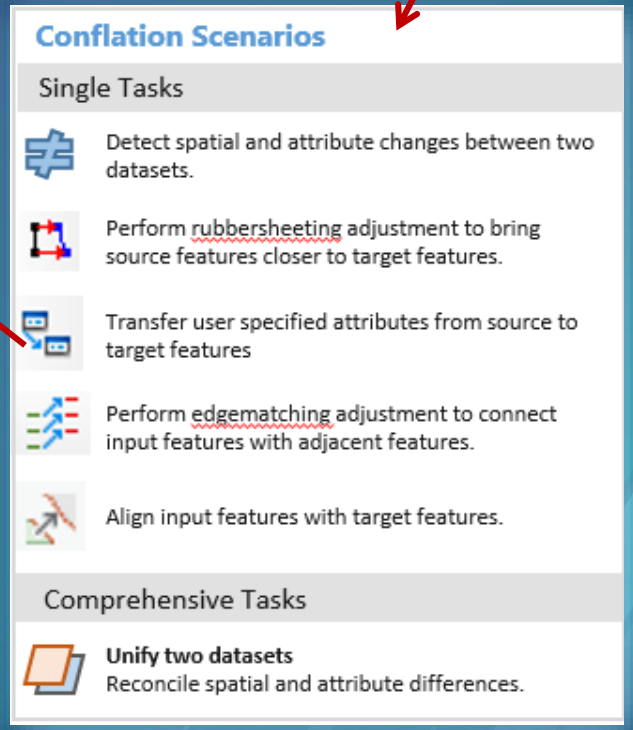
Formalization of workflows

- Common needs (e.g. multi-scale data updating, linking features at different scales)
- Incorporation of other data sources (imagery, lidar, GPS)
- Contextual conflation (spatially related features)

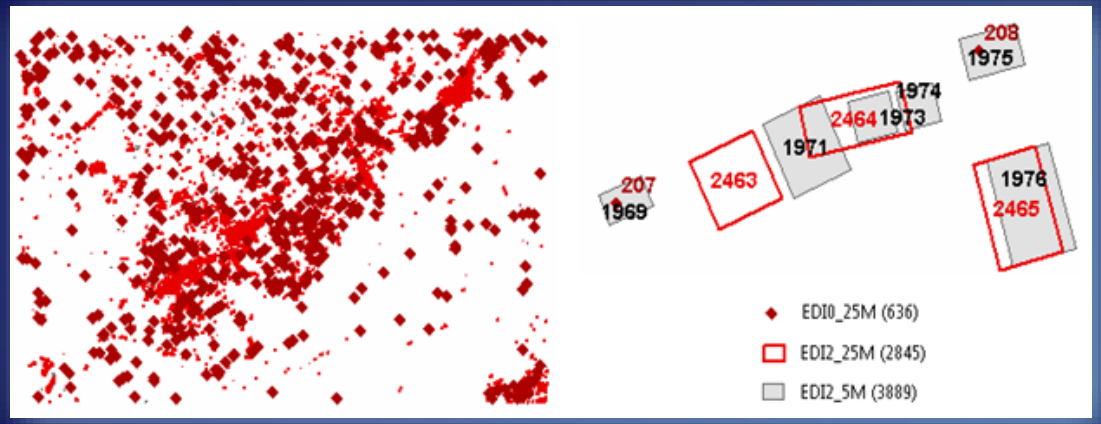
Conflation Manager (ConfMgr)

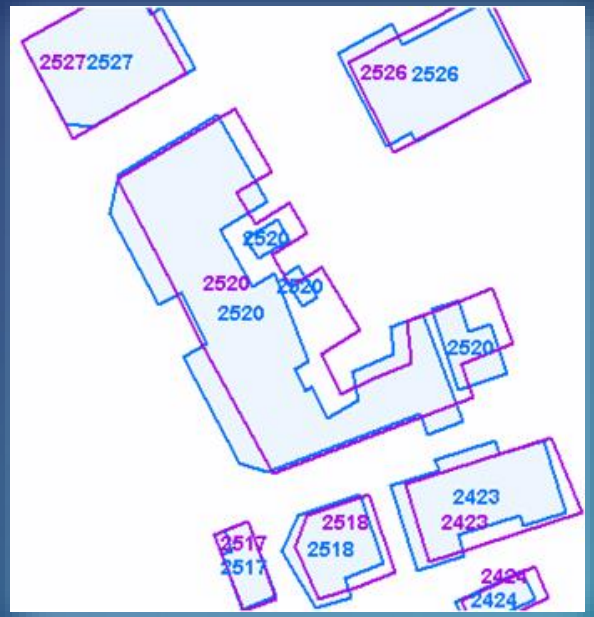


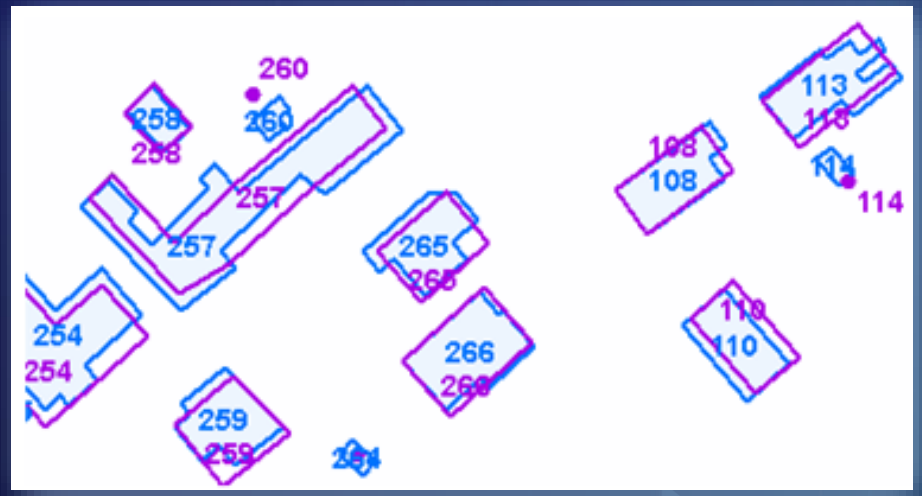


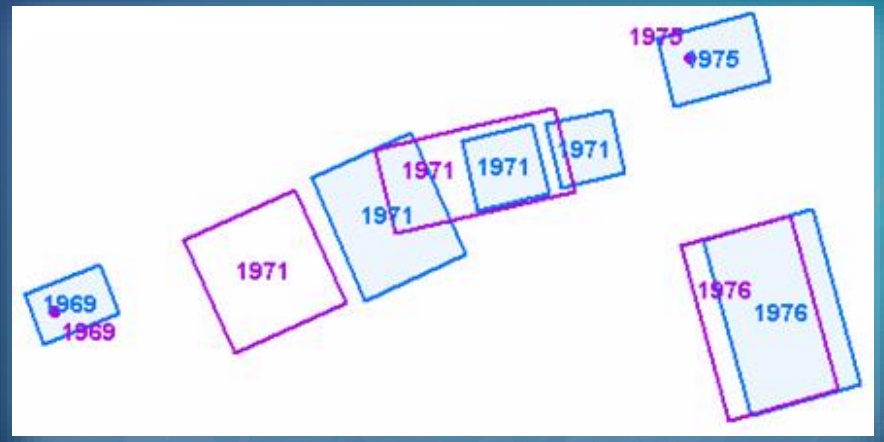


Example of linking features: ICGC 1:5k and 1:25k buildings









Recent papers

- 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. https://kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop2014/genemr2014_submission_8.pdf

- Lee D, Ahmed N, Chowdhury, I (2018), *Incorporating Changes in Multi-scale Databases*, poster presentation, AutoCarto/UCGIS Conference, 2018, Madison, Wisconsin, USA. http://www.ucgis.org/assets/docs/AutoCarto-UCGIS%202018%20Proceedings.pdf (page 95)

- Lee D, Yang W, Ahmed N (2017), *Road data conflation – the key step to geospatial data enhancement*, The 28th International Cartographic Conference, 2017, Washington DC, USA.

- Lee D (2015), *Using Conflation for Keeping Data Harmonized and Up-to-date*, ICA-ISPRS Workshop on Generalisation and Multiple Representation, 2015, Rio de Janeiro, Brazil. https://kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop2015/genemr2015_submission_8.pdf

- Lee D, Yang W, Ahmed N (2015) *Improving Cross-border Data Reliability Through Edgematching*, The 27th International Cartographic Conference, 2015, Rio de Janeiro, Brazil. http://www.icc2015.org/abstract,670.html

- Lee D, Yang W, Ahmed N (2014) Conflation in Geoprocessing Framework - Case Studies, GEOProcessing, 2014, Barcelona, Spain. http://goo.gl/iOoSGV

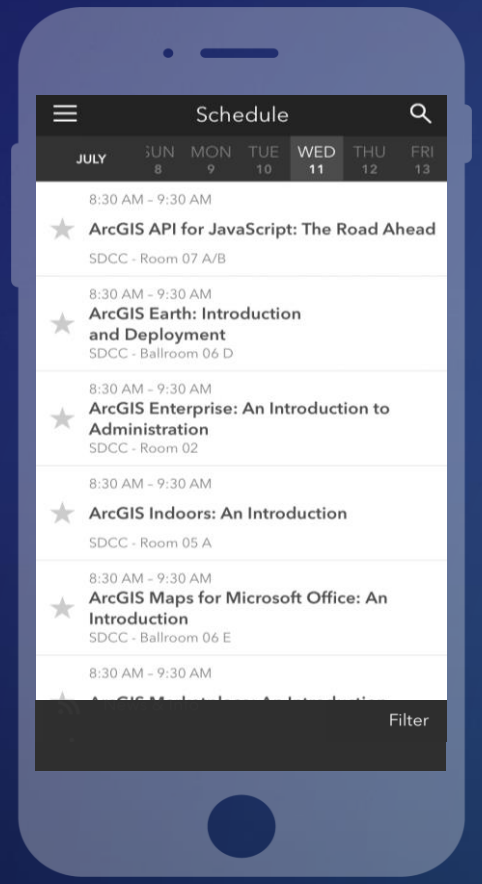
- Yang W, Lee D, and Ahmed N (2014), "Pattern Based Feature Matching for Geospatial Data Conflation", GEOProcessing, 2014, Barcelona, Spain. http://goo.gl/JKGJbo

Please Take Our Survey on the **Esri Events App!**

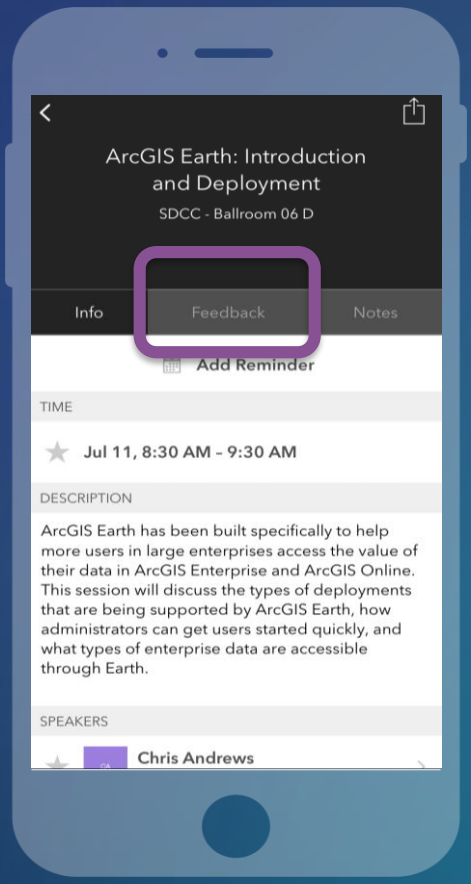
Your feedback allows us to help maintain high standards and to help presenters

Download the Esri Events app and find your event

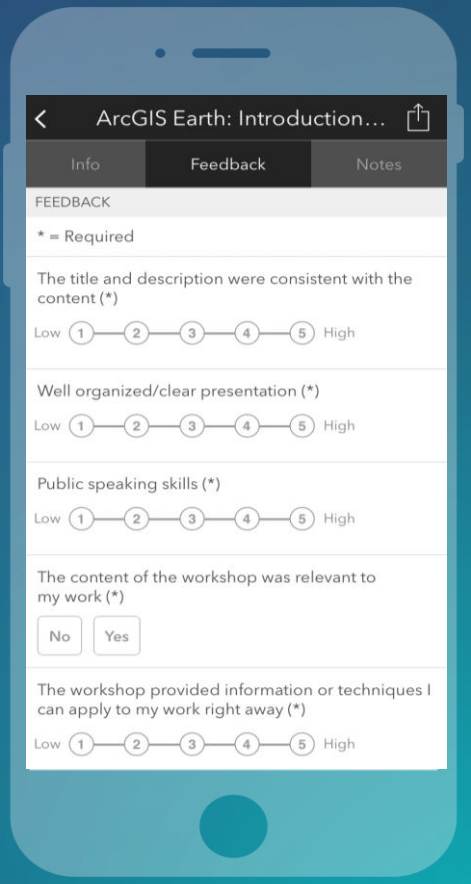
Select the session you attended



Scroll down to find the survey



Complete Answers and Select "Submit"



Thank you for attending!

Questions & Answers

This session will repeat at
2:30pm – 3:30pm, Wednesday
Room 29C

esri | THE SCIENCE OF WHERE