Toronto Police Service

Know Your Neighbourhood (KYN) Application

Implementation Guide: Data Preparation

Produced By:

Dale Langford[a], Eric McNeill[a], Eric Millan[b]

[a] GIS Cartographic Specialist, Fleming College
[b] GIS Applications Specialist, Fleming College

Client Contacts:

Debbie Verduga • Senior Crime Analyst, Analytics and Innovation, Toronto Police Service
Tyler Munn • Senior GIS & Data Integration Specialist, Toronto Police Service
Derek Cooper • Crime Analyst, Analytics and Innovation, Toronto Police Service
Gayathri Ganesan • Crime Analyst, Analytics and Innovation, Toronto Police Service

Project Advisor:

Shawn Morgan • Professor and Program Coordinator (GIS Post-Graduate Programs), Fleming College

MCI Workflow (Point Data)

The purpose of this ETL is to ensure data integrity for the point features used in the KYN application. The aim is that all point features within the neighbourhood boundaries will have the same attributes and that there will not be null values in the dataset. Setting up the ETL is a relatively repetitive process, so this guide will illustrate how to set up the connection for the Assaults Dataset, and then point out the unique situations where each crime category differs. This tool has been built with FME, and users should be familiar with the basic use of this product.
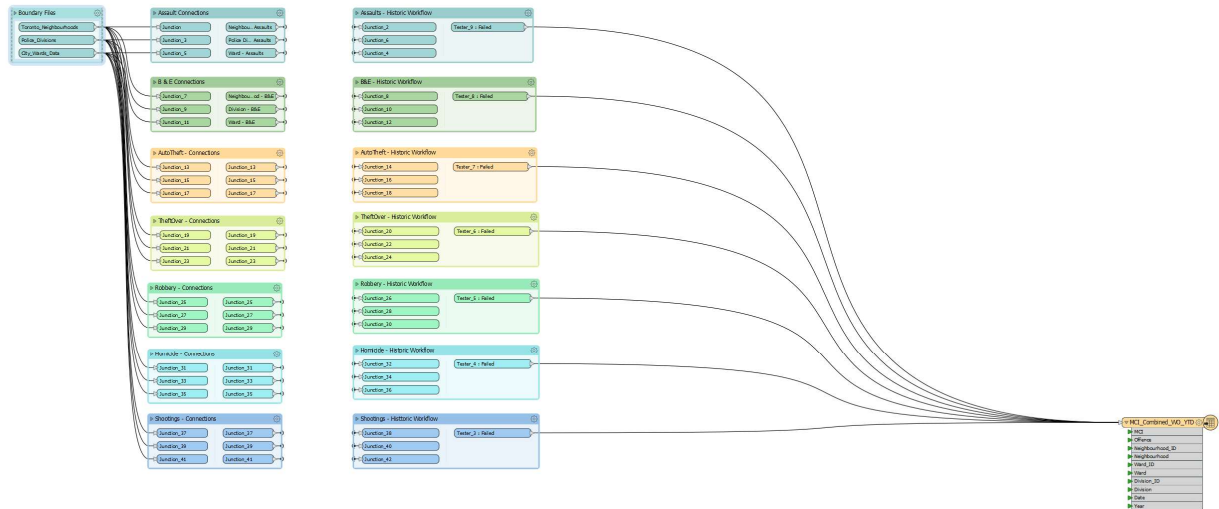
An overview of the ETL process to combine the 7 MCI datasets into a single point layer for the KYN application. Note the use of 'tunnels', separating the neighbourhood polygons and the point features to avoid cluttering the workspace.
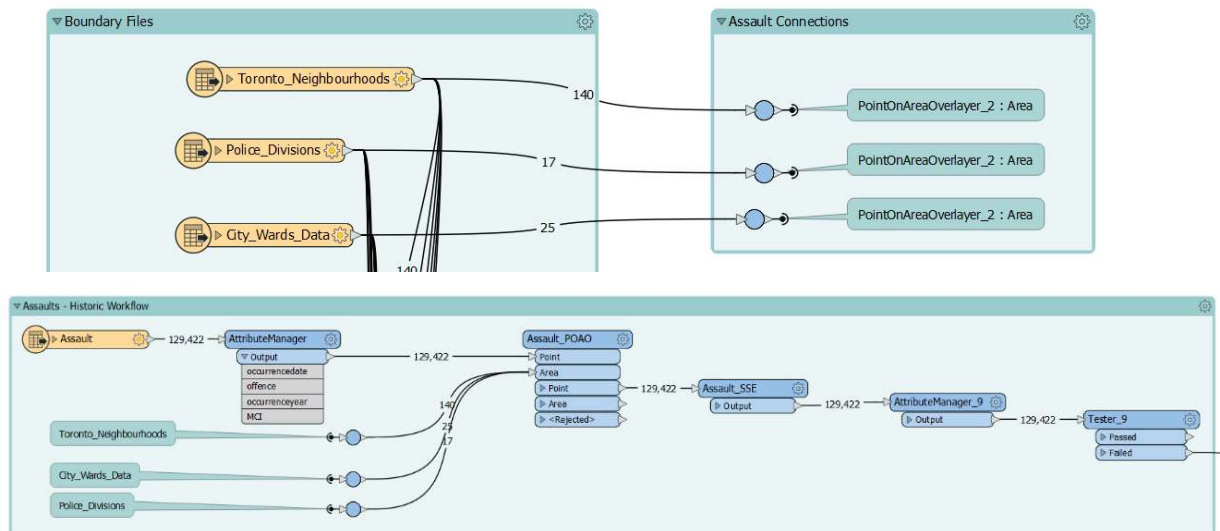

An overview of the ETL Process for a single point layer (assaults). Note the use of the tunnels to connect the boundary layers and the point feature. Using Bookmarks (the blue box surrounding all transformers) and tunnels will help organize your workflow).

1) **Create Reader Connection to ESRI ArcGIS Server Feature Server:** For this process, you will need the URL of the rest service for the feature you are connecting to. This can be acquired by navigating to the layer in your organization's ArcGIS Online Item Information page. Since this feature's sharing level are set to public, we do not require an Authentication Type (found within the Feature Service Parameters Window).

   Rest Service URLS:

   https://services.arcgis.com/S9th0jAJ7bqgIRjw/arcgis/rest/services/Assault/FeatureServer
   https://services.arcgis.com/S9th0jAJ7bqgIRjw/ArcGIS/rest/services/Break_and_Enter/FeatureServer
   https://services.arcgis.com/S9th0jAJ7bqgIRjw/ArcGIS/rest/services/Auto_Theft/FeatureServer
   https://services.arcgis.com/S9th0jAJ7bqgIRjw/ArcGIS/rest/services/Theft_Over/FeatureServer
   https://services.arcgis.com/S9th0jAJ7bqgIRjw/ArcGIS/rest/services/Robbery/FeatureServer
   https://services.arcgis.com/S9th0jAJ7bqgIRjw/arcgis/rest/services/Homicide_ASR_RC_TBL_002/FeatureServer
   https://services.arcgis.com/S9th0jAJ7bqgIRjw/ArcGIS/rest/services/Shootings_and_Firearm_Discharges/FeatureServer
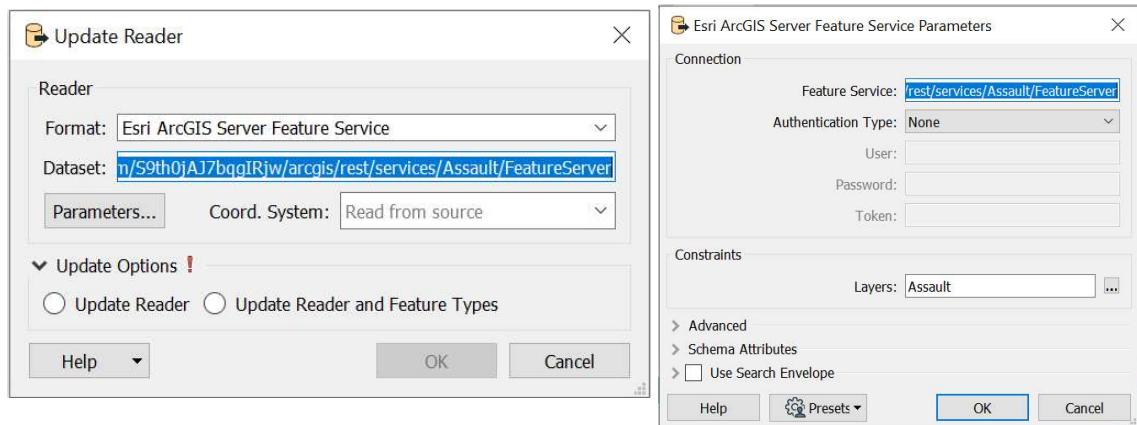
2) **Add Date Filter to Reader:** There are two different types of date attributes available in most of the datasets (occurrence date and reporting date). For the purposes of this application, only the occurrence date is relevant, as we are attempting to illustrate crime trends at the neighbourhood level. Historic crimes, while important to track, are not within the scope of this project. For this reason, navigate to the reader parameters and add the WHERE clause ["occurenceyear" >= 2014]. This will filter out all historic crimes from the dataset.
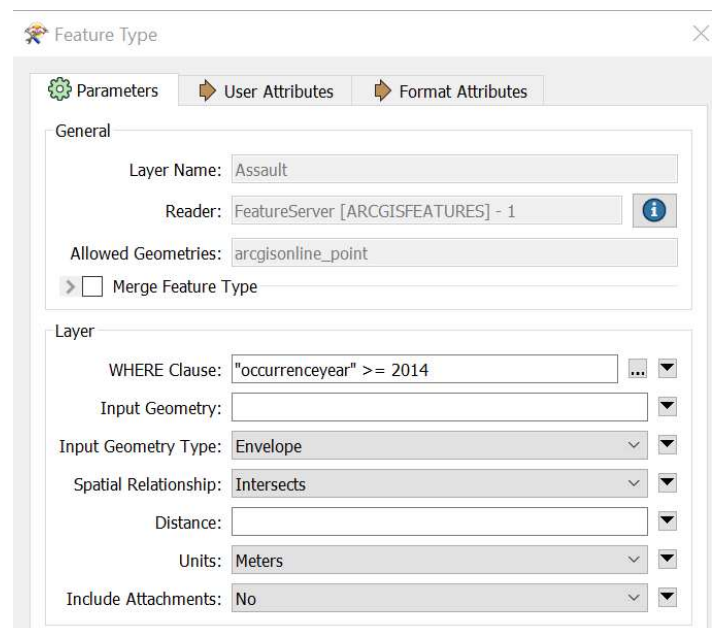
3) **Create Reader Connections to Neighbourhood, Ward and Police Division boundaries:** Since an authoritative REST service could not be found for the Polygon layers used in this project, shapefiles of the boundaries were downloaded from the City of Toronto data portal. These were then stored in a file geodatabase as polygon features.

Toronto Open Data URLS:

https://open.toronto.ca/dataset/neighbourhoods/
https://open.toronto.ca/dataset/city-wards/
https://open.toronto.ca/dataset/police-boundaries/

Once downloaded and stored in a file geodatabase, create a reader connection. Be sure to navigate to the parameters screen, click the ellipses next to the Tables box and select all. After this is completed, both sets of readers should be established.
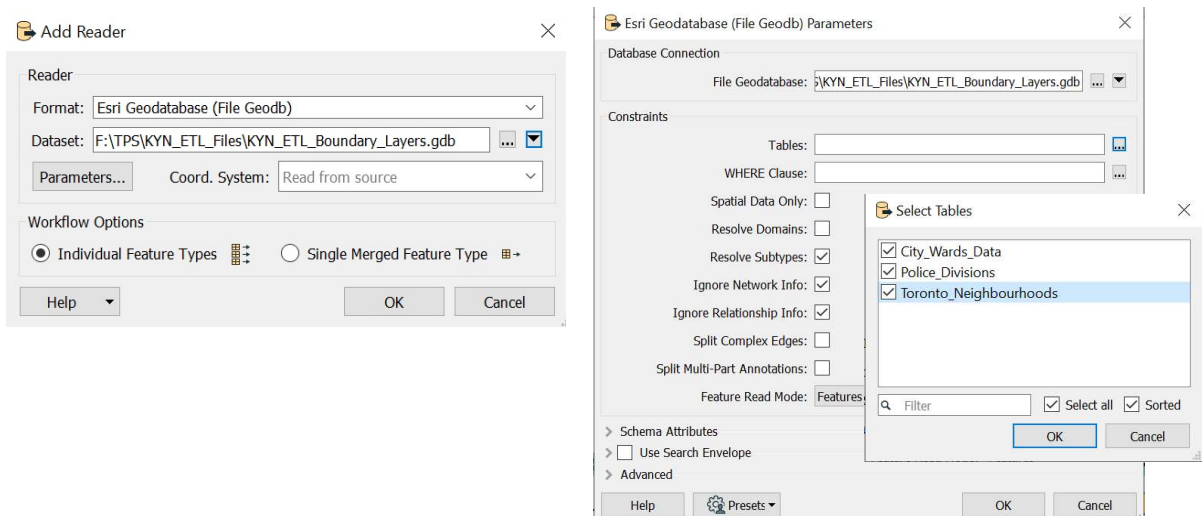
4) **Create Writer Connected to ArcGIS Online Hosted Feature:** Create a new writer and set the format to "Esri ArcGIS Online (AGOL) Feature Service. Set the Coordinate system of EPSG:4326 (The code for WGS84) and set the Layer Definition to Manual. This will be filled out in the next step. Click on parameters and select your organization account. You may be asked to enter your ESRI credentials at this stage. Add a name for your hosted feature layer.
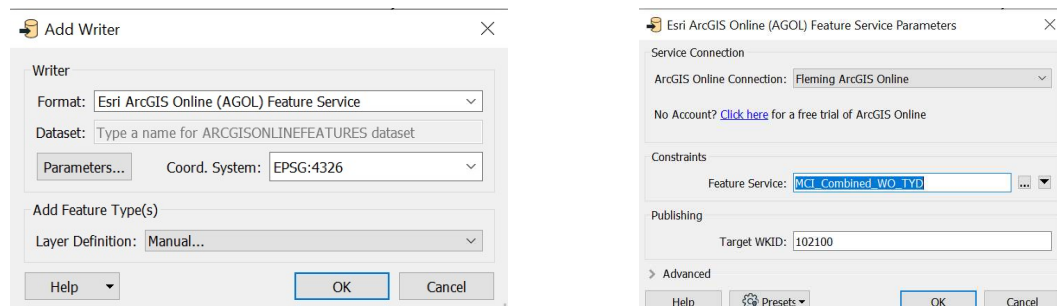
5) **Configure Writer:** To complete the creation of the writer, select "argisonline_point" as the Geometry, Set Writer Mode to "INSERT", Change Truncate First to "Yes". The click on "User Attributes" and manually fill in the fields to match the image below. Pay particular attention to the data types. You may add Widths if desired.
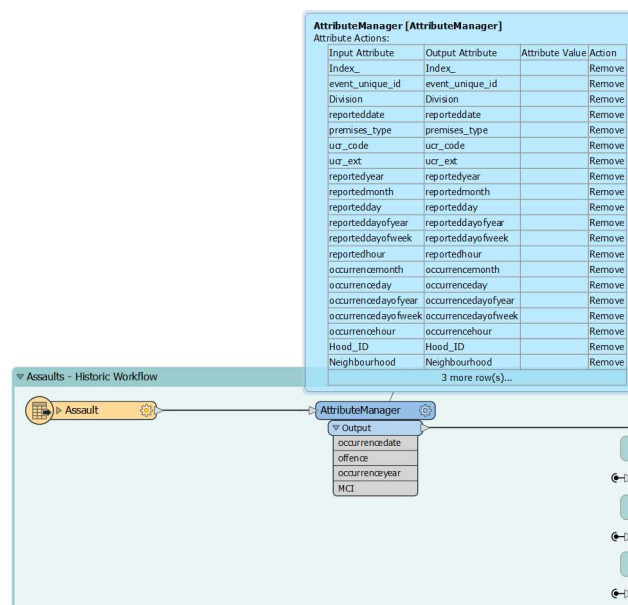

6) **Remove Unnecessary Attributes from Point Data:** To maintain a clean workspace, use the attribute manager to remove fields which will not be useful in the final dataset. Leave the occurencedate, offence, occurrence year and MCI categories in place.

7) **Add the PointOnArea Overlayer Transformer:** This tool performs a spatial join between the 3 polygon layers and the point data. Connect the point data and all three polygon layers. Be sure to navigate to the tool parameters and select "Merge Attributes". This will join the attributes of the polygon areas to the point features. Once you have connected the polygon layers, you can right click on the black line and select "Create Tunnel". This will assist in maintaining a clean workspace.
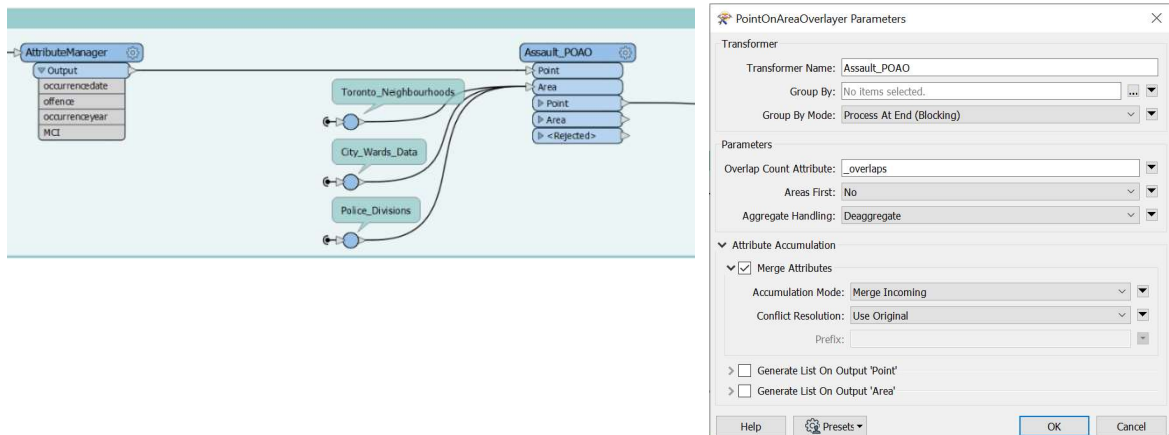

8) **Add the Substring Extractor Transformer:** The attribute for Area_Name, which contains the name of the neighbourhood from the polygon layer contains a suffix including the neighbourhood ID number (For example, "The Annex (95)"). The suffix will need to be parsed and to do this we use the substring extractor transformer. Be sure to select the "AREA_NAME" attribute and set the parameters to 0 and -6.
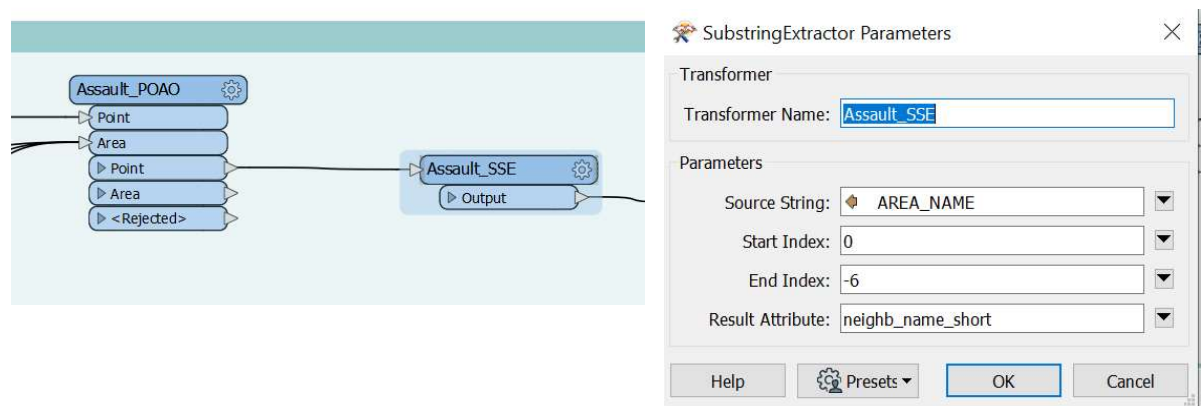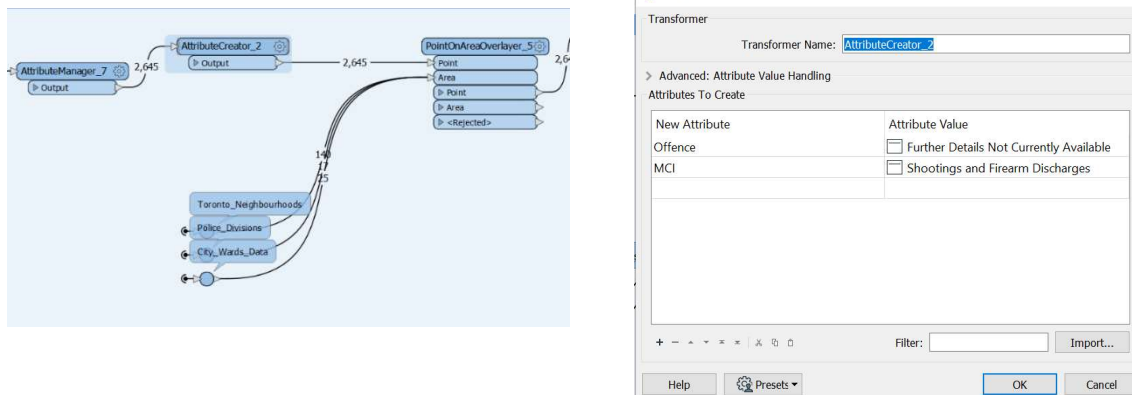
9) **Add Attribute Manager:** At this point, most of the Transformations necessary for the ETL have been completed. In order to maintain a clean workspace and make troubleshooting easier, add another Attribute Manager and delete superfluous fields. Also, rename the attributes which will be used in the final dataset to match the names in the writer attribute parameters.

10) **Add Null Value Tester:** Within the dataset are a large number of points which fall outside the boundary of Toronto. Since they are not within the neighbourhood boundaries, they are outside the scope of this project and need to be removed. Add test clauses for the Neighbourhood, Division and Ward Fields, set the Logic to "OR" and the Operator to "Attribute Is Empty String". Pass the features which fail the test to the writer for your final Hosted Feature Layer.
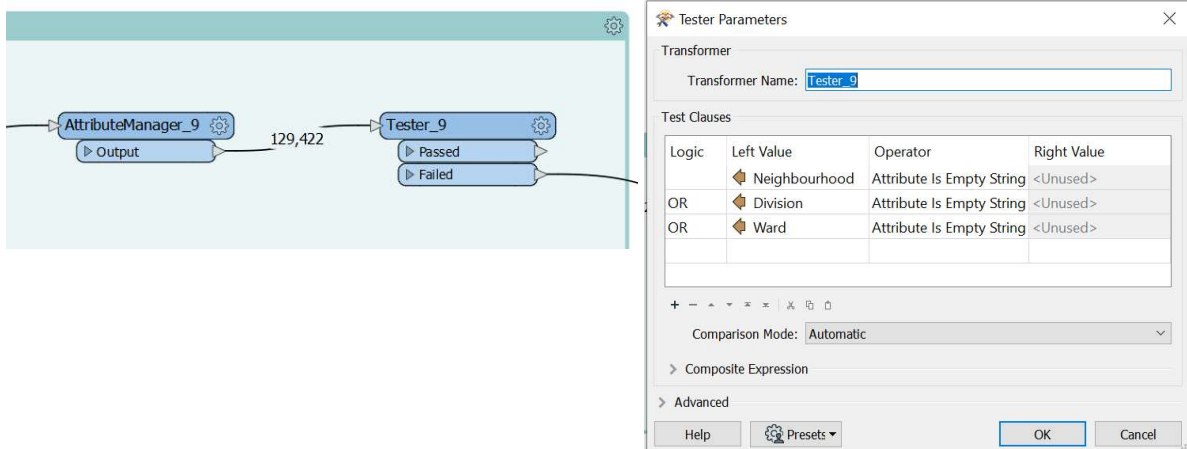
11) **Connect Process to Writer:** Because the field names should have been correctly modified during step 9, you should be able to simply connect the process to the writer. Expand the writer, and check to see if all the arrows are green. If an arrow is red or yellow, manually connect the appropriate field from the process.
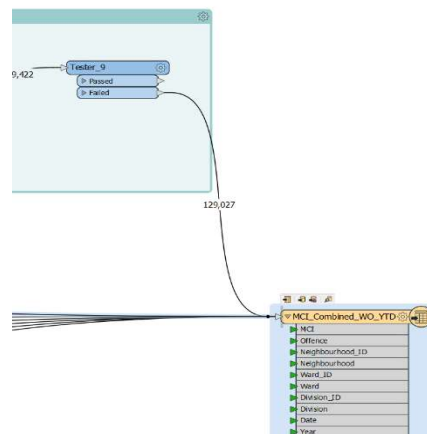
12) **Repeat for all other Data Sources:** Repeat steps 1-3 and 6-11 for the 6 remaining feature Readers. Note that for the Homicide and Shootings Layers, an additional Step is required. This process is also generally replicated for the Bicycle Theft and Auto Collision Point Layers.

13) **Add Missing Attributes:** For the shootings and Homicide layer, the data attributes are not the same as the other 5 MCI categories. There is no "offence" or "MCI" field and this data needs to be generated in order to fit the final model. Add the Attribute Creator tool and connect to the transformer in between the first attribute manager and the Point on Area Overlayer. For the Shootings Layer, configure the fields to match the image below.

Add a similar transformer in the Homicide ETL process and have it match the image below:

14) **Run the Process:** Run the process from the start in order to create the new hosted feature class in your organization. After receiving a notification in the translation log, verify the successful creation of the layer my navigating to your AGOL content page.

15) **Create View:** To use this layer in a web app, it is best to create a "View Layer". This gives you the ability to truncate the data (for example, if YTD data was incorporated in the future. Click on the Newly created layer to navigate to the item information page, click on "Create View Layer" and fill out the requested information.

Concluding Notes:

Following this process, you should have successfully created the ETL process and loaded the MCI point data for use in the Know Your Neighbourhood Application. With the Truncate Date option turned on for the writer, this application can be run frequently if new data is added to the primary datasets. This would be appropriate if YTD data were incorporated in the future versions.

To Automate this task, in the event that YTD is used, the file "Point_Feature_ETL.fmw" could be scheduled using the windows task scheduler. Alternatively, FME is available as a cloud service and this process could be run remotely independent of local architecture. Ensure that truncate first is selected in the writer parameters.

Neighbourhood, Division and Ward ETL (Polygon Data)

The other major datasets involved in this project are the neighbourhood, Division and ward polygon layers. These boundaries delineate how the crime data will be group and help to build the foundation of all analysis and cartography for the web application. As noted previously, the Neighbourhood layers are set to be modified at some point in 2021, and so this process was designed with that in mind. Overview of Polygon Layer Processing.
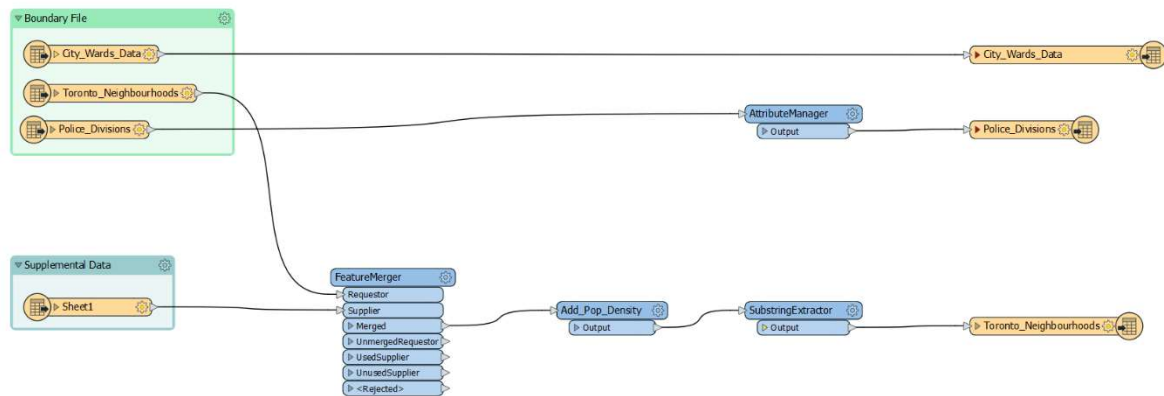
An overview of the ETL for the Polygon layers.

1) **Download Data:** This data was not readily available from a REST service, so it was important to first download the necessary data as shapefiles, open in ArcGIS Pro/Desktop and convert into a file geodatabase. The data used in this portion of the project is entirely from the City of Toronto's open data portal.

   Additionally, we wanted to include the option of showing basic demographic data on the neighbourhoods and so the Neighbourhood Profile layer was also downloaded as an excel spreadsheet. This will be joined with the Neighbourhood layer in FME. The columns and rows are oriented incorrectly, so highlight all the data, copy and open a new sheet. In the new sheet, right click and paste with the "transpose option". The data should be oriented with the Neighbourhood names on the left-most column.

   Toronto Open Data URLS:

   https://open.toronto.ca/dataset/neighbourhoods/
   https://open.toronto.ca/dataset/city-wards/
   https://open.toronto.ca/dataset/police-boundaries/
   https://open.toronto.ca/dataset/neighbourhood-profiles/

   Please note that all data which needs to be downloaded has been packaged and supplied in the folder "KYN_ETL_Files" under the name "Neighbourhood_Info.xlsx".

	A	B	C	P
1	Characteristic	Neighbourhood Numb	TSNS2020 Designation	
2	City of Toronto			
3	Flemingdon Park	44	NIA	
4	Downsview-Roding-CFB	26	NIA	
5	Clairlea-Birchmount	120	No Designation	
6	Kingsview Village-The Westway	6	NIA	
7	Malvern	132	Emerging Neighbourhood	
8	Mimico (includes Humber Bay Shores)	17	No Designation	
9	Roncesvalles	86	No Designation	
10	York University Heights	27	NIA	
11	Dovercourt-Wallace Emerson-Junction	93	No Designation	
12	Annex	95	No Designation	
13	Church-Yonge Corridor	75	No Designation	
14	Eringate-Centennial-West Deane	11	No Designation	
15	Islington-City Centre West	14	No Designation	
16	South Parkdale	85	NIA	
17	Etobicoke West Mall	13	No Designation	
18	Greenwood-Coxwell	65	No Designation	
19	Agincourt South-Malvern West	128	No Designation	
20	Banbury-Don Mills	42	No Designation	
21	Blake-Jones	69	No Designation	
22	Broadview North	57	No Designation	
23	Don Valley Village	47	No Designation	
24	East End-Danforth	62	No Designation	
25	Glenfield-Jane Heights	25	NIA	
26	High Park North	88	No Designation	
27	Lambton Baby Point	114	No Designation	
28	Mount Pleasant East	99	No Designation	
29	North St. James Town	74	No Designation	
30	Oakwood Village	107	No Designation	
31	Parkwoods-Donalda	45	No Designation	
32	Regent Park	72	NIA	
33	Rosedale-Moore Park	98	No Designation	
34	St.Andrew-Windfields	40	No Designation	
35	Waterfront Communities-The Island	77	No Designation	
36	West Hill	136	NIA	

Excel Spreadsheet has Neighbourhood Names and Numbers. Note the existence of the City of Toronto Averages included as well. This would be a useful dataset for future analysis

2) **Create Readers:** Readers will be needed for the Neighbourhood, Ward and Police Division Files as well as the Excel File containing neighborhood info. When connecting to the Esri Geodatabase, be sure to import all 3 required tables.
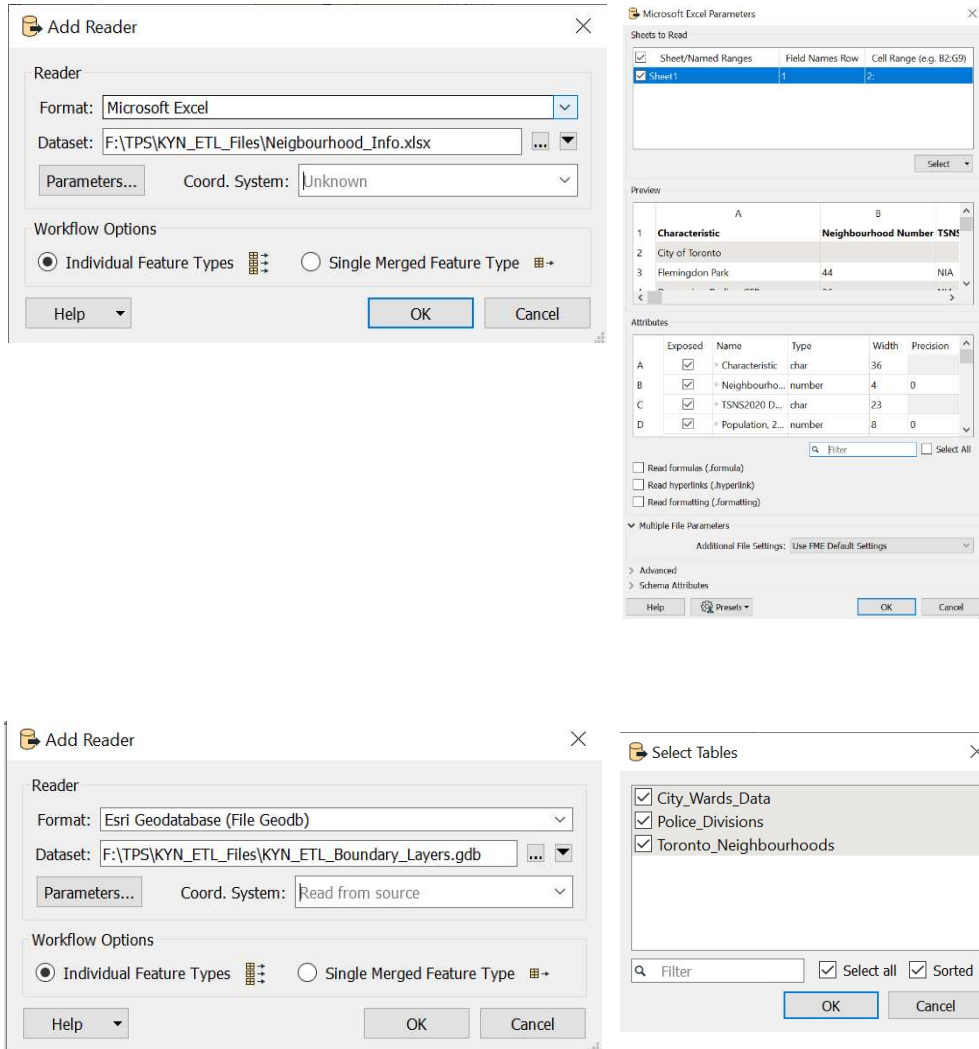
3) **Configure Feature Merger Transformer:** In order to join the data between the Toronto Neighbourhoods Feature class and the Excel spreadsheet, use the "FeatureMerger" Tool to create an attribute join. The common field between the two datasets are the "AREA_SHORT_CODE" and "Neighbourhood Number". Creating this join provides you with access to a vast amount of information regarding each neighbourhood, including data on race, income and age.
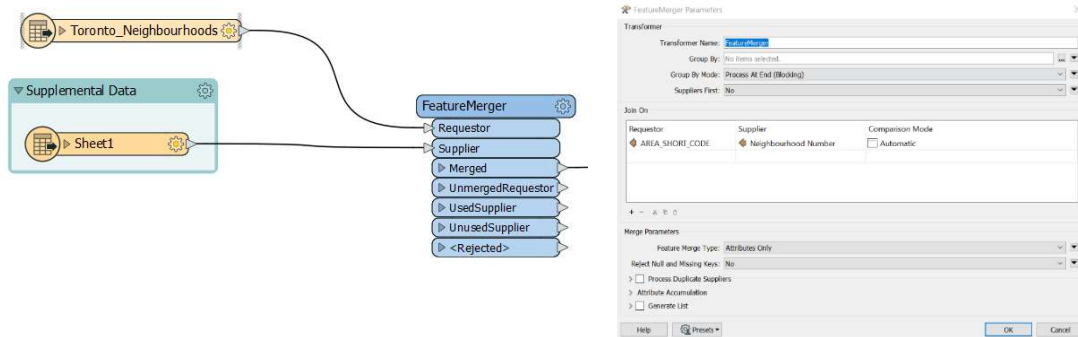

Knowing which data to include and omit can be an extremely important decision. We decided to only include basic neighbourhood population for this project to avoid any controversy or unintended perceptions of our work. Data on race, income and languages, while available, were omitted due to the limited time frame and complexity of the social context of including that information.

4) **Add Expression Evaluator:** In this stage of the data preparation, expression evaluators can also be used to perform arithmetic functions to suit the needs of the project. At an earlier stage of development, we felt that population density would be a useful dataset to include, and this tool was used to calculate this based on the population in the excel data as well as the geometry of the feature class.
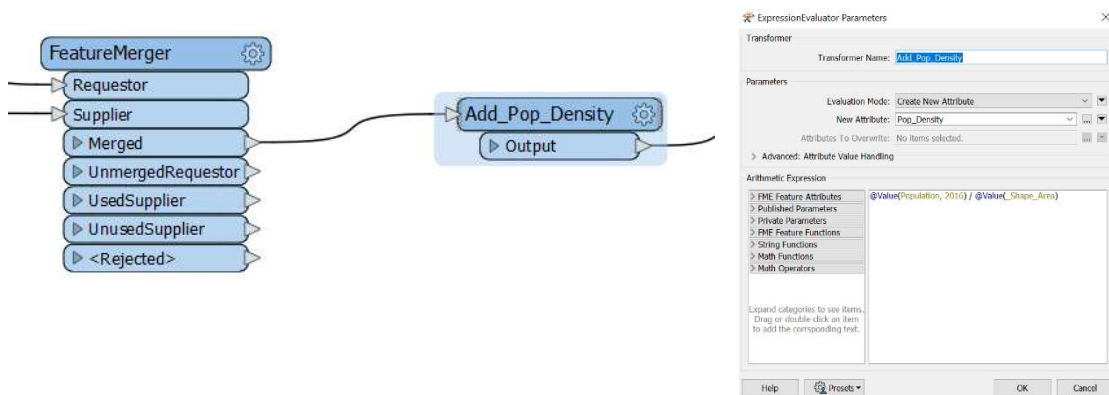

The Expression Evaluator is a powerful tool, allowing for new fields to be created within FEM and published as part of your dataset.

5) **Add Substring Extractor:** The neighbourhood names in the "Toronto_Neighbourhoods" shapefile has unnecessary information included as a suffix. This needs to be trimmed before being including in the final published dataset. To do this, add the Substring Extractor transformer and set the parameters to 0 and -6.
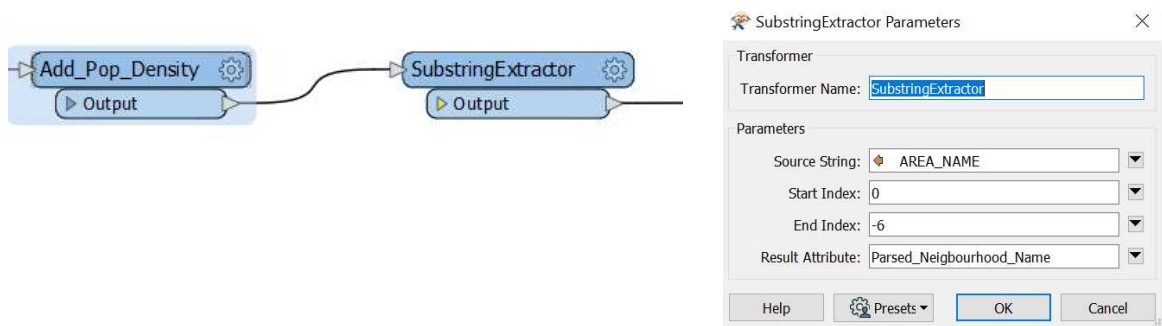

6) **Create Writer:** The writer will publish the dataset as a hosted feature layer in your organization's ArcGIS Online account. You will need your ESRI credentials in order to add your connection. Enter a new name for this layer in the "Feature Service" Parameter box.
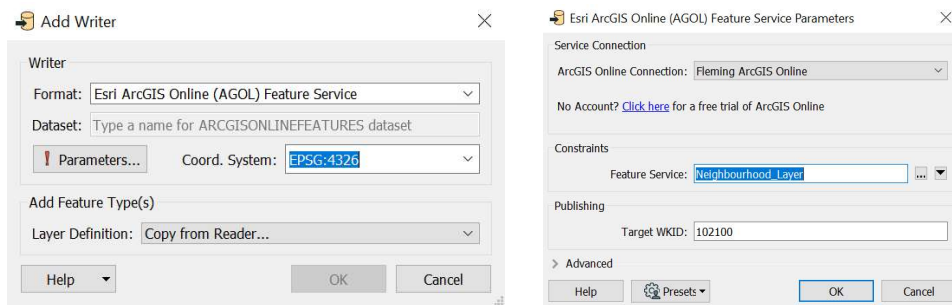
Be sure to set the writer parameters to "Insert" and select "Yes" on the truncate option. This will overwrite the data sets each time the tool is run.
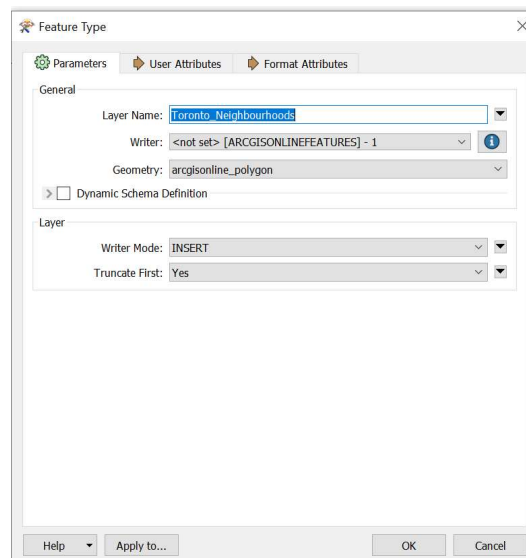
7) **Run FME Process:** Run this process and confirm the successful creation of your hosted feature layer in ArcGIS Online.

8) **Create View Layer:** In order to create web maps efficiently, create a new view layer in ArcGIS Online and give it an appropriate name. By doing this step, you are able to create web apps, and have the ability to overwrite datasets without losing the symbology you have worked on. Note that one thing you cannot do after the fact is add additional attributes to the dataset.
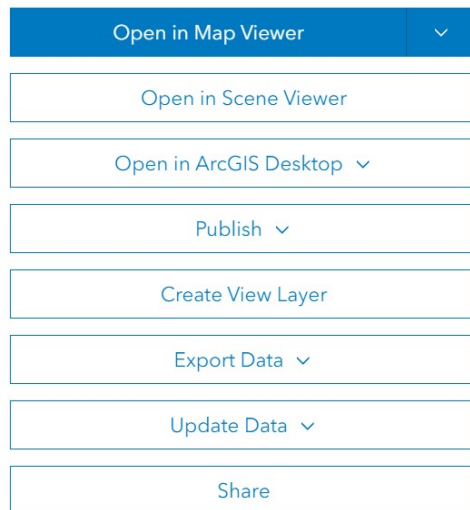
9) **Repeat with Police Division and Ward Datasets:** With the focus of this project on crime at the neighbourhood level, we did not seek out additional datasets to combine with Wards or Police Divisions. While there are lots of opportunities here, simple name and ID attributes were all that were included. Create the additional readers, transformers and writers and ensure that within ArcGIS Online, views for each dataset are also created.