

Project 1 - Bulk FME Workspace Examiner Tool

Project Description:

Design an FME Workspace that can holistically examine other FME workspaces in bulk and provide enhanced reporting & searching capabilities for long-term maintenance and upkeep.

Project Date

November/December 2021

Included Files:

Workspaces

File Name	Description
WorkspaceExaminer.fmw	Workspace Examiner Tool

HTML Reports

File Name	Description
BulkWorkspaceReport.html	Bulk report containing all individual reports. Always Generated.
SummaryStatistics.html	Workspace Summary statistics containing 1-3 reports depending on user choice
LinkedVsEmbeddedTransformers.html	Report detailing all custom transformers and whether they are linked or embedded
WorkspaceHistory.html	Report on workspaces containing no history
WorkspaceDescriptions.html	Two sections: 1. Report on workspaces containing no description/overview, 2. a table listing all the workspaces and their associated descriptions
UpgradeableTransformers.html	Report containing # of upgradeable transformers per workspace per-directory including embedded and linked transformers
InputOutputPaths.html	Report detailing input/output path types including local disk path, unc paths, url, and FME parameters. Optional to check existence of file
InputOutputFeatureTypes.html	Report allowing user to see where specific feature types (feature classes, gdb tables, excel files, etc) are used across various workspaces. Wildcard search ability allows searching for specific features

Detailed Description:

While FME is not a programming language in the traditional sense it is a form of coding in its own right and as such should follow some industry-standard best practices including participating in Code Review sessions by other authors. In my FME Journey of managing, authoring, and repairing workspaces in a production environment for a state-level DOT, I encountered the need to automate many of the FME code review steps I've established including:

1. Generating workspace statistics for both individual workspaces and overall for our project directory
2. Identifying workspaces with missing overviews/histories
3. Identifying upgradeable transformers
4. Listing all custom transformers and their current status, Linked Vs Embedded
5. Identifying NON-UNC paths used in production workflows, and checking if input/output files still exist in their intended places
6. Searching for specific input or output features. In a scenario with 400 workspaces in a production environment touching 5 different department databases, how do you identify if any workspace is currently using a feature class someone wants to remove?

Source Data:

Your own!! Test it out and have fun. Point it to a test directory containing one or more FMW/FMX files in the root folder (and subfolders if recursion is enabled) and follow the parameter and help prompts!

Proposed Solution:

Create a workspace capable of scanning one or more workspaces contained in one or more folders, generating user-specified reports in HTML format for easy visualization and distribution among team members. The HTML reports will contain common workspace issues and summary statistics allowing for a faster and more fluid code review process.

The enhanced searching capability to identify input/output features, upgradeable transformers, and linked vs embedded transformers will also assist in the long-term maintenance of all projects.

FME Solution - A Single Workspace to Rule Them All:

Input Parameters

1. Set Input Parameters
 - a. Folder Containing Workspace(s)
 - b. Recursive toggle, yes or no
 - c. Select folder to write reports, default to creating a folder called "WorkspaceReports" in the directory the main workspace is being called from
 - d. Option to output individual reports along with the bulk report
 - e. Regex filter to discard directories, multiple entries separated by pipe (| = regex "OR")
 - f. Regex filter to find specific directories/workspaces, multiple entries separated by pipe (| = regex "OR")
2. Generate Workspace Statistics
 - a. Choose Report Type, choose one or more option
 - b. Genericize transformer counts or provide detail reporting?
 - i. If this is checked, only "Transformers, Reader, Writers" will be output, unchecked will produce a lengthy report with specific transformer names
3. Custom Transformers
 - a. Outputs table of custom transformers used in workspaces and whether they are linked or embedded
4. Find Input/Output Features
 - a. Option to search in inputs
 - b. Option to search in outputs
 - c. Wildcard search string to limit results. IF you wanted to find every process using a database table called "Counties" you could put the string Counties into the search box
5. Identify upgradeable transformers
 - a. Produces count of upgradeable transformers by workspaces
6. Workspace History / Description
 - a. Option to generate report for workspaces missing history
 - b. Option to generate report for workspaces missing description
7. Input/Output Paths
 - a. Option to generate report
 - b. Option to report filepaths that include an FME Parameter
 - c. Option to check if input file path exists
 - d. Option to check if output file path exists

Translation Parameter Values

1. Reporting Parameters

Select a Workspace Folder: ...

Include workspaces in sub folders? ☒ Yes

Select a folder to write HTML reports: ...

Individual reports? ☒

Enter directories to ignore separate by pipe character (|) *:

Enter specific directory/ workspace to target **:

* Regex Clause - Leave blank to not filter out any directories, use | when adding multiple filters
 ** To read everything, leave filter blank.
 Ex: To find all workspaces containing string "GDM" replace .* with GDM
 EX2: to read a single workspace, supply the name of it such as Example1.fmw

2. Generate Workspace Statistics

Select Report Type(s) *: ...

Generalize Transformers to Reader/Writer/Transformer? **: ☒

* One of more options can be chosen and will be generated independently of each other
 ** If this option is not checked, Specific FME Transformer names will be used. This could lead to a very LARGE report.

3. Custom Transformers

List Custom Transformers? ☒

4. Find Input/Output Features?

Search for Feature Type Inputs? ☒

Search for Feature Type Outputs? ☒

Enter String to search (optional) *:

* supplying a string to search will limit the output to only matching features. Blank will pass all features through
 Ex: If you were looking for all workspaces using a specific Database feature class, type that in here such as 'dbo.counties'
 - Multiple features can be searched for together by separating with the pipe character |

5. Upgradeable Transformers

Identify Upgradeable Transformers? ☒

6. Workspace History / Description

Generate Workspace History *: ☒

Generate Workspace Description **: ☒

* History will only report on workspaces missing history
 ** Description will list workspace descriptions (if they have one) and report on workspaces missing description

7. Input/Output Paths

Report Path Types? ☒

Include Paths using FME Parameters? ☐

Check Existence of Input Datasets *: ☒

Check Existence of Output Datasets *: ☐

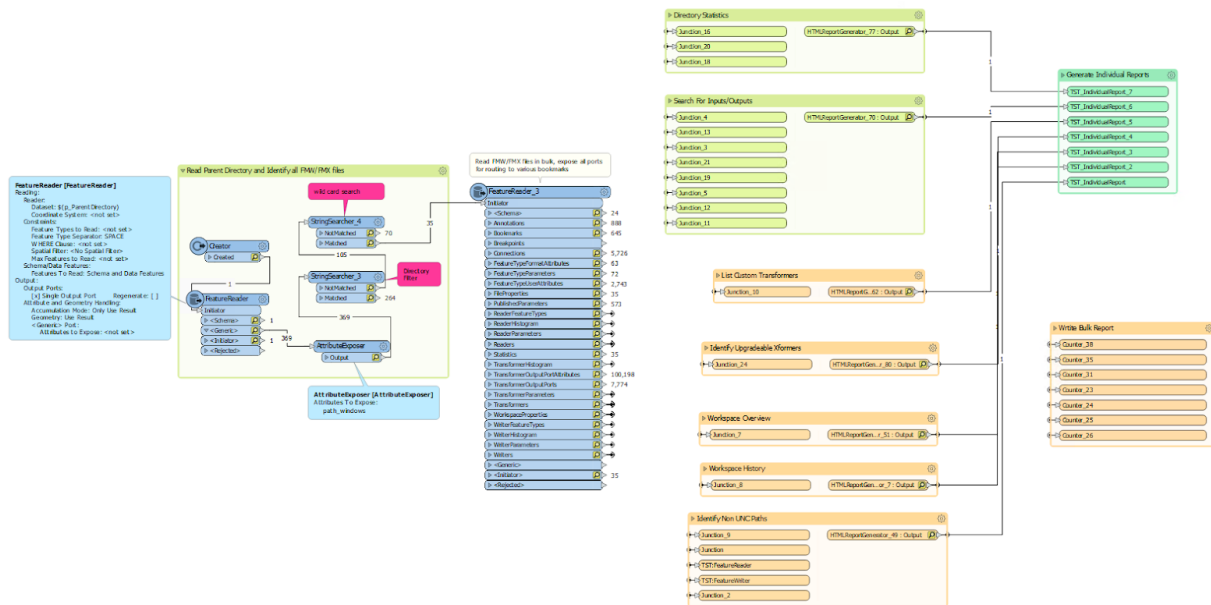
* FME Parameters and URL's will never be checked for existence

Options Presets Run Cancel

Main Overview

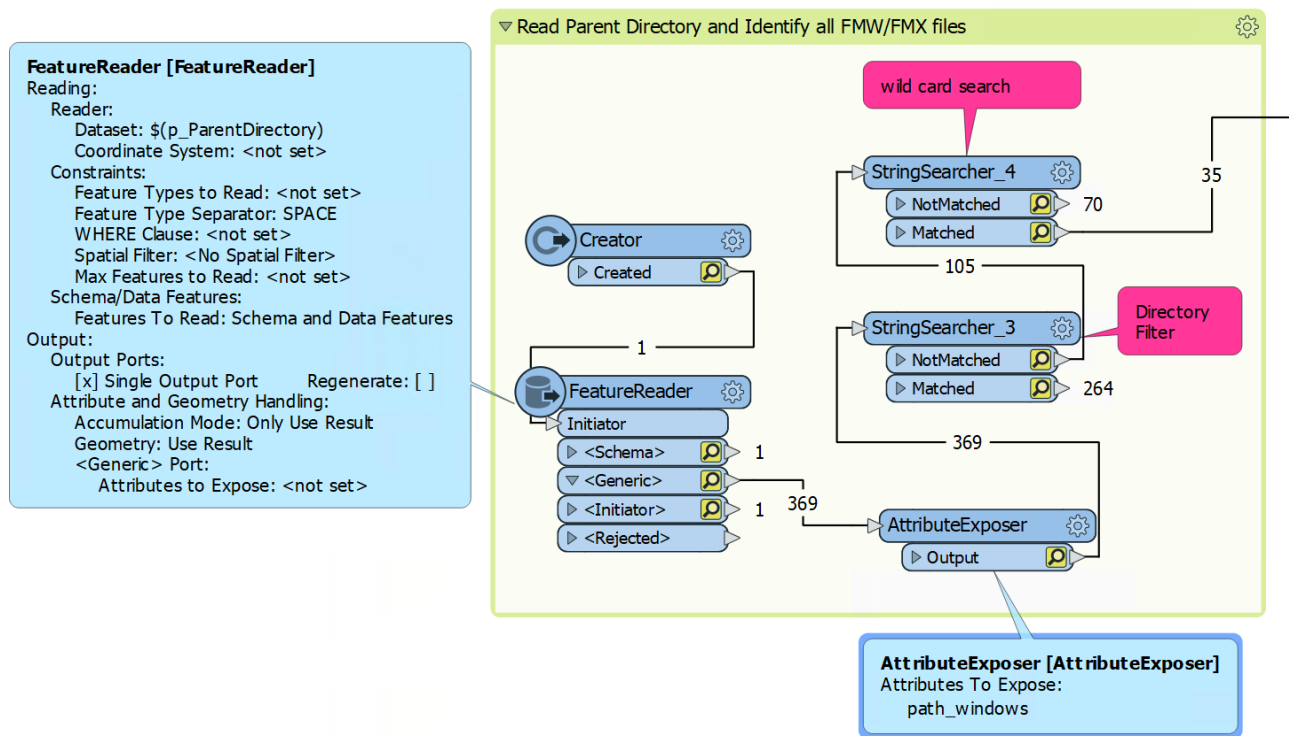
In the image below we can see the main workspace. It is broken up into 4 logical groups:

1. Identify all FMW/FMX files in the target directory and apply designated filters
2. Read remaining FMW/FMX files in bulk, exposing all attributes
3. Process features through user selected reports
4. Write results, bulk and/or individual.



1) Identifying FMW/FMX files

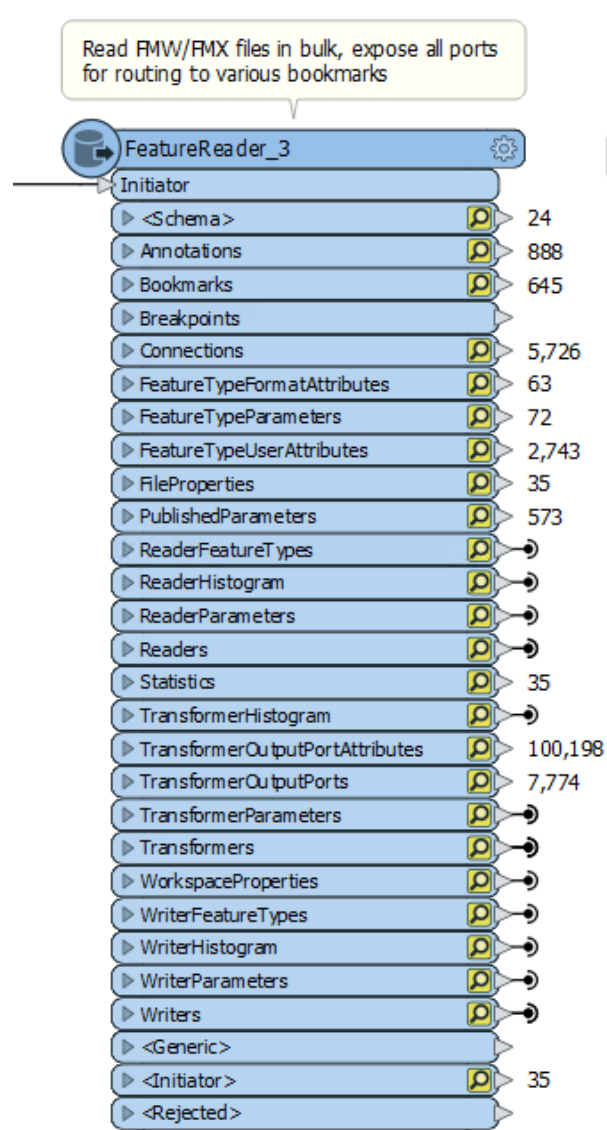
In the image below we can see the process of recursively identifying all FMW/FMX files in a directory, by allowing the user to specify a parent directory and a yes/no recursion option via parameter inputs. Additionally, we use regex to search the file paths for the wildcard filters



2) Bulk Reading Workspaces

We use a second feature reader, this time passing the file paths generated by the first feature reader, to read in all FMW/FMX files in bulk and expose the relevant FME attributes for report processing. We also expose specific attributes including basename and dataset for easily joining and relating feature streams

Read FMW/FMX files in bulk, expose all ports for routing to various bookmarks



Port Name	Count
<Schema>	24
Annotations	888
Bookmarks	645
Breakpoints	
Connections	5,726
FeatureTypeFormatAttributes	63
FeatureTypeParameters	72
FeatureTypeUserAttributes	2,743
FileProperties	35
PublishedParameters	573
ReaderFeatureTypes	
ReaderHistogram	
ReaderParameters	
Readers	
Statistics	35
TransformerHistogram	
TransformerOutputPortAttributes	100,198
TransformerOutputPorts	7,774
TransformerParameters	
Transformers	
WorkspaceProperties	
WriterFeatureTypes	
WriterHistogram	
WriterParameters	
Writers	
<Generic>	
<Initiator>	35
<Rejected>	

FeatureReader Parameters

Transformer Name: FeatureReader_3

Reading

Reader

Format: FME Workspace (FMW)

Dataset: path_windows

Parameters...

Coord. System: Unknown

Constraints

Feature Types to Read: <All Feature Types>

WHERE Clause:

Spatial Filter:

Max Features to Read:

Schema/Data Features

Features To Read: Schema and Data Features

Enable Cache

Output

Output Ports

One per Feature Type

Select 'Additional Attributes to Expose' Items

fme_basename

fme_dataset

fme_feature_type

fme_color

Filter

Select all

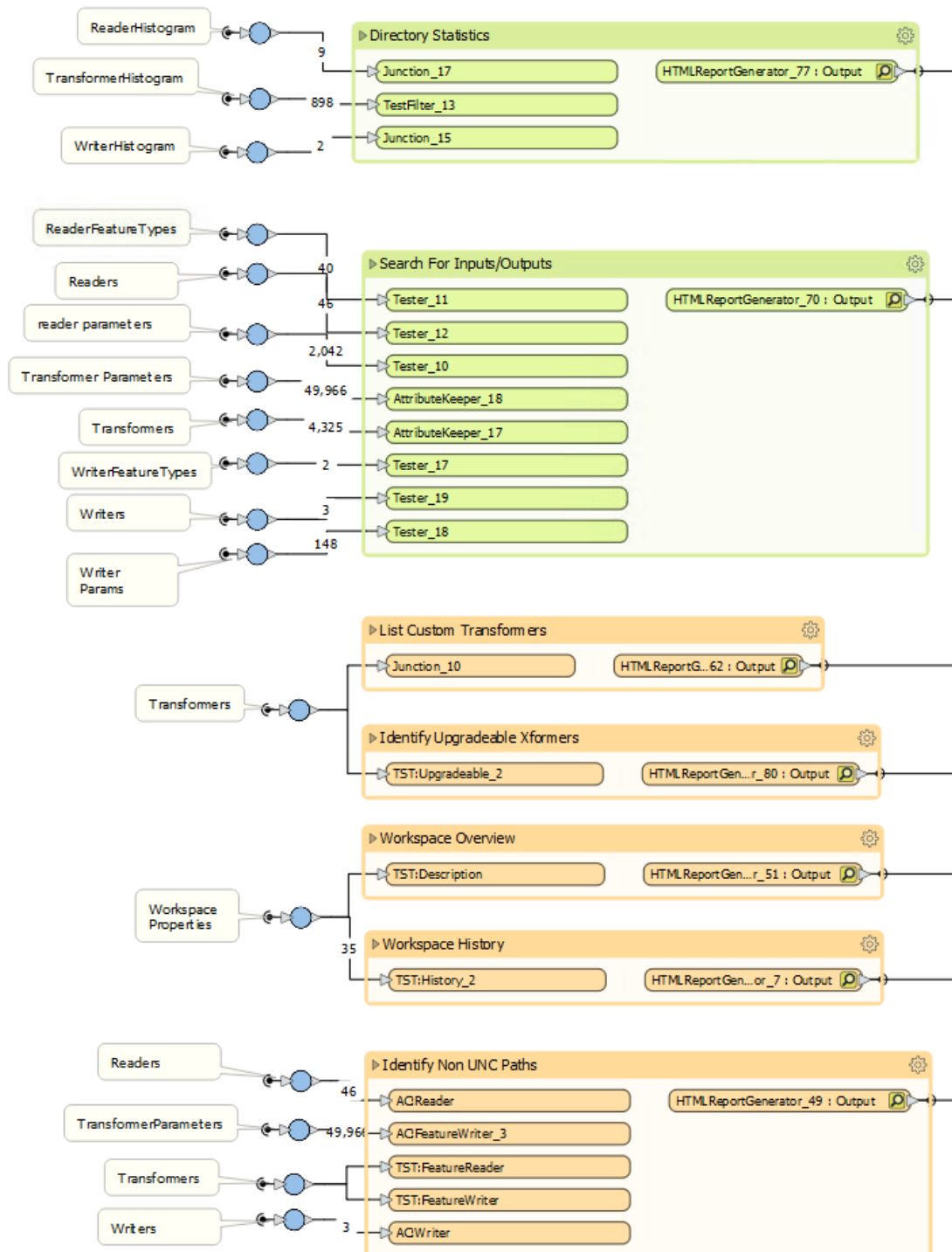
Sorted

OK

Cancel

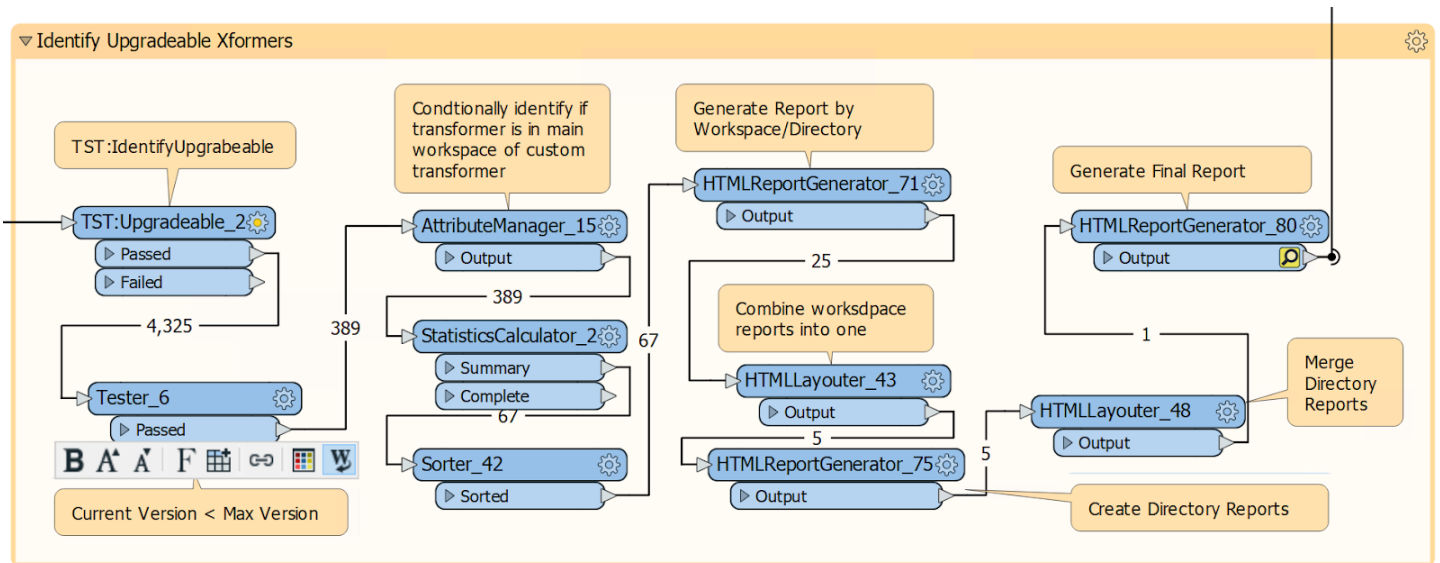
3) Generating Reports

Each report type is contained within its own bookmark and controlled by its own parameter set allowing for full modularity and control when choosing which reports to generate. Additionally, features have been combined into singular junctions to avoid duplication of features where possible. All transformers follow the same basic pattern: Identify needed attributes, join between parameters and feature types if necessary, generate HTML reports based on specified groupings, and pass final HTML report for writing to a tester to determine if we are writing individual reports.



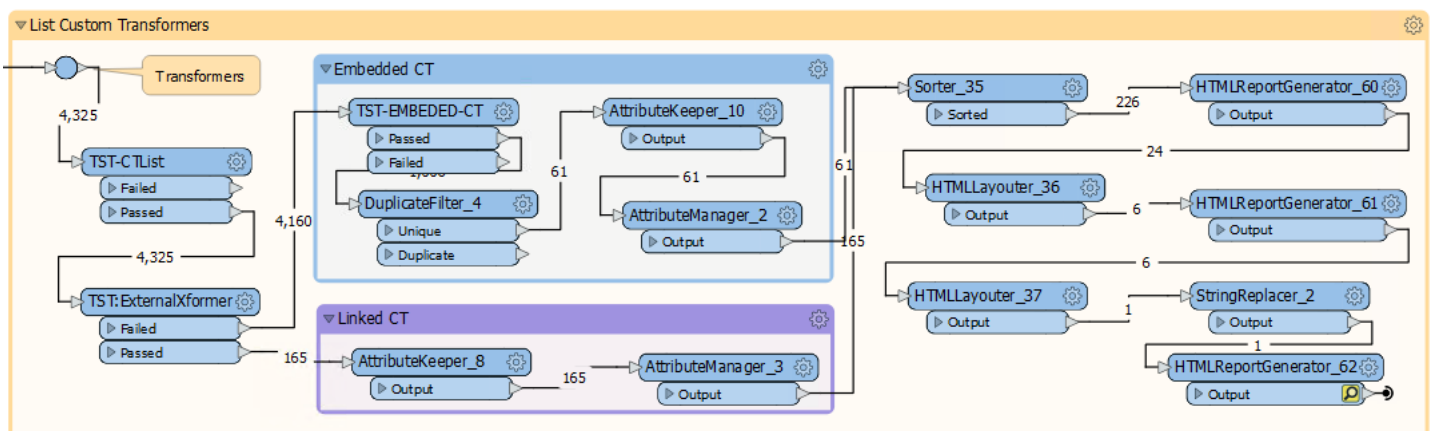
3.1) Report Generation Example 1 - Identify Upgradeable Transformers

In this example, the first step is to use a tester to determine whether this report option is selected. Next, we use another tester to identify if the current transformer version is less than the max transformer version. Next, we conditionally assign a location (Main Workspace VS Custom Transformer) depending on whether it has a transformer parent name. Next, we generate summary statistics of counts, then we begin the multi-level HTML report generation. First, generating individual reports for each workspace, then aggregating those by directory, and then aggregating the directory reports into a single report.



3.2) Report Generation Example 2 - List Custom Transformers

Similar to the previous example, we first test to ensure this report type is selected, then we test to determine whether it is embedded or linked. We process the feature separately, keeping relevant attributes and filtering duplicates so we don't list the same transformer if it's used multiple times per workspace. We assign new attributes based on the current attributes with the attribute manager and then sort the results before generating the multi-tiered HTML reports similar to the first example



Results

The results of this workspace produce a bulk HTML report with the aggregated reports of all metrics and outputs chosen by the user. Optionally, is selected, individual reports are also generated. Below are small screenshots from each report, however, the full reports are included in my application for your review.

Summary Statistics

Workspace Statistics

WORKSPACE HISTORY

WORKSPACE OVERVIEWS

C:/Users/JohnJ/Desktop/FME%20Certification/Project%201/Reports/SummaryStatistics.html

Workspace Statistics

Transformer Statistics - Overall

Parent Directory: I:\Data\TableauVisualization\FMEConversion

Transformer	Count
Total	4371
Transformer	4099
Reader	186
Writer	86

Transformer Statistics by Directory

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\General\

Transformer	Count
Total	2014
Transformer	1881
Reader	95
Writer	38

Missing Workspace History / Descriptions

WORKSPACE HISTORY

C:/Users/JohnJ/Desktop/FME%20Certification/Project%201/Reports/WorkspaceHistory.html

MISSING WORKSPACE HISTORY

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\

Workspace	Folder
TEMPLATE_WORKSPACE	I:\Data\TableauVisualization\FMEConversion\PRODUCTION\

WORKSPACE OVERVIEWS

MISSING WORKSPACE DESCRIPTIONS

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Bridge\

Workspace	Folder
DB_ATIS_STRUC_BRM_LOADER	I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Bridge\
DB_NBI_ALL2SQL	I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Bridge\
OwnerMaintGeometry_StructurePrimaryKeyValidation	I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Bridge\

Linked Vs Embedded & Upgradeable Transformers

Linked & Embedded Custom Transformer Reporting

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\

- Workspace: TEMPLATE_WORKSPACE

Transformer Type	Transformer Name
Linked	ADOT_Filter
Linked	ADOT_Filter_4
Linked	ADOT_Filter_2
Linked	ADOT_Filter_3

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\HighwayLog\

- Workspace: SHS_HIGHWAY_LOG_SUMMARY

Transformer Type	Transformer Name
Embedded	ADOT_SHS_Log_Pivot
Embedded	ADOT_Overlay_Overlap_Filter
Embedded	ADOT_LRS_Overlays
Embedded	RH_DomainChecker
Embedded	ADOT_RampConnectorExtractor
Embedded	ADOT_BeforeAfterLengthDiff

UPGRADEABLE TRANSFORMERS

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Bridge\

- Workspace: DB_ATIS_STRUC_BRM_LOADER

Location	# Upgradeable
Main Workspace	8
Custom Transformer: LRS_Event2SL_2	4
Custom Transformer: LRS_Event2SL_3	4
Custom Transformer: LRS_Event2SL_4	4
Custom Transformer: LRS_Event2SL_5	3

- Workspace: OwnerMaintGeometry_StructurePrimaryKeyValidation

Location	# Upgradeable
Main Workspace	3

Input/Output Dataset Reporting with Existence Check Enabled on Inputs

INPUT/OUTPUT DATASET REPORTING

INPUT DATASETS

NON-EXISTENT DATASETS

UNC Paths

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\General\

Workspace Name	Transformer Name	Transformer Type	Dataset
GDM_RAH_FETCHER	FRD-GDM-MAINT	FeatureReader	\\dot.state.az\AdotFiles\GIS\GISTeam\Data\GISDataMart\GDM_Events_1.gdb

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Utility_Workspaces\

Workspace Name	Transformer Name	Transformer Type	Dataset
UTL_CustomXfmrExaminer	SHS_SUMMARY_LOG [FMW]	Reader	\\dot.state.az\AdotFiles\GIS\GISTeam\Data\TableauVisualization\FMEConversion\PRODUCTION\General\GDM_MMS_Polygonizer_SQL.fmw
UTL_CustomXfmrExaminer	SHS_SUMMARY_LOG [FMW]	Reader	\\dot.state.az\AdotFiles\GIS\GISTeam\Data\TableauVisualization\FMEConversion\PRODUCTION\General\SHS_SUMMARY_LOG.fmw

EXISTING DATASETS

Local Disk Paths

Directory: I:\Data\TableauVisualization\FMEConversion\PRODUCTION\Bridge\

Workspace Name	Transformer Name	Transformer Type	Dataset
DB_NBI_ALL2SQL	AZ2009 [CAT]	Reader	I:\Data\TableauVisualization\DataSources\Bridge\AZ2009.NBI
DB_NBI_ALL2SQL	AZ2016 2nd [CAT]	Reader	I:\Data\TableauVisualization\DataSources\Bridge\AZ2016 2nd.NBI