

Integrating Real-time GIS and Social Media for Qualitative Transportation Data Collection

PI: Dr. Hongmian Gong

Co-PI: Dr. Carsten Kessler

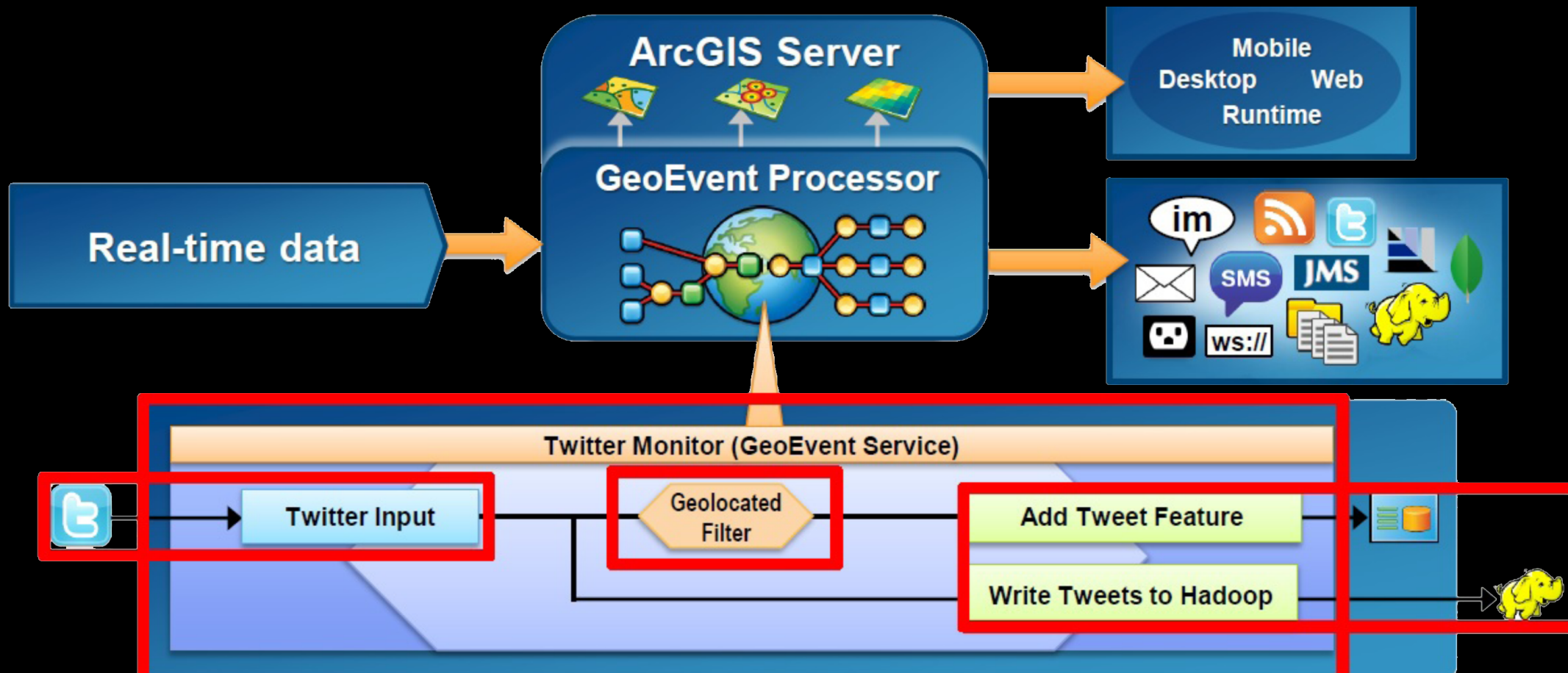
Mario Giampieri

Why social media data?

- Transportation planning has traditionally focused on collecting quantitative data for modelling, focusing on how much people travel
- Qualitative data allows us to model why people use different modes for different purposes
- This study seeks to determine how real-time transportation information affects daily travels

Research Participants

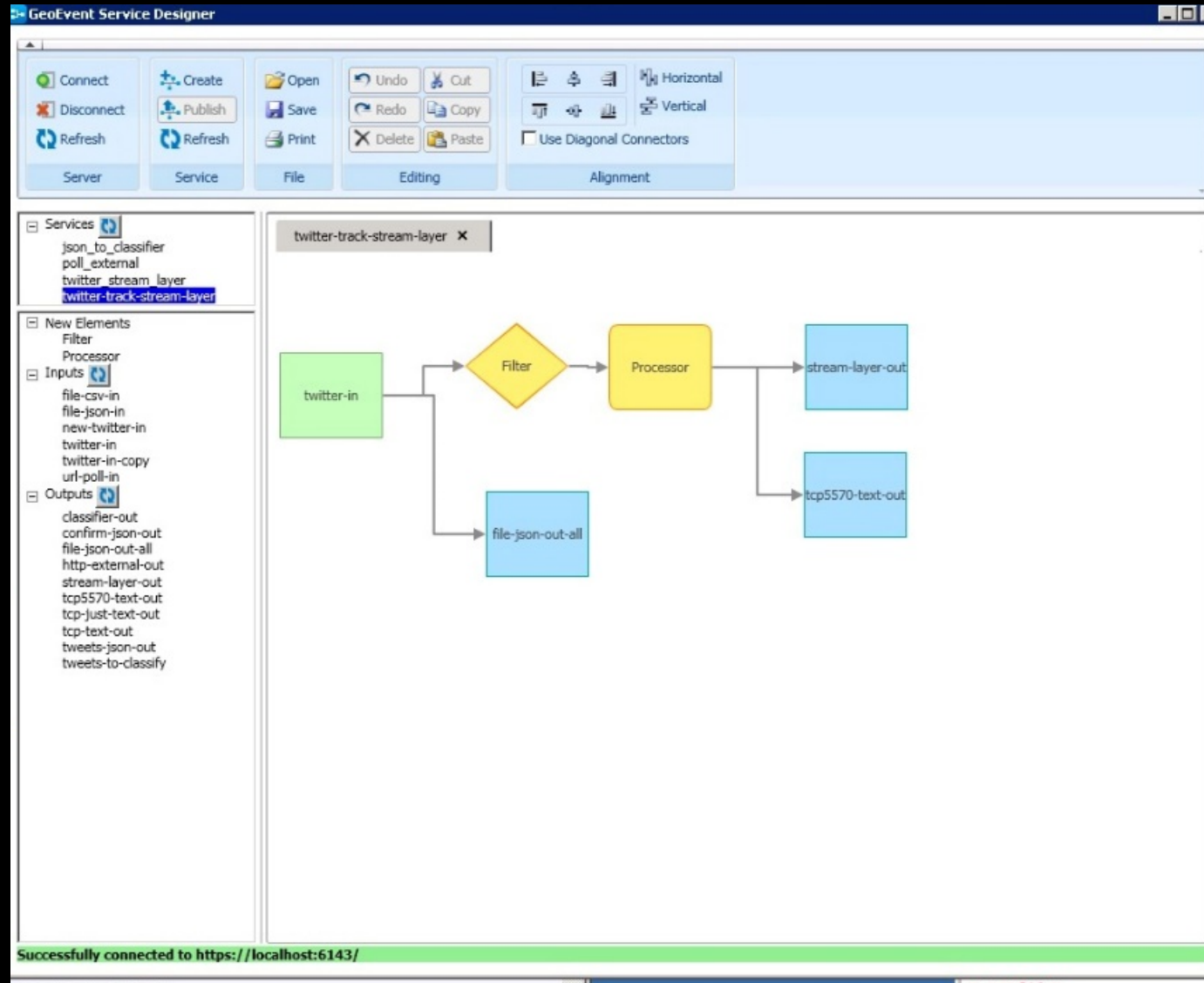
- Participants have installed a GPS tracking app that report their locations
- Users allow us to inspect data they publish on Twitter, and have that data analyzed using machine learning algorithms



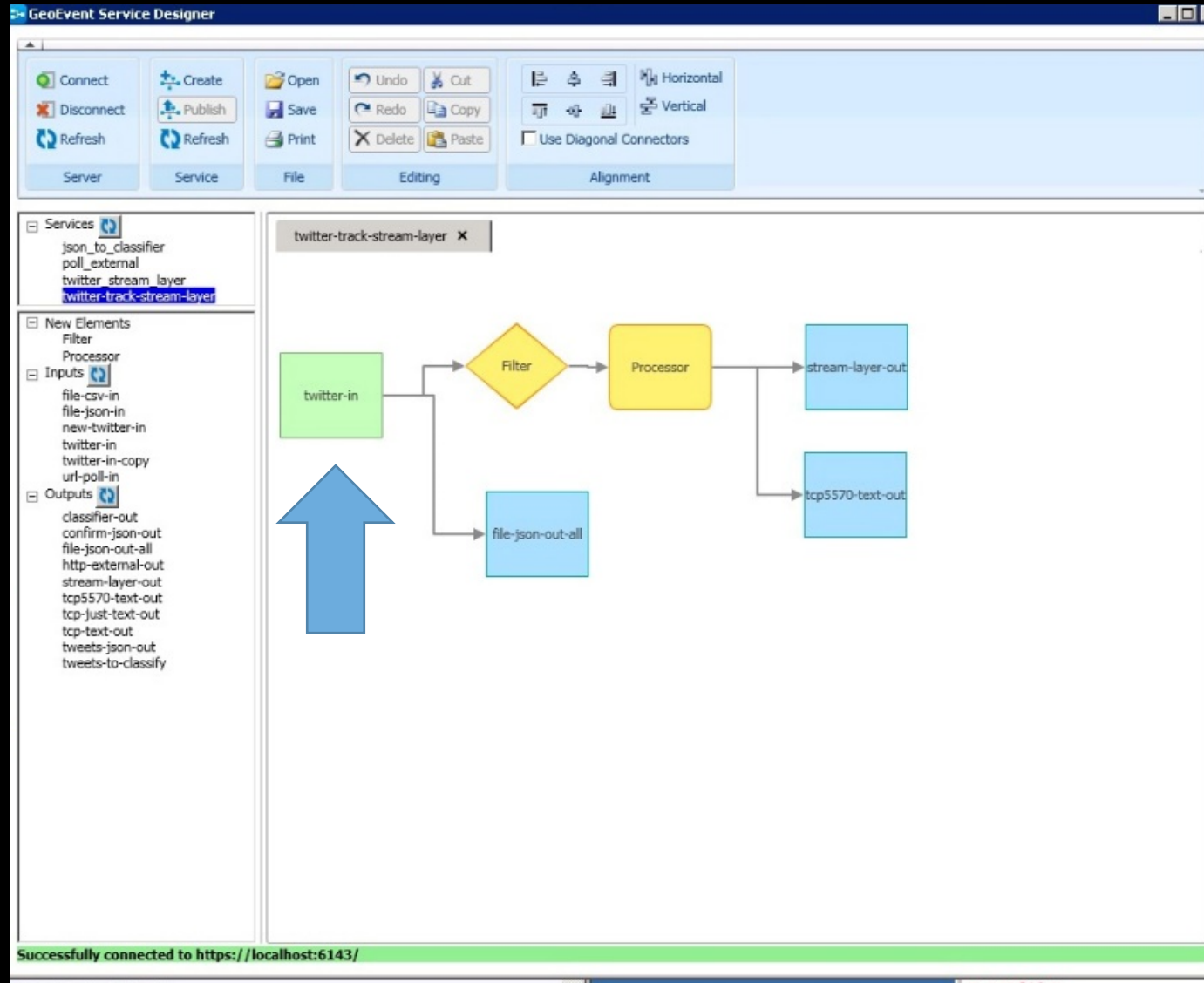
Server infrastructure setup

- Amazon Virtual Machine Instance running Windows Server 2008
- ESRI ArcServer 10.2
- Esri Geoevent Processor for Server 10.2
- mySQL database

ESRI Geoevent Processor



Ingest Twitter data



Twitter developer API

The screenshot shows a web browser window with the URL <https://apps.twitter.com/app/8054663>. The page is titled "Application Management" and features the "utrc" application logo. The application is identified as "UTRC response app" with the website <http://www.utrctweettracker.com>. The "Details" tab is selected, showing the application's organization information and settings. The organization is listed as "None". The application settings include the access level, consumer key (API key), callback URL, and various OAuth URLs.

utrc [Test OAuth](#)

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

 UTRC response app
<http://www.utrctweettracker.com>

Organization

Information about the organization or company associated with your application. This information is optional.

Organization	None
Organization website	None

Application Settings

Your application's Consumer Key and Secret are used to [authenticate](#) requests to the Twitter Platform.

Access level	Read and write (modify app permissions)
Consumer Key (API Key)	64TSew7rityQqxCIIVwvA7Nq (manage keys and access tokens)
Callback URL	None
Callback URL Locked	No
Sign in with Twitter	No
App-only authentication	https://api.twitter.com/oauth2/token
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize
Access token URL	https://api.twitter.com/oauth/access_token

Twitter developer API

The screenshot shows the Twitter Application Management interface for an application named 'utrc'. The browser address bar displays the URL <https://apps.twitter.com/app/8054663/keys>. The page has a blue header with the Twitter logo and the text 'Application Management'. Below the header, the application name 'utrc' is displayed, along with a 'Test OAuth' button. A navigation bar contains links for 'Details', 'Settings', 'Keys and Access Tokens', and 'Permissions'. The 'Settings' tab is active, showing 'Application Settings'. A warning message states: 'Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.' The settings table includes: Consumer Key (API Key) as '64TSew7rityQxC1IVvvvA7Nq', Consumer Secret (API Secret) as 'ZGn2tN9BeGsESQLI7AXBtbcqM5QiWYpJQ9gZ7udbbeNpdx8J2O', Access Level as 'Read and write (modify app permissions)', Owner as 'mario_giampieri', and Owner ID as '298685377'. Below the settings is the 'Application Actions' section with buttons for 'Regenerate Consumer Key and Secret' and 'Change App Permissions'. The 'Your Access Token' section follows, with a warning: 'This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.' The access token details table shows: Access Token as '298685377-WRyLdns3zaJN3Mc9owHOqn3l5SR3mJdYY60xTot', Access Token Secret as 'TKqTe9Ypl2bDlkOCS5FxDfDgfhTmOeNEXzeMyY30uTmcU', Access Level as 'Read and write', and Owner as 'mario_giampieri'.

utrc [Test OAuth](#)

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	64TSew7rityQxC1IVvvvA7Nq
Consumer Secret (API Secret)	ZGn2tN9BeGsESQLI7AXBtbcqM5QiWYpJQ9gZ7udbbeNpdx8J2O
Access Level	Read and write (modify app permissions)
Owner	mario_giampieri
Owner ID	298685377

Application Actions

[Regenerate Consumer Key and Secret](#) [Change App Permissions](#)

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	298685377-WRyLdns3zaJN3Mc9owHOqn3l5SR3mJdYY60xTot
Access Token Secret	TKqTe9Ypl2bDlkOCS5FxDfDgfhTmOeNEXzeMyY30uTmcU
Access Level	Read and write
Owner	mario_giampieri

GeoEvent processor Twitter connector

ArcGIS GeoEvent Processor Manager - Windows Internet Explorer

https://localhost:6143/geoevent/manag

ArcGIS GeoEvent Processor ...

Help | Sign Out

ArcGIS GeoEvent Processor Manager

Services Site Security Logs

Monitor Inputs GeoEvent Services Outputs

twitter-in (Receive Tweets) Save Cancel

Name*: ? twitter-in

ConsumerKey*: ?

ConsumerSecret*: ?

AccessToken*: ?

AccessTokenSecret*: ?

Follow: ? 4082387427

Track: ?

Locations: ?

Advanced

Count: ? 0

GeoEvent processor Twitter connector

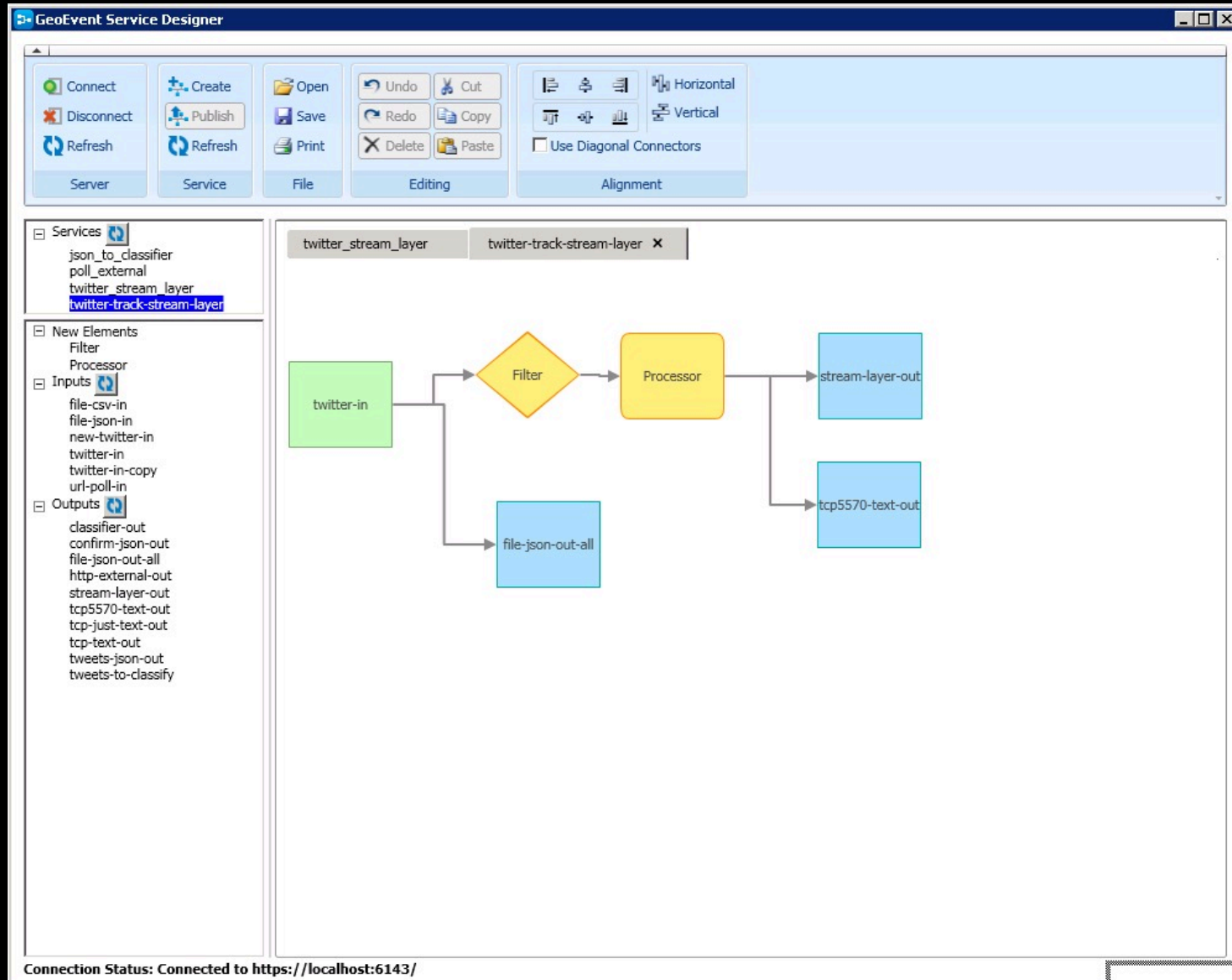
The screenshot shows the ArcGIS GeoEvent Processor Manager web interface in a Windows Internet Explorer browser. The browser's address bar displays `https://localhost:6143/geoevent/manag`. The page title is "ArcGIS GeoEvent Processor Manager". The navigation bar includes tabs for "Monitor", "Inputs", "GeoEvent Services", and "Outputs", with "Inputs" currently selected. A secondary navigation bar contains links for "Services", "Site", "Security", and "Logs". In the top right corner, there are links for "Help" and "Sign Out".

The main content area is titled "twitter-in (Receive Tweets)" and includes "Save" and "Cancel" buttons. The configuration fields are as follows:

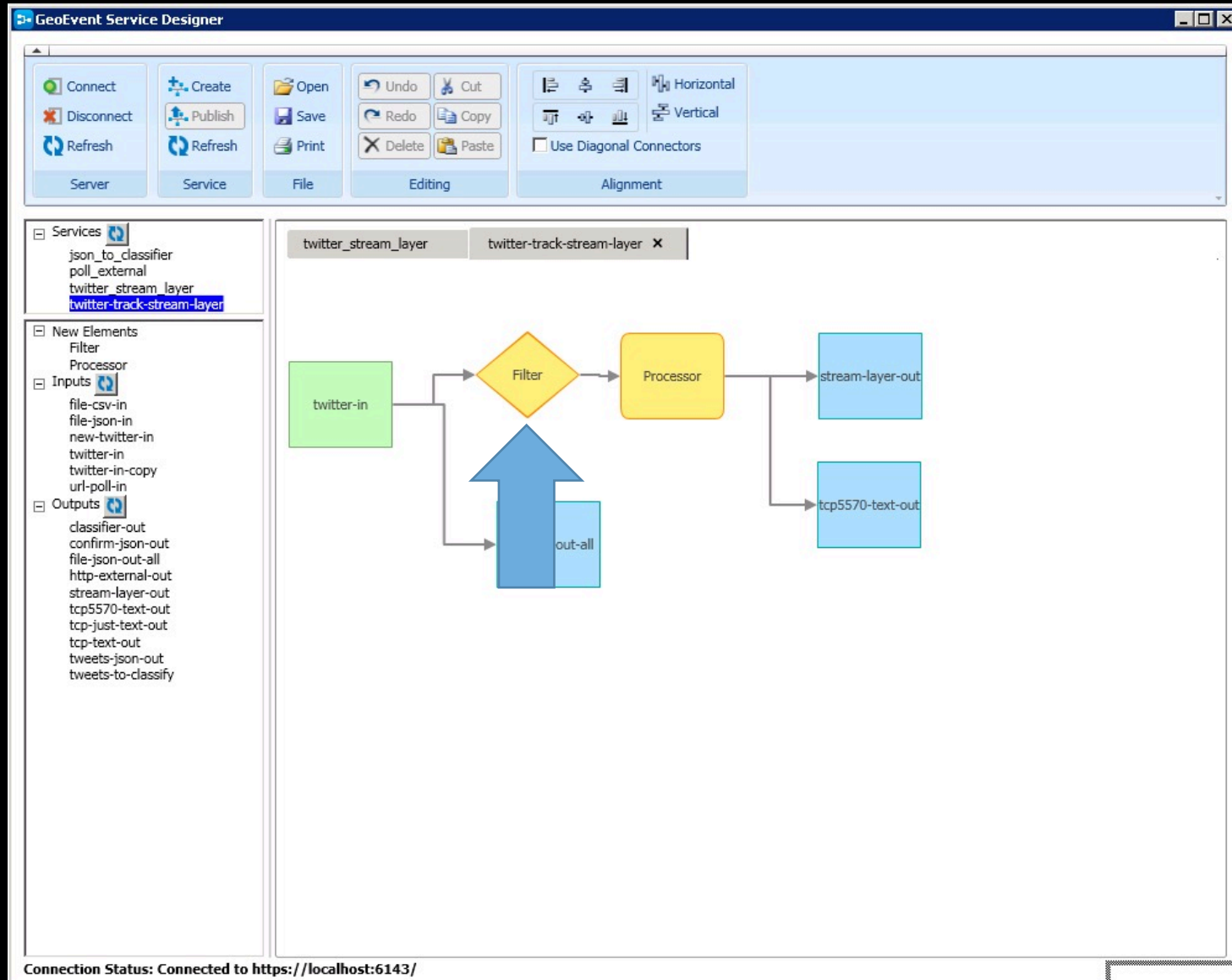
- Name***: A text field containing "twitter-in".
- ConsumerKey***: A text field containing six dots.
- ConsumerSecret***: A text field containing six dots.
- AccessToken***: A text field containing six dots.
- AccessTokenSecret***: A text field containing six dots.
- Follow:** A text field containing "4082387427".
- Track:** An empty text field.
- Locations:** An empty text field.

Two large blue arrows point from the left towards the "Follow:", "Track:", and "Locations:" fields. Below these fields is an "Advanced" section, which is currently collapsed. The "Advanced" section contains a "Count:" field with a value of "0".

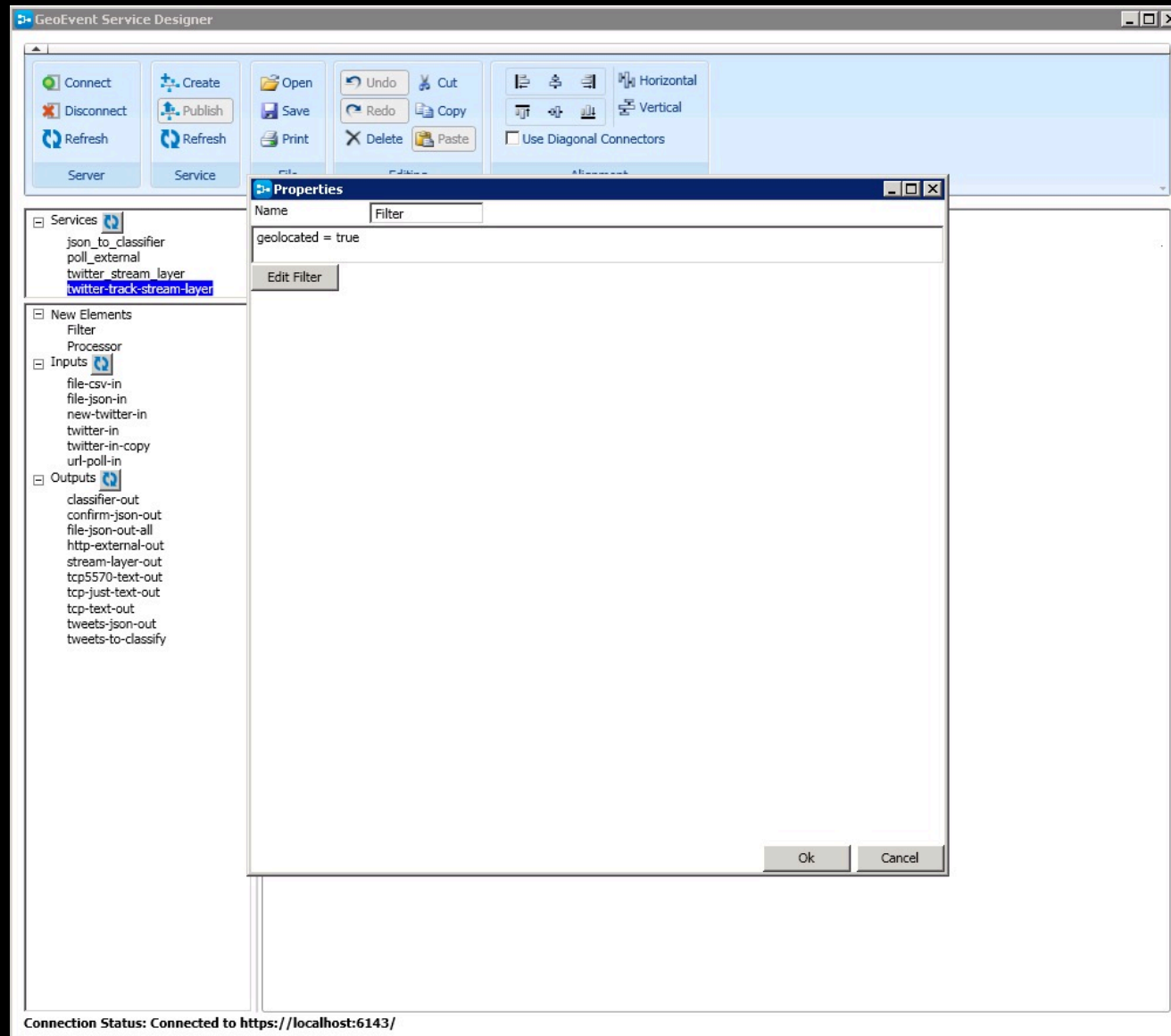
GeoEvent processor



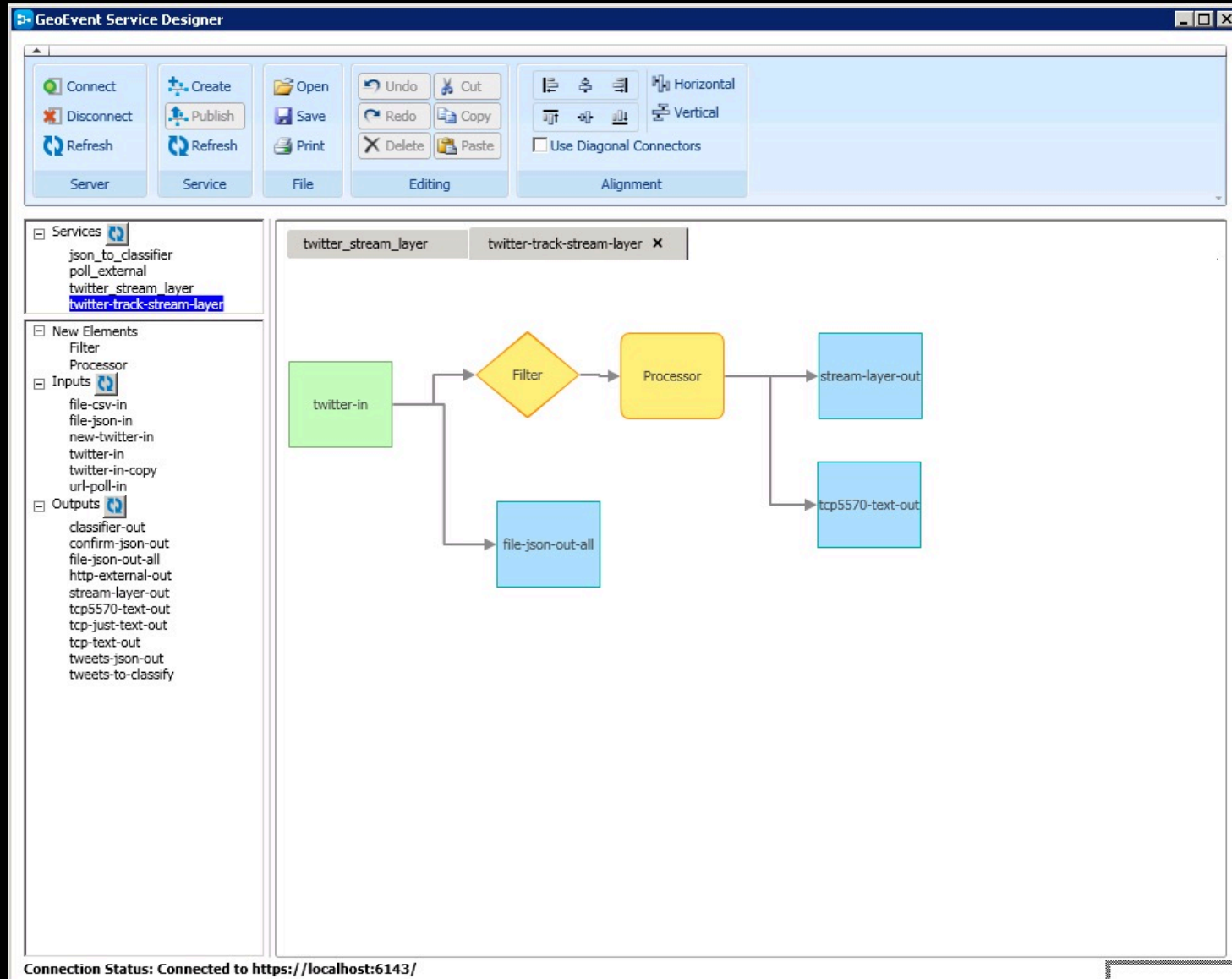
Add filter



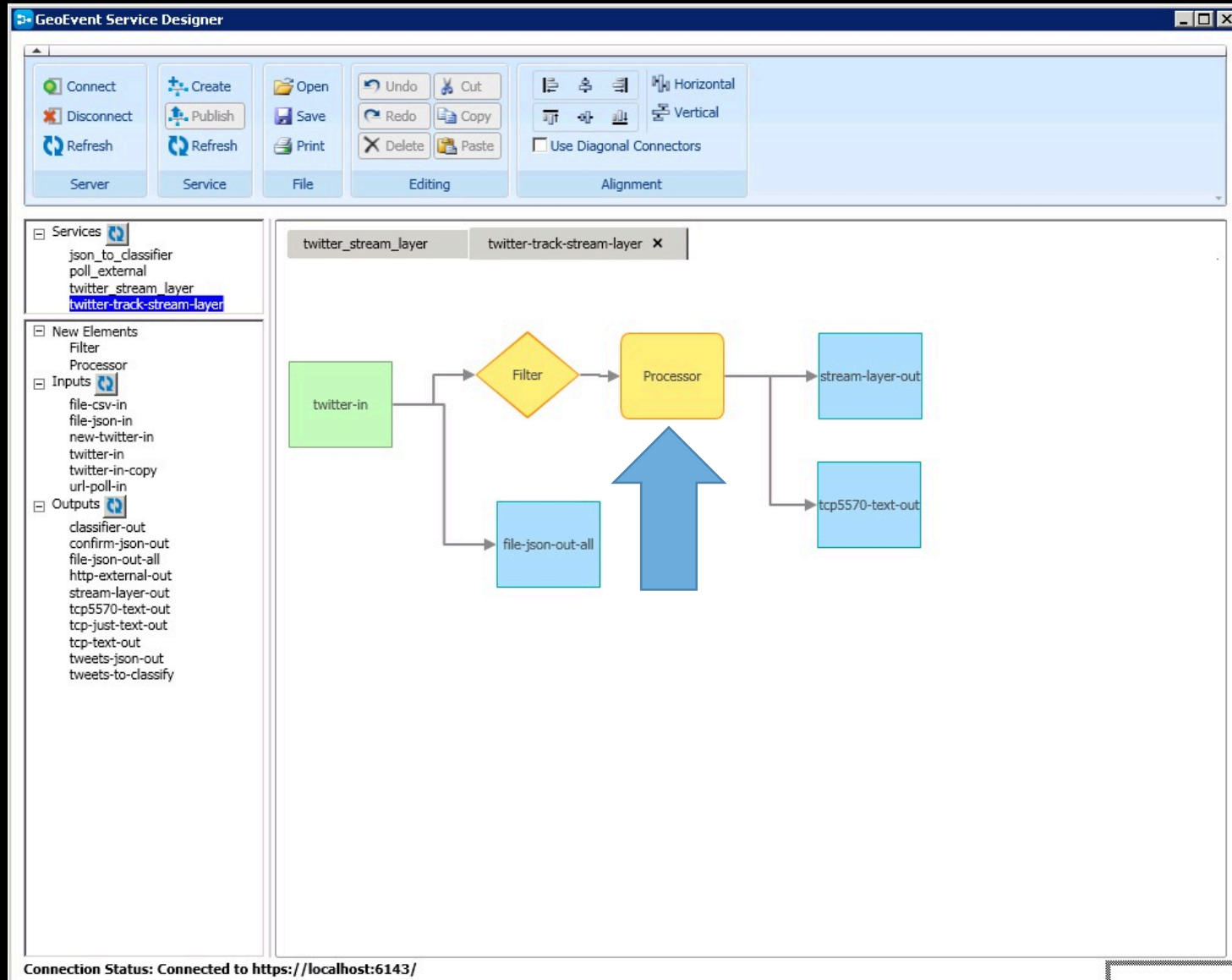
Filter by geolocation



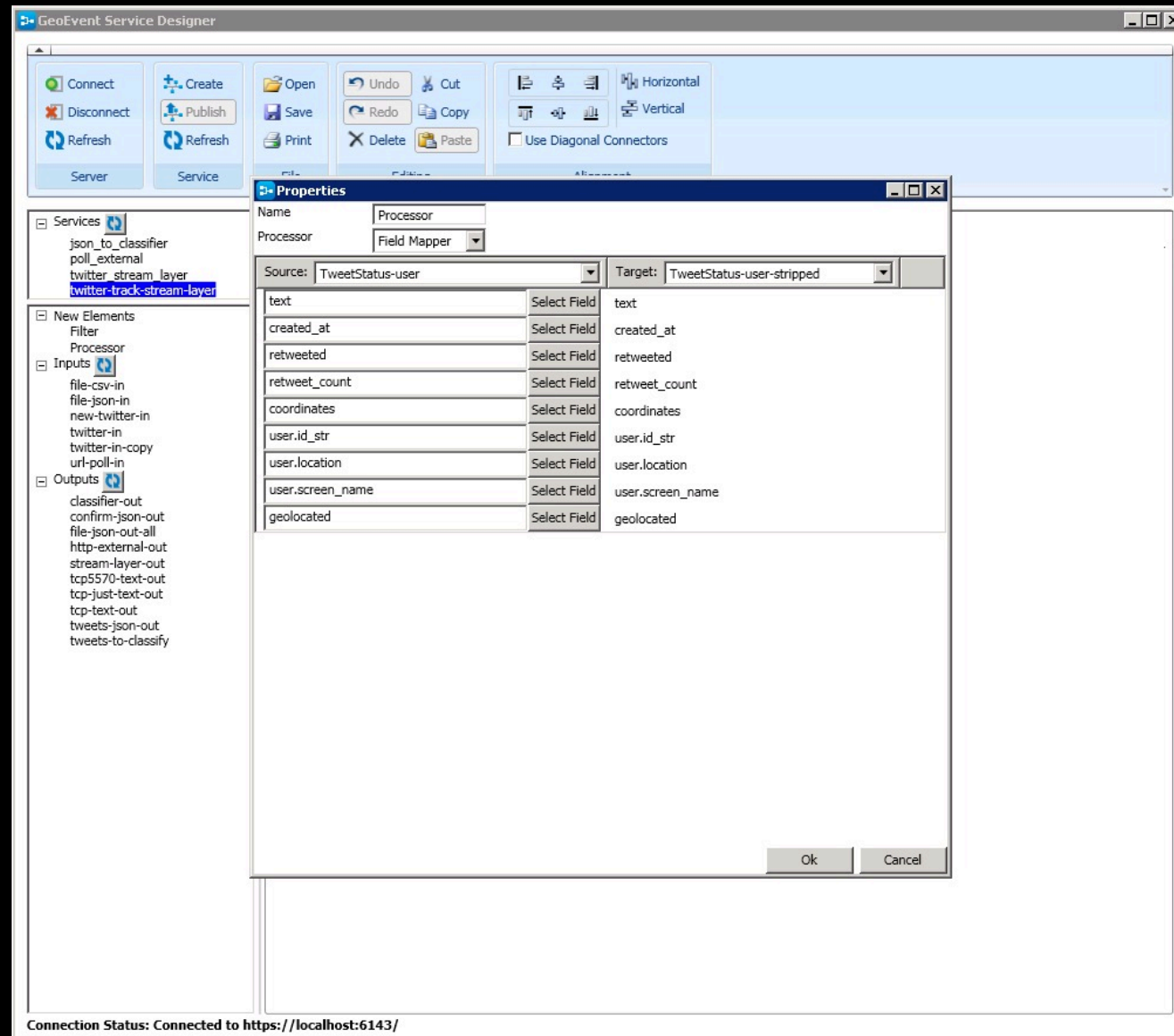
GeoEvent processor



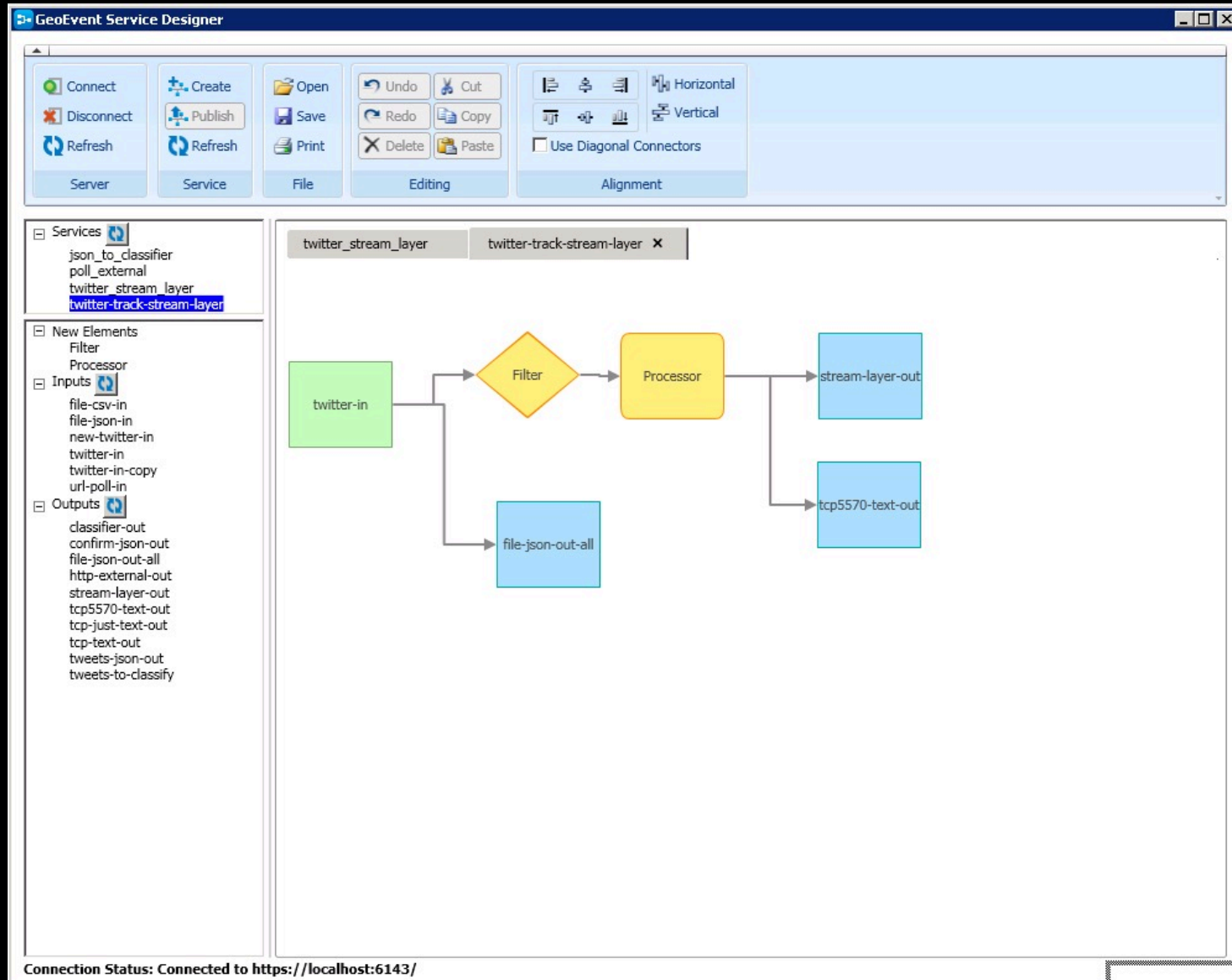
Add processor



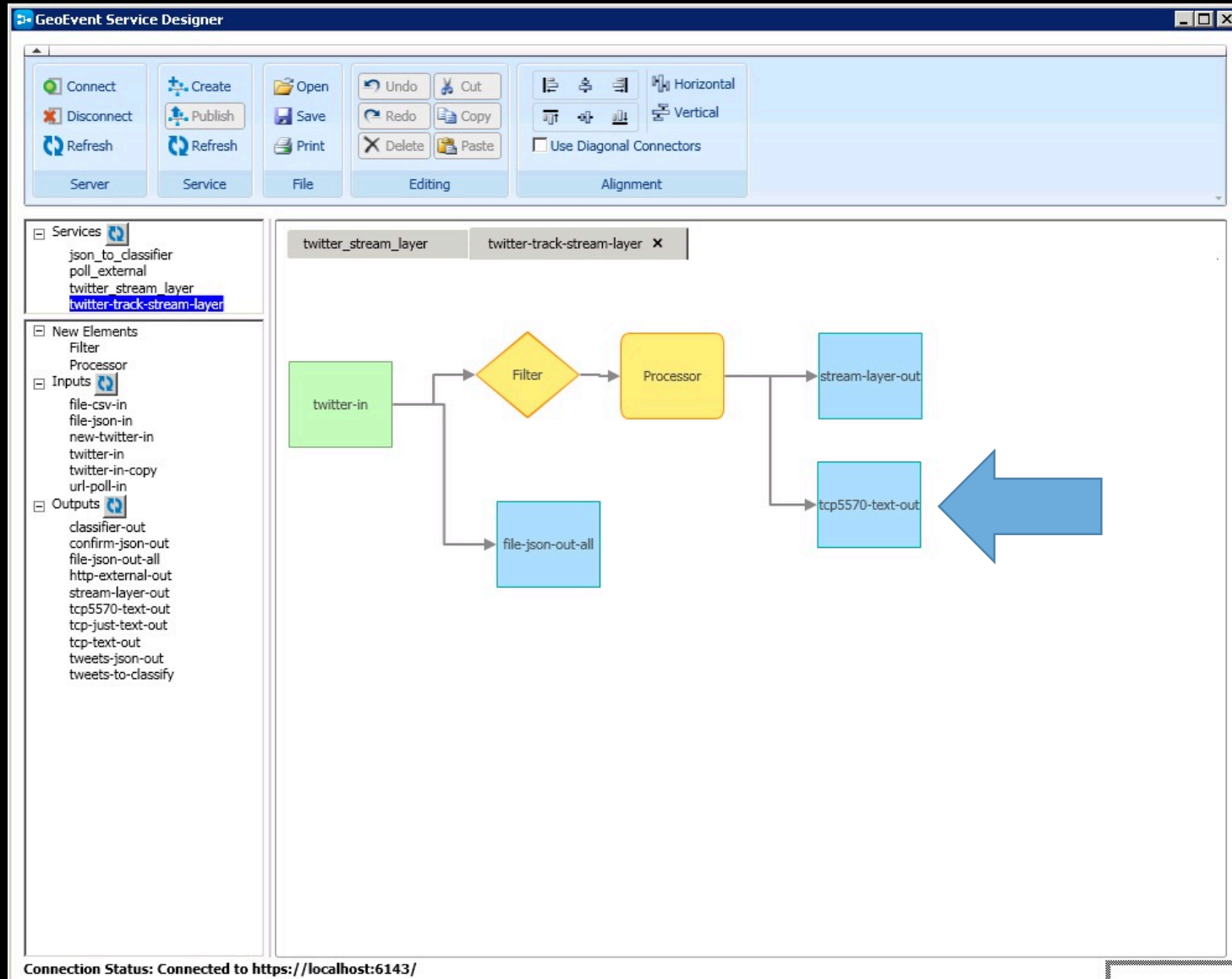
Map fields to remove extraneous data



GeoEvent processor



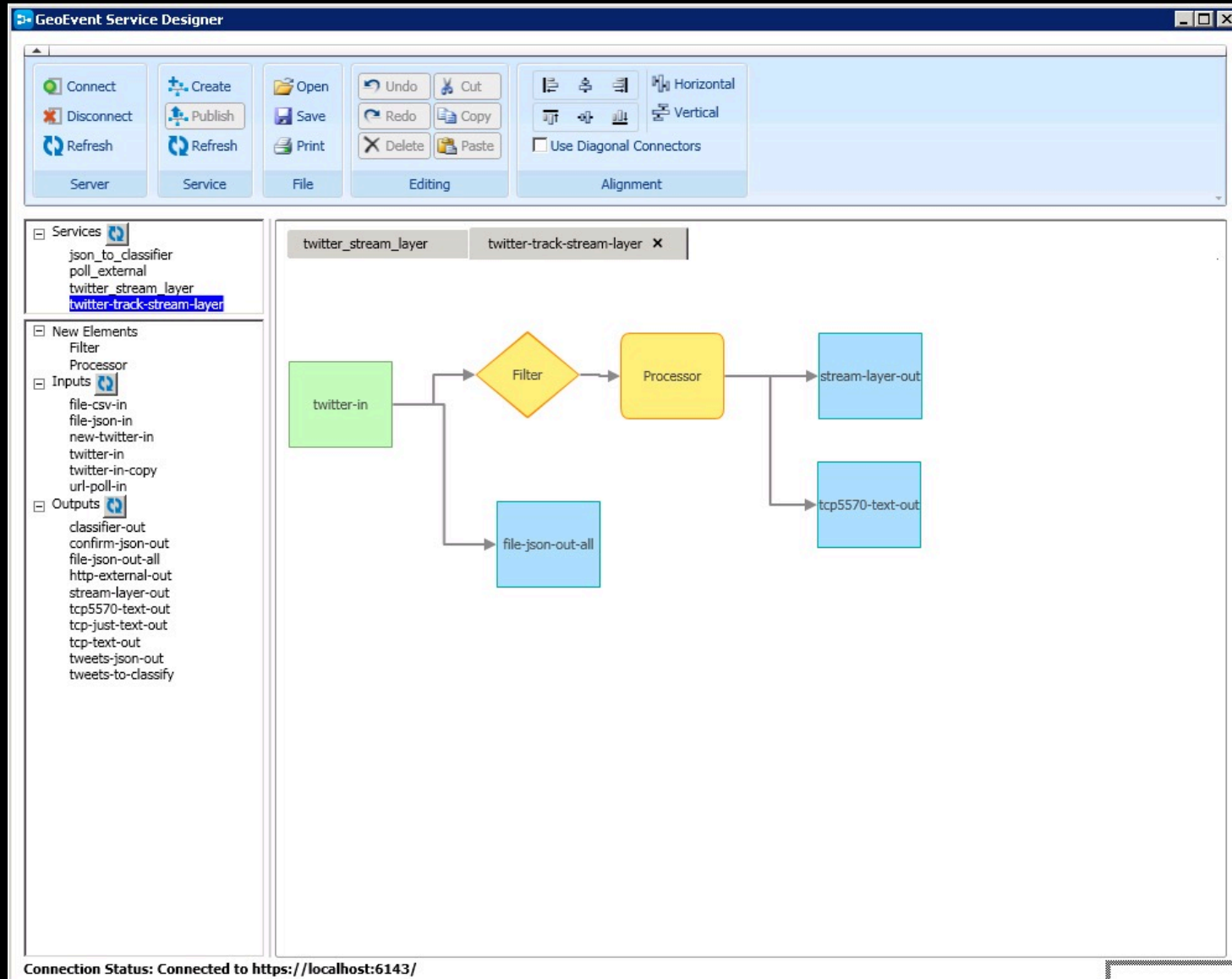
Print tweet data to console



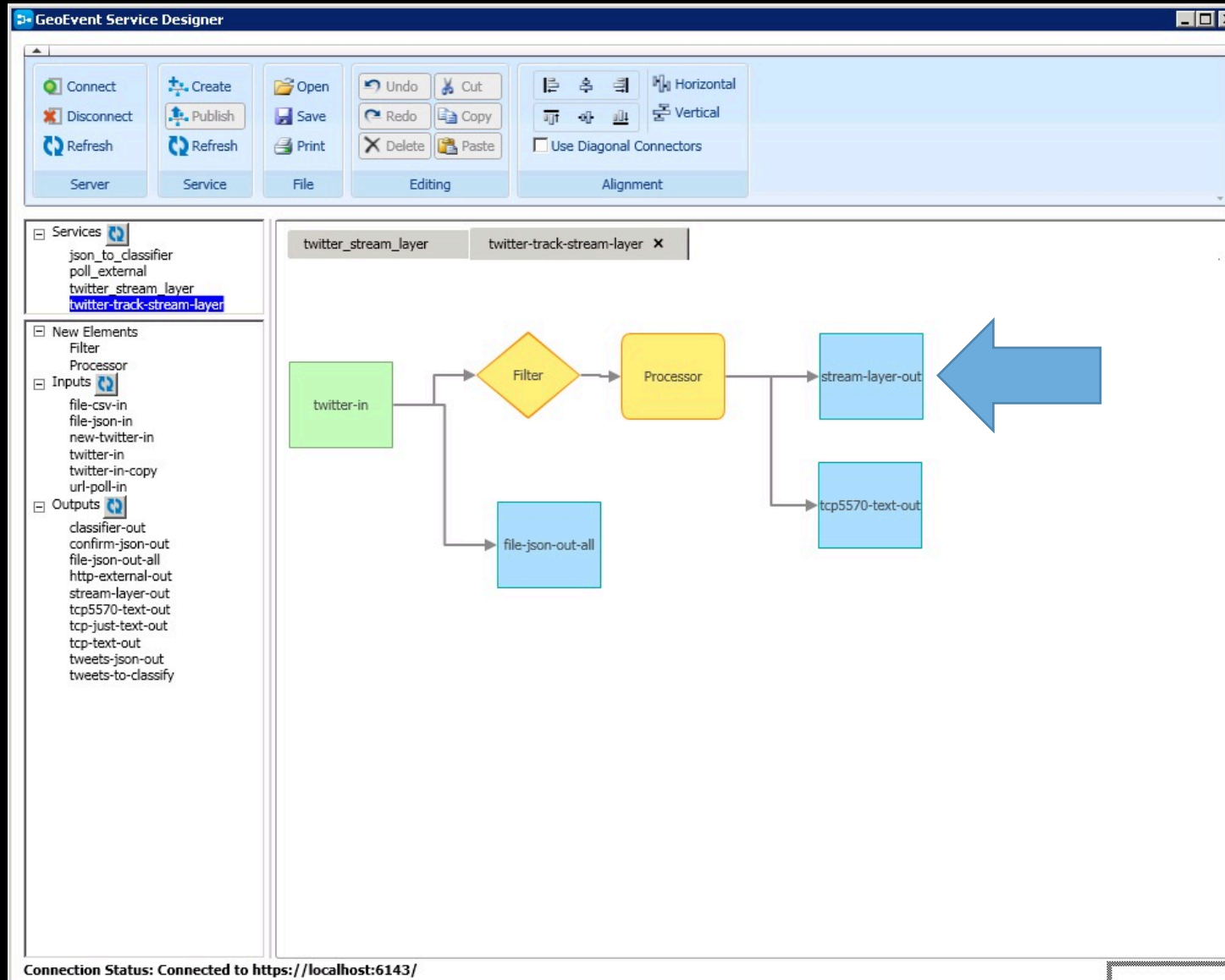
Print tweet data to console

```
C:\Windows\system32\cmd.exe
ia https://t.co/7FSGc7zi1P,11/20/2015 02:08:12 AM,false,0,"-122.400067,45.53389
4000000004,0.0",667524816010768384,,,false,false,false,,,true
Received Event : TweetStatus,,@DanielWickham93 @sunny_hundal @MazMHussain @mehd
irhasan Assuming genuine, Paris rationale outlined in statement: https://t.co/v6
NxXiKoD5",11/20/2015 02:08:14 AM,false,0,"-2.23348,53.4569525,0.0",6675248220800
32768,DanielWickham93,667515571962028037,false,false,false,476654184,,,true
Received Event : TweetStatus,,@maddow What Trump is proposing is what the Nazis
did to Jews.,11/20/2015 02:08:20 AM,false,0,"-121.3187105,38.641305,0.0",6675248
46524325888,maddow,,,false,false,,16129920,,,true
Received Event : TweetStatus,,My pretty city from the 19th Fl of the Hippodrome
Bldg 1120 6th Ave #nyc #nycraalestate #nycviewsà https://t.co/BBfZRUhlg2,11/20/2
015 02:08:26 AM,false,0,"-73.9828491,40.7555885,0.0",667524873682599936,,,false,
false,false,,,true
Received Event : TweetStatus,,What Trump said doesn't scare me as much as the su
pport he'll get for saying it. #maddow #rwnj,11/20/2015 02:08:28 AM,false,0,"-85
.5900935,38.25192,0.0",667524880859062272,,,false,false,,,,true
Received Event : TweetStatus,,Trump veut enregistrer les identités des musulmans
américains. Attendez. Qui a déjà eu cette idée? J'ai peur d'un coup. #heuessom
bres,11/20/2015 02:08:28 AM,false,0,"0.31429950000000042,42.308346,0.0",667524880
003272705,,,false,false,,,,true
Received Event : TweetStatus,,@_simbuh bush senior I presume? ... Buchanan, Car
ter, Bush, Obama. Obama being the worst.",11/20/2015 02:08:33 AM,false,0,"-85.69
5933,37.9995225,0.0",667524903118249984,_simbuh,667524055315177472,false,false,,
340185682,,,true
```

GeoEvent processor



Stream Tweet data to web map



Stream Tweet data to web map

ArcGIS GeoEvent Processor Manager - Windows Internet Explorer

https://localhost:6143/geoev... ArcGIS GeoEvent Processor ...

Help | Sign Out

ArcGIS GeoEvent Processor Manager

Services Site Security Logs

Monitor Inputs GeoEvent Services Outputs

stream-layer-out (Publish features to a stream layer) Save Cancel

Name: stream-layer-out

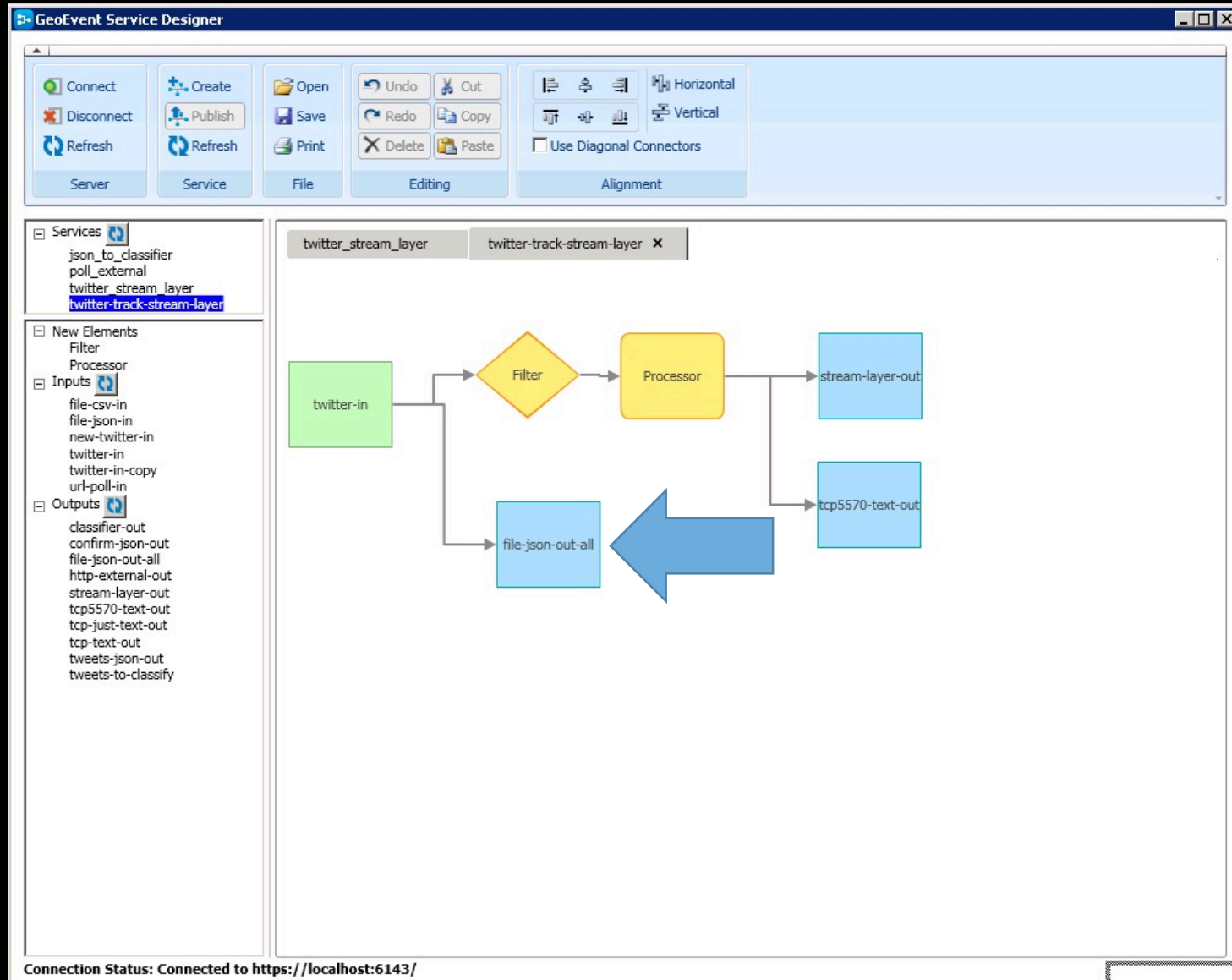
URL Path: /StreamService



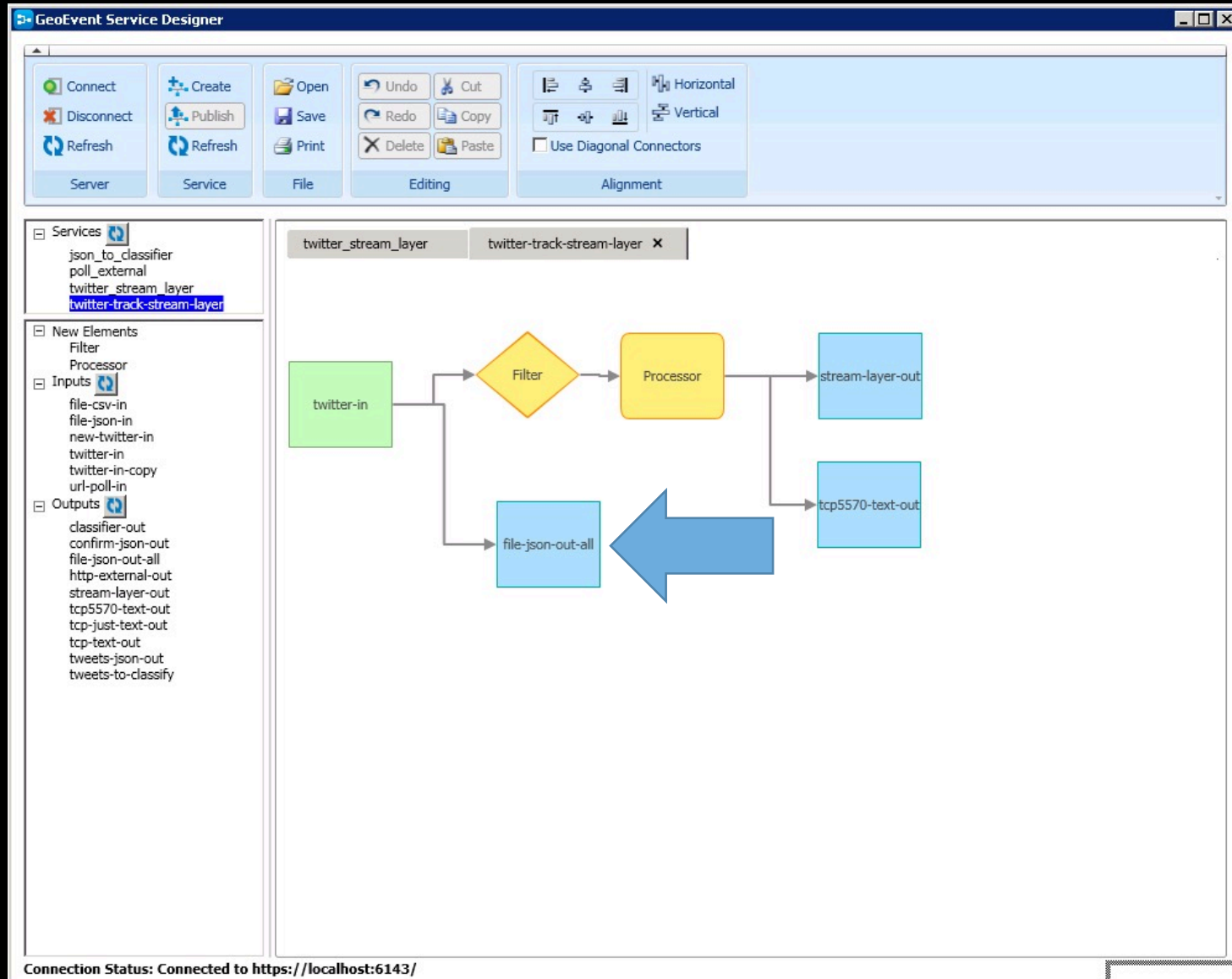
Enter websocket connection: ws://localhost:6180/StreamService

Make Stream Layer Disconnect Stream Layer

GeoEvent processor



Write Tweet data to JSON



Write Tweet data to JSON

ArcGIS GeoEvent Processor Manager - Windows Internet Explorer

https://localhost:6143/geoever

ArcGIS GeoEvent Processor ...

Help | Sign Out

ArcGIS GeoEvent Processor Manager

Services Site Security Logs

Monitor Inputs GeoEvent Services Outputs

file-json-out-all (Write to a .json file) Save Cancel

Name*:

Folder*: Register Folder

Filename Prefix:

Advanced

Subfolder:

File Extension:

File Rollover Method:

File Rollover Size (KB):

File Cleanup Age (Minutes):

File Cleanup Method:

Formatted JSON*: ☐ Yes ☒ No

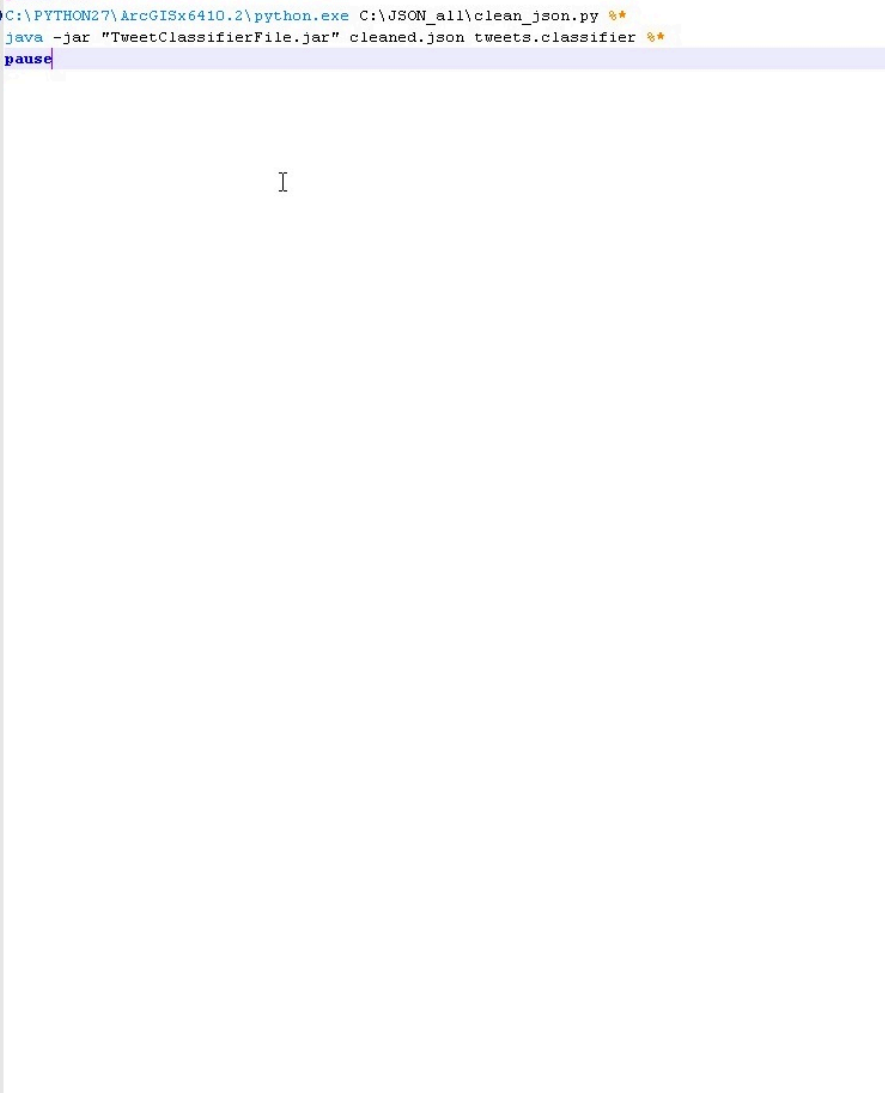
Classify tweets

```
C:\dir > java -jar TweetClassifier.jar tweets.json tweets.classifier
```

Classify tweets

```
C:\dir > java -jar TweetClassifier.jar tweets.json tweets.classifier
```

```
"classifier":{"none":0.011179990893958146,"shopping":  
0.3480679872338228,"workschool":0.08498662517955215,"social":  
0.01098021382705481,"eat":0.01570851139120672,"workother":  
0.0021255868285692445,"dropoff":  
0.14357428557518584,"recreation":  
0.23121059429062144,"otherfam":0.0823702210443366}
```

C:\JSON_all\batch_classify.bat - Notepad++

File Edit Search View Encoding Language Settings Macro Run Plugins Window ?

TwitterStreamLayer.html batch_classify.bat

```
1 @echo off
2 C:\PYTHON27\ArcGISx6410.2\python.exe C:\JSON_all\clean_json.py **
3 java -jar "TweetClassifierFile.jar" cleaned.json tweets.classifier **
4 pause
```

Batch file length: 154 lines: 4 Ln: 4 Col: 6 Sel: 0 | 0 Dos\Windows ANSI as UTF-8 INS

```

Hostname      : IP-0A4CDA72
Instance ID   : i-fd991492
Public IP Address : 54.196.55.93
Private IP Address : 10.230.209.194
Availability Zone : us-east-1a
Instance Size : c3.large
Architecture   : AMD64

```

```

Python Shell
File Edit Shell Debug Options Windows Help

str": "660474904958124033", "favorited": "false", "truncated": "false", "place": {"id":
"002e24c6736f069d", "full_name": "Bronx, NY", "url": "https://api.twitter.com/1.1/ge
o/id/002e24c6736f069d.json"}, "user": {"id_str": "4082387427", "name": "mgia", "follow
ers_count": 0, "screen_name": "utrc_test_tweet"}, "geolocated": "true"}
>>> ===== RESTART =====
>>>

Traceback (most recent call last):
  File "C:\JSON_all\clean_json.py", line 5, in <module>
    line = myfile.readline()
IOError: File not open for reading
>>> ===== RESTART =====
>>>

Traceback (most recent call last):
  File "C:\JSON_all\clean_json.py", line 9, in <module>
    myfile.write(["+line+""])
NameError: name 'line' is not defined
>>> ===== RESTART =====
>>>

Traceback (most recent call last):
  File "C:\JSON_all\clean_json.py", line 7, in <module>
    for line in myfile.readlines():
IOError: File not open for reading
>>> ===== RESTART =====
>>>

Traceback (most recent call last):
  File "C:\JSON_all\clean_json.py", line 2, in <module>
    with open('output_classified.json', "r") as file, open('cleaned.txt', "w") a
s outfile:
IOError: [Errno 2] No such file or directory: 'output_classified.json'
>>> ===== RESTART =====
>>>
>>> ===== RESTART =====
>>>
>>> ===== RESTART =====
>>>
>>>
>>>

Ln: 182 Col: 4

```

Next steps

- Send classified tweet information to mySQL database and associate with user
- Build public facing web map application that will allow users to modify ML classifications if necessary

Questions?

Thanks!

mariogiampieri@gmail.com

