****



**NGA Disparate Data Challenge**

**Stage 1 Report**

**19 September 2016**

Table of Contents

[Demonstration Synopsys 4](#_30j0zll)

[Background Information 4](#_1fob9te)

[Safe Software 4](#_3znysh7)

[Spatial Business Systems 4](#_44sinio)

[FME - Feature Manipulation Engine 4](#_2et92p0)

[NGA Disparate Data Challenge 5](#_tyjcwt)

[Key Points for FME 5](#_3dy6vkm)

[Viewing Data with FME 5](#_2jxsxqh)

[Workflows 10](#_1t3h5sf)

[World Food Program Metadata Loader 11](#_4d34og8)

[Navy BlueMarble - low res WMS Metadata Loader 12](#_2s8eyo1)

[Sample NITF Data Metadata Loader 13](#_17dp8vu)

[Large Data Metadata Loader 14](#_3rdcrjn)

[Metadata Extract 15](#_26in1rg)

[Additional Work 16](#_lnxbz9)

[Summary 16](#_35nkun2)

# Demonstration Synopsys

Our submission focuses on the extraction of metadata and loading it into a queryable database. The submission is based on the FME product from Safe Software, which is a commercial off-the-shelf (COTS) product, and involves no custom coding.

# Background Information

## Safe Software

Safe Software is the industry leader in spatial ETL (extract, transform and load) and the makers of FME, which delivers the most extensive format support for data translation and integration and provides unlimited flexibility in data model transformation and data distribution. Thousands of customers worldwide use FME’s powerful technology to help them solve their spatial data access challenges.

## Spatial Business Systems

SBS provides software and service solutions that help their clients streamline their business processes, reduce costs, and maximize their investment in CAD, GIS and other information systems. SBS is a Gold Partner of Safe Software and a full service Solution provider spanning from spatial data integration services to strategic consulting and training.

## FME - Feature Manipulation Engine

Safe Software manufactures three primary products:

[**FME Desktop**](http://www.safe.com/fme/fme-desktop/)– Authoring tool to transform, load, integrate, validate, and share data

[**FME Server**](http://www.safe.com/fme/fme-server/) – Enterprise automation product that leverages FME Desktop

[**FME Cloud**](http://www.safe.com/fme/fme-cloud/) – AWS-hosted version of FME Server

# NGA Disparate Data Challenge

Our goal for the data challenge submission is to show that the FME product is capable of extracting the relevant metadata and loading it into an appropriate database.

## Key Points for FME

FME was selected for this challenge because it:

* Can easily view data using FME Data Inspector (see Figure 1)
* Is a no-coding-required COTS product with an intuitive GUI for creating workflows
* Supports all data formats provided within the NGA disparate dataset
* Has the ability to run the workflows from a web GUI via FME Server

## 

## Workflows

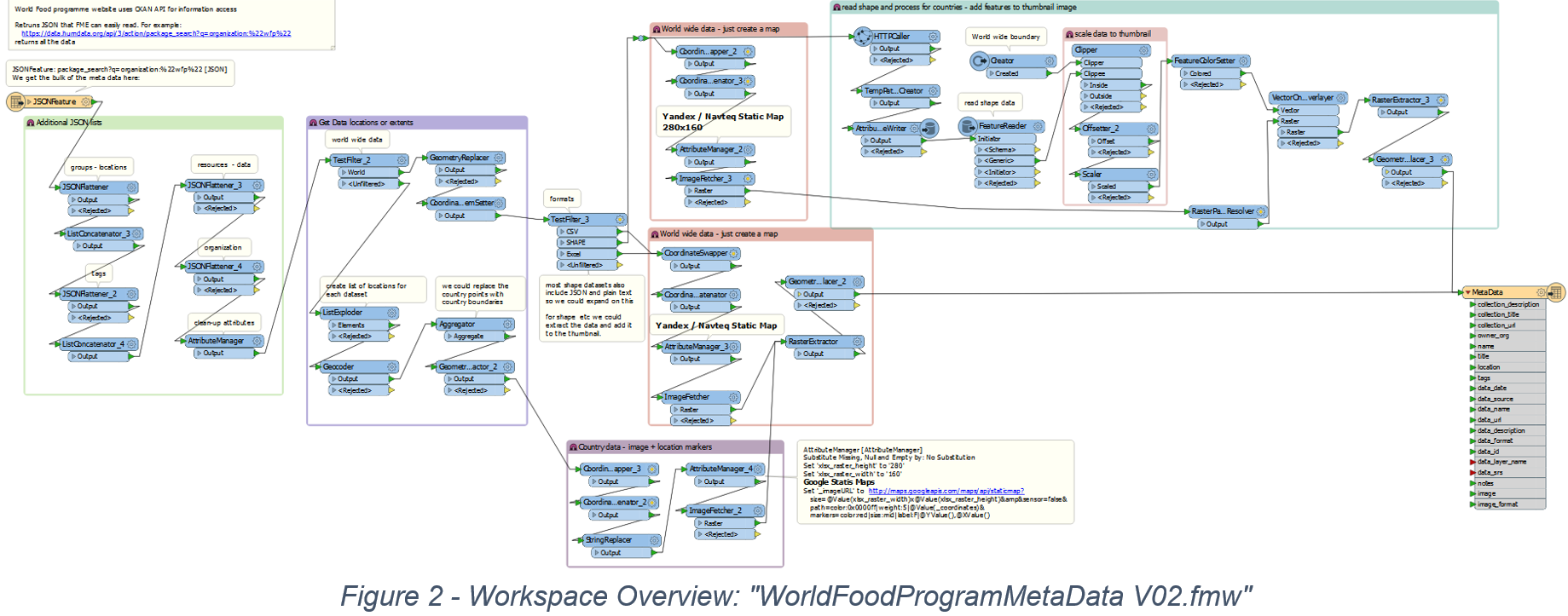
The following describes the workflows created for this submission. The primary workflows read the metadata from the Data Challenge websites and load it into a PostGIS database. The target database could be any FME-supported database, as listed [here](http://www.safe.com/fme/format-search/#!).

The metadata loader workflows illustrate FME’s flexibility in reading metadata using either an API, WMF, scraping the HTML or direct data link.

We have created workflows for four datasets from the data challenge [website](http://disparatedata.s3-website-us-east-1.amazonaws.com/):

* [World\_Food\_Program](https://data.humdata.org/organization/wfp)
* [Navy\_Blue\_Marble\_-\_low\_res](http://geoint.nrlssc.navy.mil/nrltileserver/wms)
* [Sample\_NITF\_Data](http://tinyurl.com/joo8cmu)
* Large\_Data

### World Food Program Metadata Loader



This workflow loads metadata from the WFP website:

<https://data.humdata.org/organization/wfp>

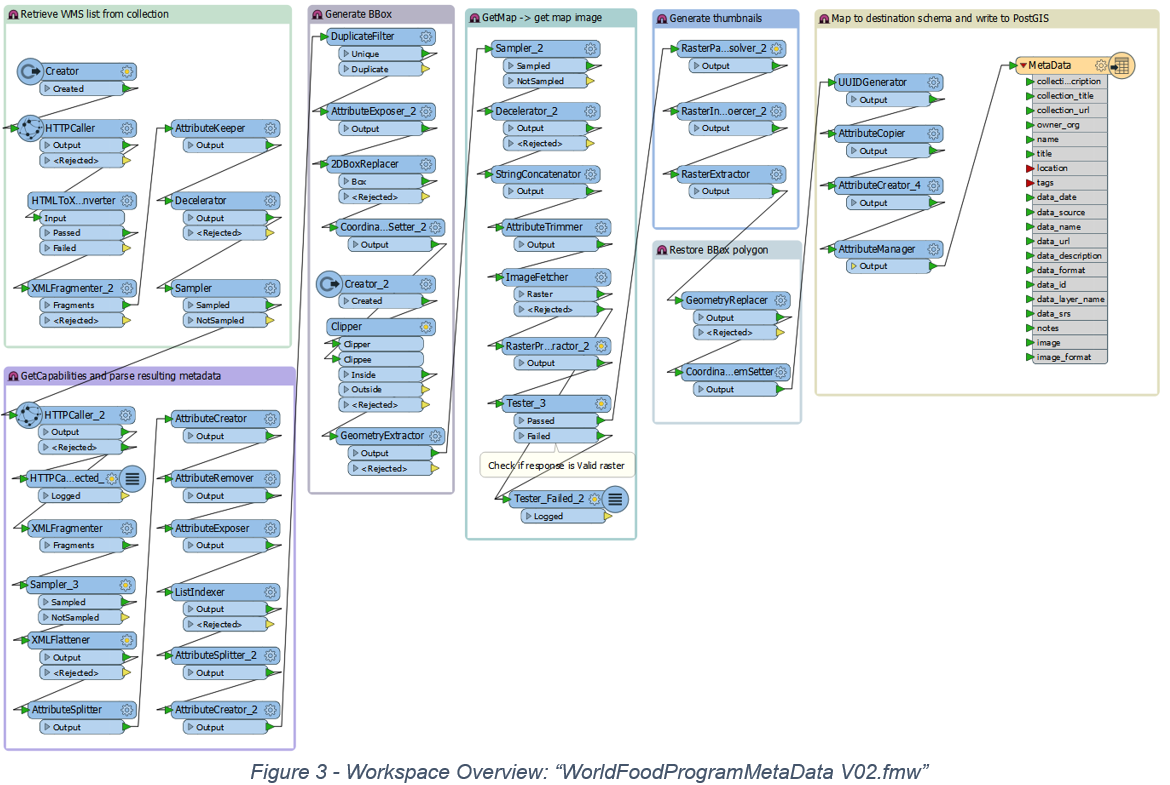
Using the FME JSON reader, and leveraging the WFP Website API, the workflow requests the metadata and processes the resulting JSON as follows:

* Identifies locations for each dataset
* Extracts tags
* Retrieves the data URL for downloadable resources
* Extracts organization details

Additionally, the workspace “WorldFoodProgramMetaData V02.fmw”:

* Creates a thumbnail
* Adds country markers or spatial data to the thumbnail depending on source data format
* Adds data extents to the PostGIS database so the data can be extracted using a spatial query

### Navy BlueMarble – Low-Res WMS Metadata Loader



This workflow loads metadata from the Blue Marble WMS service, which contains links to the US Navy’s NRL's WMS Tileserver:

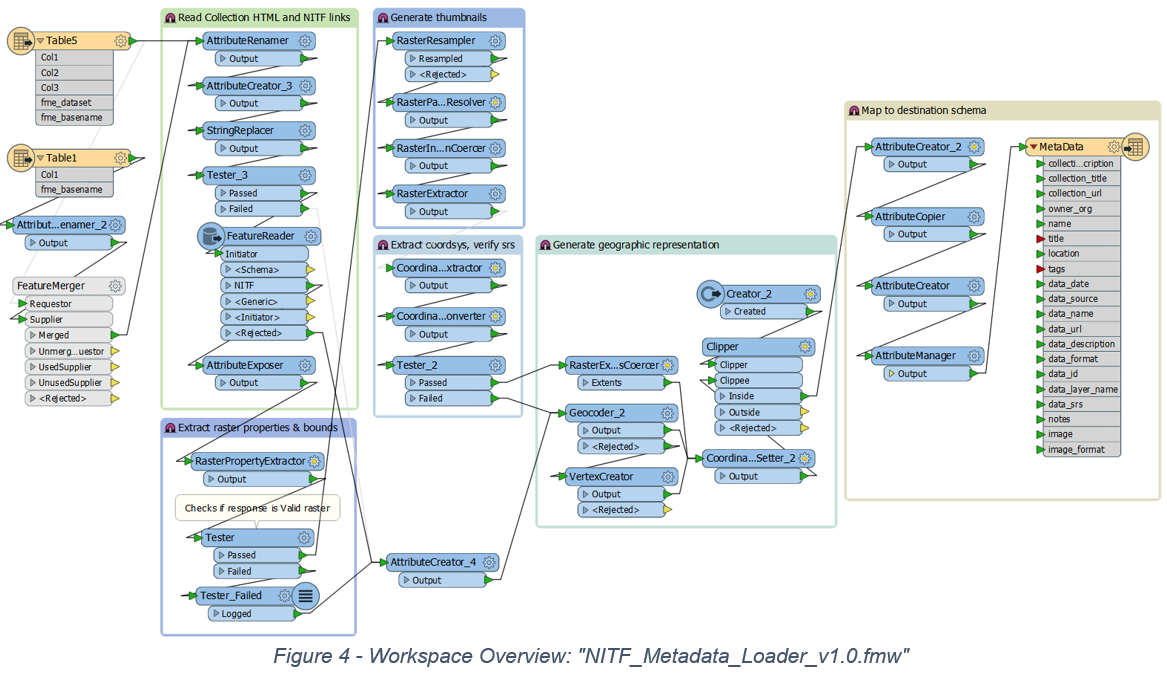
<http://geoint.nrlssc.navy.mil/nrltileserver/wms>

The page consists of 35 WMS getCapabilities links.

NGA\_nrlssc.navy.WMS\_v1.0.fmw harvests metadata from the listed WMS servers, as follows:

1. Reads the web page with HTTPCaller to generate a list of WMS URL’s.
2. Composes a GetCapabilities request to get a list of available layers.
3. Composes a GetMap request to get the overview thumbnail images.
4. Writes a record representing each layer's metadata to the metadata table.

### Sample NITF Data Metadata Loader



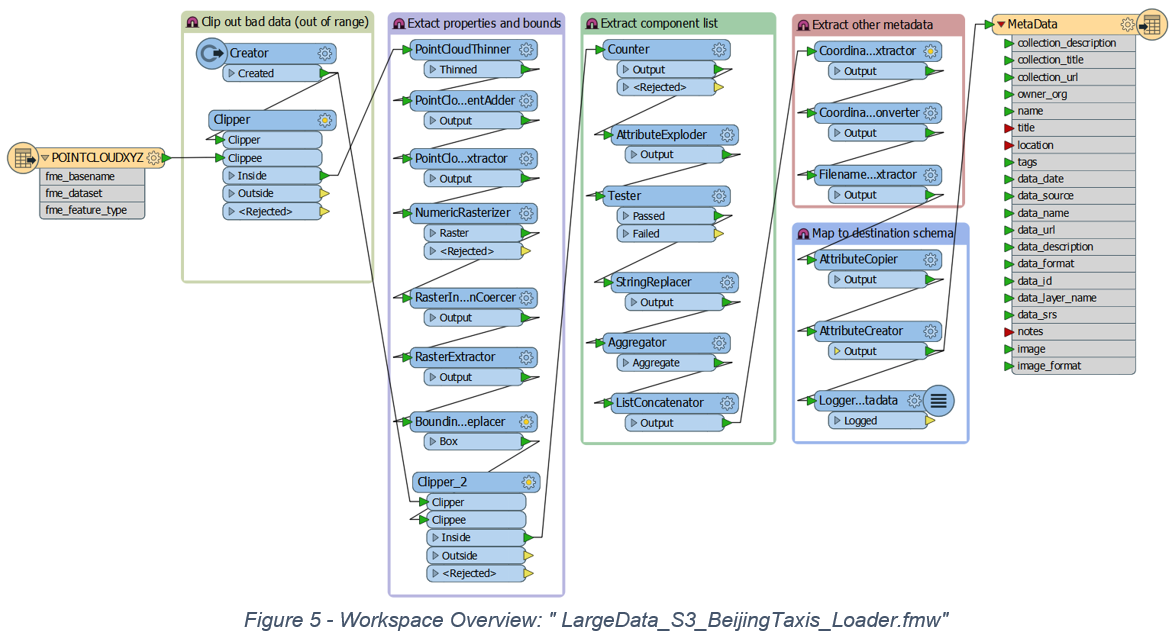
This workflow loads metadata from the Sample NITF data link:

<http://www.gwg.nga.mil/ntb/baseline/software/testfile/Nitfv2_0/scen_2_0.html>

NITF\_Metadata\_Loader\_v1.0.fmw workflow:

1. Reads NGA’s NITF Sample Data webpage using FME’s HTML reader
2. Generates URLs to the individual NITF files
3. Uses FeatureReader to read the NITF datasets
4. Generates thumbnails for each NITF dataset
5. Composes metadata and writes to the database

### Large Data Metadata Loader



This workflow loads metadata from the S3-hosted NGA Large data store:

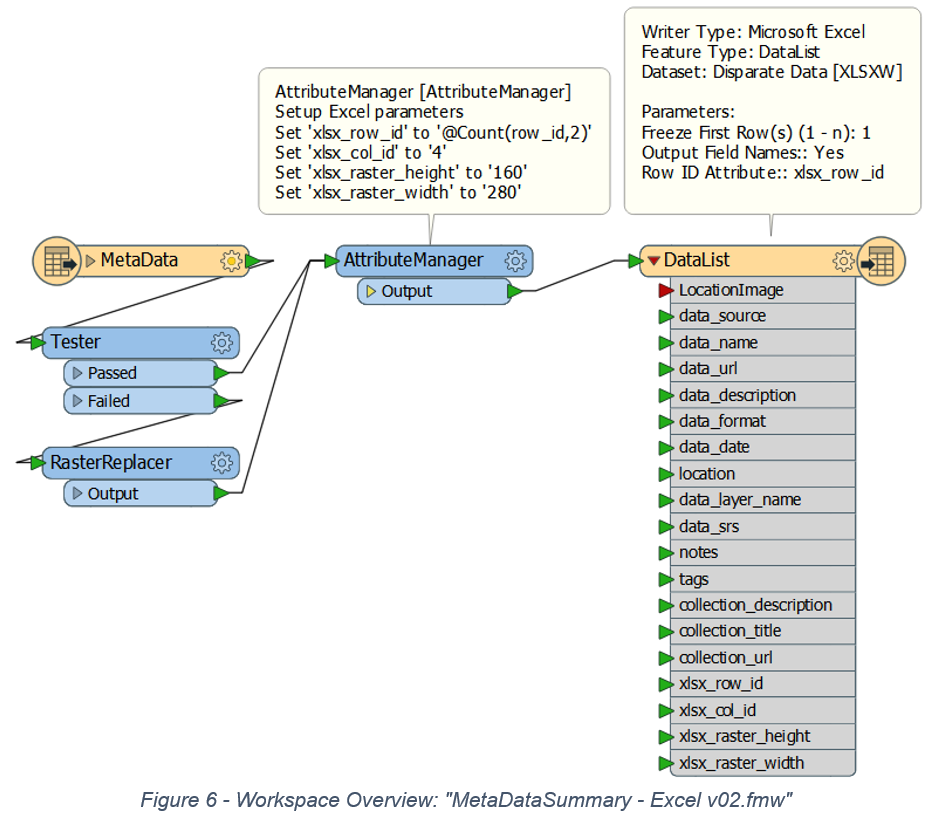
<https://s3.amazonaws.com/data-challenge-june2016/RawData/Large+Data/taxi_beijing.csv>

For efficiency, FME uses the point cloud reader to read the dataset from the csv file into one array feature. This workflow demonstrates how FME can handle large datasets and work with AWS.

LargeData\_S3\_BeijingTaxis\_Loader.fmw workflow:

1. Read the csv with the Point Cloud XYZ reader
2. Clip the data to remove extreme values
3. Extract point cloud properties
4. Convert to raster for thumbnail
5. Generates metadata including spatial extents, srs, and time range

### Metadata Extract



“MetaDataSummary - Excel v02.fmw” allows the user to send a spatial query to the PostGIS database and extracts a metadata report as an Excel spreadsheet.

The REST API can be used if the reporting workflow is to be added to a webpage or other app.

This workflow could also be extended to allow:

* Different outputs (e.g. HTML report)
* Data selection by country, tag, etc.
* Reading metadata from any of the FME-supported RDBMS or NoSQL databases.
* Integration into a web map environment to include graphical/attribute queries. See this [example](http://demos.fmeserver.com/datadistribution-webmap/index.html?map=google).

## Additional Work

The examples included in this submission can be extended in a variety of ways, such as:

* Extract layers and attributes from data.
* Produce cleaner thumbnails.
* Extract country borders from OpenStreetMap or equivalent, to add to thumbnail.
* Use FME Server’s scheduler for automation.
* Add tasks that check for data changes and reload metadata for the changed data.
* Implement unique dataset ID to allow metadata updates for single datasets.
* Utilize the scalability of FME Server to improve data processing and performance.
* Provide a data streaming / download geoportal and REST API to search for and retrieve needed data. FME Server provides data streaming / download services with minimal configuration and the ability to manage request traffic and data volumes.
* Extend the metadata workflows to download data via download geoportal, rather than reading each disparate dataset provided.
* Explore integration with other systems and web pages.

# Summary

We have deployed a system that extracts the metadata from various disparate data websites and loads it into a queryable database. The entire implementation uses FME technology and requires no custom code.

# Appendix A. How to Access the Demo

The above workflows can be run using the following URL on the FME Cloud server that is running on Amazon AWS.

<https://disparatedatachallenge-safe-software-inc.fmecloud.com>

User: NGA

Password: fmenga

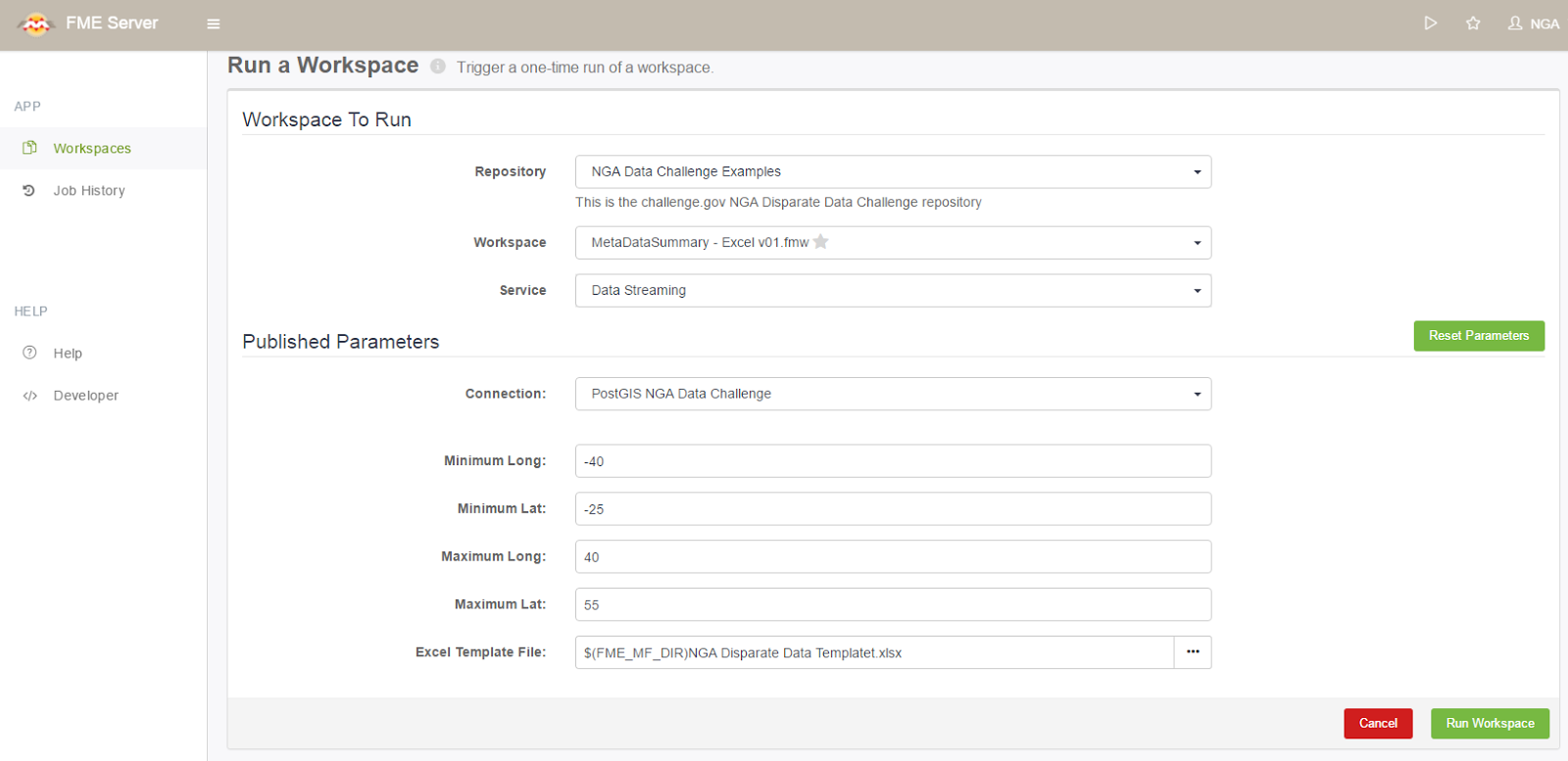
The following describes how to run the NGA Data Challenge Examples on FME Cloud.

## Metadata Extract

This workflow can be run using the following URL:

<https://disparatedatachallenge-safe-software-inc.fmecloud.com/fmeserver/#/workspaces/run/NGA%20Data%20Challenge%20Examples/MetaDataSummary%20-%20Excel%20v01.fmw/fmedatastreaming>

Clicking on this link will bring up the login page. After logging in you should see the workspace run form shown below:



Simply enter the extents for the area of interest, and the Metadata Extract will generate an Excel file containing a table of metadata records and thumbnails for the available datasets for that area. The output can then be reviewed from within Microsoft Excel.

## Metadata Loading Workflow

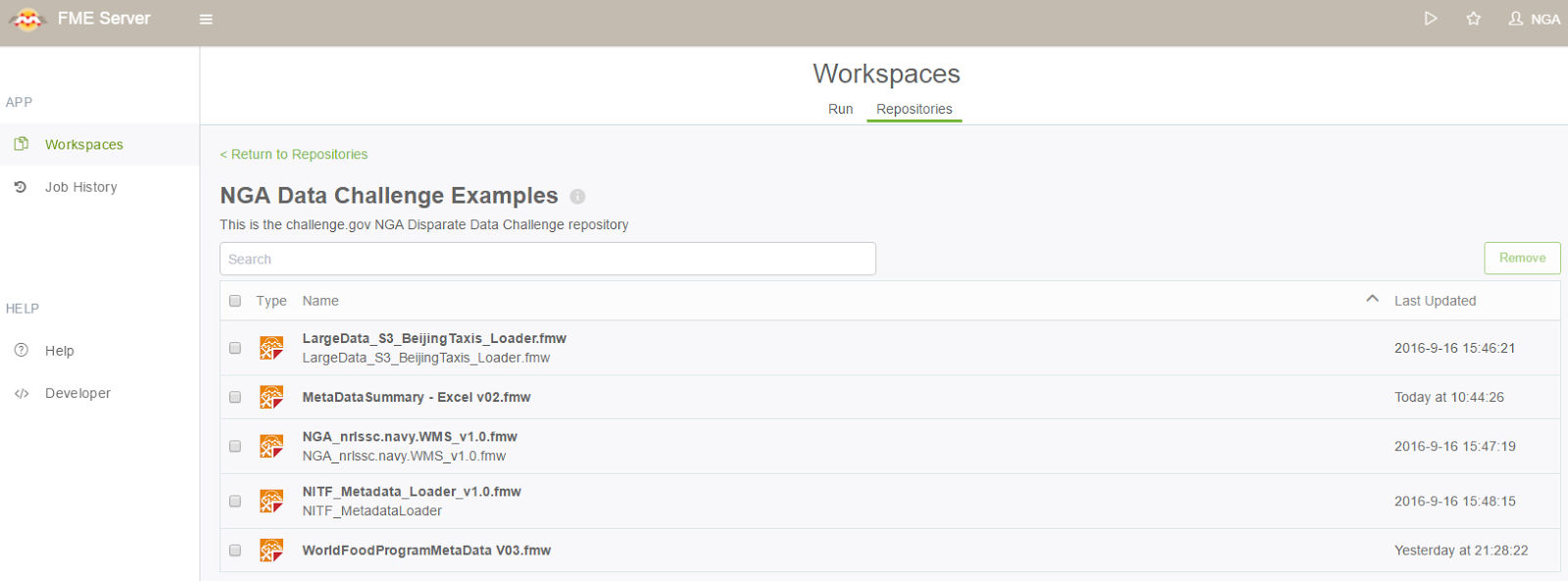
The other workflows simply read from selected NGA disparate data sources provided, generate metadata, and load the results into PostGIS. Running these will overwrite the existing data in PostGIS. To reload all the metadata, run the loaders in the following order:

1. WorldFoodProgramMetaData V02.fmw first as it gives you the option to truncate and reload.
2. NGA\_nrlssc.navy.WMS\_v1.0.fmw
3. NITF\_Metadata\_Loader\_v1.0.fmw
4. LargeData\_S3\_BeijingTaxis\_Loader.fmw

These workspaces can be run from the following url using the same login above:

<https://disparatedatachallenge-safe-software-inc.fmecloud.com/fmeserver/#/workspaces/repo/NGA%20Data%20Challenge%20Examples>

The NGA Data Challenge Examples repository contains all the workflows (FME workspaces) associated with this submission as shown below.



FME Server repository for NGA Data Challenge Examples workflows.

### **Data Inspector Demo of FME Reading NGA OGC WMS and WFS Services**

The following link demonstrates reading OGC services via FME’s Data Inspector:

<https://www.youtube.com/watch?v=1AY_76HELG8&feature=youtu.be>

The Blue Marble Navy Tile service WMS is added as a background map and then the 2 provided NGA Foreign GeoNames WFS services are added in the foreground. The demo also shows FME rendering the GML feature elements according to the GeoNames and Gazetteer Profile (WFS-G) application schemas.