Setup for a Generic Linked Data Platform

Christophe Debruyne, Kris McGlinn, ADAPT @ Trinity College Dublin

# Introduction

This document aims to provide the reader with a step-by-step tutorial for creating a generic Linked Data platform from the uplift process – generating RDF from non-RDF sources – to connecting a Linked Data frontend with content negotiation.

# Choosing an appropriate URI

Remember that one needs to provide HTTP URIs to retrieve useful information. So the first step is to think about a domain that will serve as the base URI for your resource identifiers. For this tutorial, we will assume that the base URI for all resources will be <http://data.example.org/>.

# Short BAckground Information

The data was generated to enable a tool which supports the Facility Manager (FM) in the task of comparing predicted (simulated) energy consumption against measured (real) energy consumption in a building.

This required the following tasks to be completed:

* The FM had to install PlugWise meters at plugs sources. These monitored the energy consumption of the devices plugged into those sockets.
* An energy modeler modeled the building in an energy simulation tool and configured the simulator to generate simulated energy consumption for the different rooms.
* A model of the building geometry to exist, which includes the position of rooms and sensors, so that the building may be visualized and sensors with sensor measurements made selectable.

PlugWise provides a webapp to access and monitor the PlugWise meters. PlugWise does not have an open API. On a monthly basis the FM must log into a web app and export the months energy readings into a CSV file so that they can be accessed by external tools. Also, as there are a number of different energy modelling tools, the developers of the monitoring tool do not wish to develop multiple sets of interfaces to connect to these different tools.

As such, the energy modeler configures the energy simulator to output the simulated energy consumption as a CSV file for the zone in question. **The model of the building geometry is generated elsewhere, and will not be examined in this tutorial.**

# STEP 1: Analysing the Data

We will annotate two CSV files with R2RML[[1]](#footnote-1) to generate RDF with an appropriate processor. Let’s use R2RML to describe the tabular data and provide a means to uplift it into RDF, adding semantics and supporting linking data to other RDF models.

First, let us look at the simulated data. In this case we see it is a file called ‘forum\_room2\_TAS.csv’ (https://www.scss.tcd.ie/~mcglink/lectures/CS7012/tutorials/r2rml/csv/ forum\_room2\_TAS.csv)

There are seven columns in the file.

* First gives a unique ID.
* Second a date, third a day, fourth a time.
* We then have lighting gains based on three models of usage. The first, 100% on during working hours. The second, based upon occupancy rate. The third, when autodimming (dims based upon solar light) is in effect. These each need their own column.

Now, let us examine the second PlugWise file (https://www.scss.tcd.ie/~mcglink/lectures/CS7012/tutorials/r2rml/csv/ forum\_room2\_plugwise.csv).

* We can use the 2nd and 4th columns as per the simulation data.
* We also have one recording data from the rooms lighting.
* Finally, we have the outputs in kWh.

# Annotating CSV FILES with R2RML

Looking at the examples provided by the R2RML W3C Recommendation, create a mapping that

* Uses SSN (you can use ns “ssn”: https://www.w3.org/ns/ssn/ )
  + <https://www.w3.org/TR/vocab-ssn/>
* Uses time (you can use ns “xsd” for date and time )
* Declares a TriplesMap for the CSV table “forum\_room2\_tas”
  + Each record is an instance of ssn:Observation
  + Each record has a xsd:date and a xsd:time
* Declares a TriplesMap for the CSV table “forum\_room2\_plugwise”
  + Each record is an instance of ssn:Observation

Each record has a xsd:date and a xsd:time.

The implementation we are using[[2]](#footnote-2) has a configuration file and provides support for named graphs, which is not the case for db2triples. If you don’t want to install maven, you can use the following version[[3]](#footnote-3). A configuration file could look as follows:

connectionURL = ./forum\_room2\_TAS.csv

user = root

password = XXX

mappingFile = mapping.ttl

outputFile = output.ttl

filePerGraph = true

And the RDF is generated with the following command:

$ java -jar r2rml-0.0.1-SNAPSHOT.jar config.properties

Though I urge you to try to create your own mapping, the examples in the recommendation should be sufficient, a mapping is provided in Appendix 3. The RDF generated with that mapping is provided in Appendix 4.

# Setting up the triplestore and SPARQL endpoint

Setting up a triplestore and SPARQL endpoint is straightforward. May suites exist:

* Stardog – <http://stardog.com/> (free for non-commercial purposes)
* Virtuoso – <https://github.com/openlink/virtuoso-opensource>
* Parliament – <http://parliament.semwebcentral.org/>
* Apache Jena and Fuseki – <https://jena.apache.org/documentation/serving_data/>
* …

For this tutorial, we are going to use Apache Jena and Fuseki. If you wish to quickly try out a SPARQL endpoint, you can quickly start a Fuseki instance using the batch file fuseki-server.bat:

Fuseki runs by default on localhost:3030. You can now add the triples manually in the dataset with the interface provided (see below). You can choose whether to make it persistent or in memory.

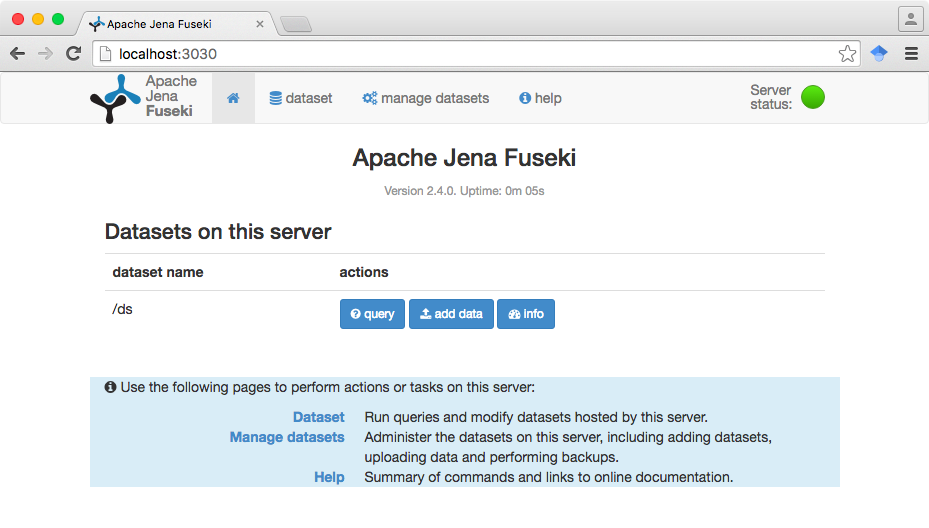


Figure Click on “add data” to manually load triples in the dataset “ds” that we have created in the terminal.

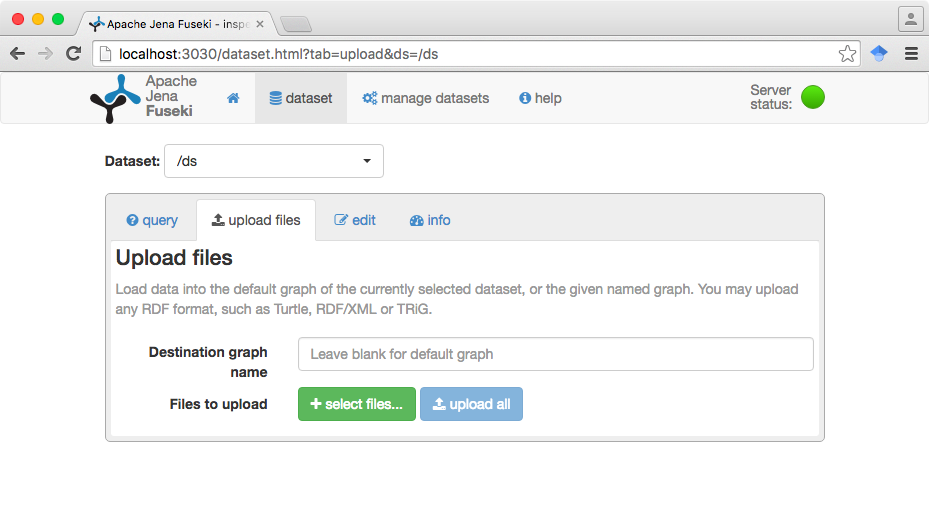


Figure You can upload one or more files and even specify the graphs in which the triples have to be loaded.

With the triples loaded, you can start playing with SPARQL queries. In the next tutorial, we we look at this in more detail and you will create a small JavaScript web app to display the data in a chart.

# Appendix 1: CSV DATA DUMP TAS SIMULATOR

ID,DATE,DAY,TIME,R2LIGHT100,R2LIGHTOCCUPANCY,R2LIGHTAUTODIMMING

1,20/08/2012,Mon,00:00,0,0,0

2,20/08/2012,Mon,01:00,0,0,0

3,20/08/2012,Mon,02:00,0,0,0

4,20/08/2012,Mon,03:00,0,0,0

5,20/08/2012,Mon,04:00,0,0,0

6,20/08/2012,Mon,05:00,0,0,0

7,20/08/2012,Mon,06:00,0,0,0

8,20/08/2012,Mon,07:00,173.17,43.29,38.91

9,20/08/2012,Mon,08:00,173.17,86.58,38.91

10,20/08/2012,Mon,09:00,173.17,173.17,38.91

11,20/08/2012,Mon,10:00,173.17,173.17,38.91

12,20/08/2012,Mon,11:00,173.17,173.17,38.91

13,20/08/2012,Mon,12:00,173.17,129.88,38.91

14,20/08/2012,Mon,13:00,173.17,129.88,38.91

15,20/08/2012,Mon,14:00,173.17,173.17,59.58

16,20/08/2012,Mon,15:00,173.17,173.17,38.91

17,20/08/2012,Mon,16:00,173.17,173.17,122.87

18,20/08/2012,Mon,17:00,173.17,86.58,145.34

19,20/08/2012,Mon,18:00,173.17,43.29,164.45

20,20/08/2012,Mon,19:00,0,0,0

21,20/08/2012,Mon,20:00,0,0,0

22,20/08/2012,Mon,21:00,0,0,0

23,20/08/2012,Mon,22:00,0,0,0

24,20/08/2012,Mon,23:00,0,0,0

# Appendix 2: CSV DATA DUMP PLUGWISE

ID,DATE,DAY,TIME,NAME,ROOM,TYPE,KWH

1,20/08/2012,Mon,00:00,PC,2,lighting,0

2,20/08/2012,Mon,01:00,PC,2,lighting,0

3,20/08/2012,Mon,02:00,PC,2,lighting,0

4,20/08/2012,Mon,03:00,PC,2,lighting,0

5,20/08/2012,Mon,04:00,PC,2,lighting,0

6,20/08/2012,Mon,05:00,PC,2,lighting,0

7,20/08/2012,Mon,06:00,PC,2,lighting,0

8,20/08/2012,Mon,07:00,PC,2,lighting,0

9,20/08/2012,Mon,08:00,PC,2,lighting,0.0411

10,20/08/2012,Mon,09:00,PC,2,lighting,0.0578

11,20/08/2012,Mon,10:00,PC,2,lighting,0.0577

12,20/08/2012,Mon,11:00,PC,2,lighting,0.0579

13,20/08/2012,Mon,12:00,PC,2,lighting,0.0575

14,20/08/2012,Mon,13:00,PC,2,lighting,0.0564

15,20/08/2012,Mon,14:00,PC,2,lighting,0.0564

16,20/08/2012,Mon,15:00,PC,2,lighting,0

17,20/08/2012,Mon,16:00,PC,2,lighting,0

18,20/08/2012,Mon,17:00,PC,2,lighting,0

19,20/08/2012,Mon,18:00,PC,2,lighting,0

20,20/08/2012,Mon,19:00,PC,2,lighting,0

21,20/08/2012,Mon,20:00,PC,2,lighting,0

22,20/08/2012,Mon,21:00,PC,2,lighting,0

23,20/08/2012,Mon,22:00,PC,2,lighting,0

24,20/08/2012,Mon,23:00,PC,2,lighting,0

# Appendix 3: Example of the mapping file FOR PLUGWISE

@prefix rr: <http://www.w3.org/ns/r2rml#> .

@prefix ex: <http://data.example.com/plugwise/#> .

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

@prefix ssn: <http://www.w3.org/ns/ssn/>.

@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.

<#TriplesMap1>

rr:logicalTable [ rr:tableName "forum\_room2\_plugwise" ];

rr:subjectMap [

rr:template "http://data.example.com/plugwise/{ID}";

rr:class ssn:Observation;

];

rr:predicateObjectMap [

rr:predicate xsd:date;

rr:objectMap [ rr:column "DATE"; rr:datatype xsd:date ];

];

rr:predicateObjectMap [

rr:predicate ex:day;

rr:objectMap [ rr:column "DAY" ];

];

rr:predicateObjectMap [

rr:predicate ex:time;

rr:objectMap [ rr:column "TIME"; rr:datatype xsd:time ];

];

rr:predicateObjectMap [

rr:predicate ex:name;

rr:objectMap [ rr:column "NAME" ];

];

rr:predicateObjectMap [

rr:predicate ex:room\_number;

rr:objectMap [ rr:column "ROOM" ];

];

rr:predicateObjectMap [

rr:predicate ex:type;

rr:objectMap [ rr:column "TYPE" ];

];

rr:predicateObjectMap [

rr:predicate ssn:hasValue;

rr:objectMap [ rr:column "KWH" ];

]

.

# APPENDIX 4: GENERATED RDF EXAMPLE (PLUGWISE)

<http://data.example.com/plugwise/64>

a <http://www.w3.org/ns/ssn/Observation> ;

<http://data.example.com/plugwise/#day>

"Wed" ;

<http://data.example.com/plugwise/#name>

"PC" ;

<http://data.example.com/plugwise/#room\_number>

"2" ;

<http://data.example.com/plugwise/#type>

"lighting" ;

<http://www.w3.org/2001/XMLSchema#date>

"22/08/2012" ;

<http://www.w3.org/2001/XMLSchema#time>

"15:00" ;

<http://www.w3.org/ns/ssn/observationValue>

"0" .

<http://data.example.com/plugwise/36>

a <http://www.w3.org/ns/ssn/Observation> ;

<http://data.example.com/plugwise/#day>

"Tue" ;

<http://data.example.com/plugwise/#name>

"PC" ;

<http://data.example.com/plugwise/#room\_number>

"2" ;

<http://data.example.com/plugwise/#type>

"lighting" ;

<http://www.w3.org/2001/XMLSchema#date>

"21/08/2012" ;

<http://www.w3.org/2001/XMLSchema#time>

"11:00" ;

<http://www.w3.org/ns/ssn/observationValue>

"0.0564" .

<http://data.example.com/plugwise/77>

a <http://www.w3.org/ns/ssn/Observation> ;

<http://data.example.com/plugwise/#day>

"Thu" ;

<http://data.example.com/plugwise/#name>

"PC" ;

<http://data.example.com/plugwise/#room\_number>

"2" ;

<http://data.example.com/plugwise/#type>

"lighting" ;

<http://www.w3.org/2001/XMLSchema#date>

"23/08/2012" ;

<http://www.w3.org/2001/XMLSchema#time>

"04:00" ;

<http://www.w3.org/ns/ssn/observationValue>

"0" .

<http://data.example.com/plugwise/49>

a <http://www.w3.org/ns/ssn/Observation> ;

<http://data.example.com/plugwise/#day>

"Wed" ;

<http://data.example.com/plugwise/#name>

"PC" ;

<http://data.example.com/plugwise/#room\_number>

"2" ;

<http://data.example.com/plugwise/#type>

"lighting" ;

<http://www.w3.org/2001/XMLSchema#date>

"22/08/2012" ;

<http://www.w3.org/2001/XMLSchema#time>

"00:00" ;

<http://www.w3.org/ns/ssn/observationValue>

"0" .

<http://data.example.com/plugwise/5>

a <http://www.w3.org/ns/ssn/Observation> ;

<http://data.example.com/plugwise/#day>

"Mon" ;

<http://data.example.com/plugwise/#name>

"PC" ;

<http://data.example.com/plugwise/#room\_number>

"2" ;

<http://data.example.com/plugwise/#type>

"lighting" ;

<http://www.w3.org/2001/XMLSchema#date>

"20/08/2012" ;

<http://www.w3.org/2001/XMLSchema#time>

"04:00" ;

<http://www.w3.org/ns/ssn/observationValue>

"0" .

1. <https://www.w3.org/TR/r2rml/> [↑](#footnote-ref-1)
2. <https://opengogs.adaptcentre.ie/debruync/r2rml/> [↑](#footnote-ref-2)
3. <https://www.dropbox.com/sh/3plas6sfjs3kcoh/AABgMCXoHIjJt-JrmYJ1eyU4a?dl=0>  [↑](#footnote-ref-3)