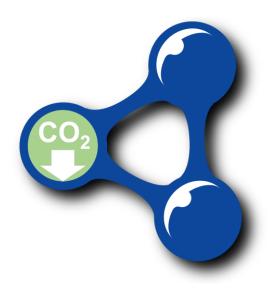
Building linked open data about carbon savings

Table of Contents

1. Measurement data → CSV	2
1.1. Sources	2
1.2. Deriving the same information from the dissimilar sources	2
1.3. The process.	3
1.4. Design decisions	3
1.5. Per-source walk-throughs	4
1.5.1. Alloa Community Enterprises - reused furniture	4
1.5.2. Stirling Community Food - redistributed food	5
1.5.3. Stirling Council - household recycling bins	7
1.5.4. The Fair Share - reused, student-oriented items	8
2. Reference data → CSVs	9
2.1. Carbon metric	9
2.2. Enabler organisations	11
3. CSVs → linked data	11
3.1. CSVs may be good enough	11
3.2. Describing our data using linked data vocabularies	11
3.3. Using CSVW	11
3.3.1. CSVW for the carbon-metric reference data	12
3.3.2. CSVW for the enablers reference data	13
3.3.3. CSVW for the carbon-savings measurement data	13
3.4. Generating an RDF graph	15
4. Using the linked data	15
5. Next?	17
References	18

DRAFT

An incomplete, early draft - still lots of TODOs.



This is a *walk-through* of how we can build **linked open data** (LoD) about **carbon savings** from dissimilar data sources.

1. Measurement data → CSV

1.1. Sources

We take examples of meaurement data from 4 sources:

- Alloa Community Enterprises reused furniture
- Stirling Community Food redistributed food
- Stirling Council household recycling bins
- The Fair Share reused, student-oriented items.

1.2. Deriving the same information from the dissimilar sources

Their data is dissimilar, for example:

- 3rd sector reporting vs. local government reporting
- emailed Excel files vs. CKAN hosted CSV files
- kilograms vs. tonnes
- · various means of indicating dates
- various ways of categorising the stuff that gets reused.

...but our end goal is to derive the same information from each of them. This information can be seen in the CSV file that contains the results of our efforts in this section: carbon-savings.csv. Here's

a snippet of data from that file:

occurrence-date	scottish-carbon- metric-material	tonnes-weight	enabler	enabler-specific
2018-02-28	Wood	0.385	Alloa Community Enterprises	55 items of Furniture (Chair, Kitchen, Dining or Wooden)
2021-01-29	Food and Drink Waste (wet AD)	0.28012	Stirling Comunity Food	Used for human- food, bio-etc & sanctuary
2021-01-03	Glass (mixed colours)	3.5	Stirling Council	Balfron
2019-12-31	Textiles & Footwear	0.57644	The Fair Share	autumn semester

1.3. The process

To derive that CSV file, we do the following for each source:

- a. parse the essential values out of the source files
- b. derive and standardise values
 - i. mapping text values into controlled, canonical *code lists* (e.g. the Scottish Carbon Metric's list of materials; and our list of carbon-savings enabler organisations)
 - ii. appropriately scaling numbers (e.g. weight amounts to tonnes)
 - iii. applying a standard format to dates
 - iv. encode source-specific data that would be useful to propogate, into the enabler-specific field
- c. output as CSV.



Our process is basically: messy data → *tidy* CSV data.

The Tidy Data paper provides good rationale for and examples of this process, generalised.

1.4. Design decisions

- Use The Scottish Carbon Metric as the basis for measuring carbon savings. This is referred to via the scottish-carbon-metric-material column; and we will discuss it further in section 2.
- Accounting & reporting procedures often batch into a single carbon-savings record, all of the instances of *same material-class* reuse/recycling that have occurred within an interval of time. The occurrence-date column captures the end date of those intervals.
- Include the enabler-specific column to allow enablers to: propagate additional information;

and to report measurements at an additional level of granularity (see the primary key explanation, later).

• Reduce "double accounting" by trying to ensure that: the data contains no records where one carbon savings enabler 'feeds' the same reuse/recycling item to another enabler, within some (significant-for-use) duration.

1.5. Per-source walk-throughs

In this section we walk-through the *measurement data* \rightarrow *CSV* process, for each source; outlining some per-source specifics.

1.5.1. Alloa Community Enterprises - reused furniture



Alloa Community Enterprises' (ACE) furniture reuse initiative has been running since 1984, helping prevent furniture from becoming waste.

ACE is in the process of publishing its data as open data. The following examples of measurement data have been taken from a **draft** of that work.

Here's a snipped image of ACE's measurement data:

	Α	В	С	D	F
	Category	Items			
			Average	Feb 18	Feb 19
1			Weight	nos	nos
7	Furniture	Chair, Kitchen, Dining or Wooden	7	55	82
46	Soft Furniture	Mattress, single	21	7	8

The main steps in processing ACE's measurement data (the snippet shown above) are:

- a. Parse the essential values out of columns A-D & F
- b. Derive and standardise values ...
 - i. We map each pair of Category & Items values, to a scottish-carbon-metric-material value. We build this mapping table as a CSV file: ace-to-carbon-metric.csv. Here's a snippet of data from that file:

category	item	material
Furniture	Chair, Kitchen, Dining or Wooden	Wood
Soft Furniture	Mattress, single	Textiles and Footwear

- ii. Multiple columns D & F by column C and divide by 1000, to calculate the tonnes-weight value.
- iii. Map the headers of columns D & F, to establish the occurrence-date value.
- iv. Capture furniture type & count information in the enabler-specific value.
- c. Output the end result as the CSV rows:

occurrence-date	scottish-carbon- metric-material	tonnes-weight	enabler	enabler-specific
2018-02-28	Wood	0.385	Alloa Community Enterprises	55 items of Furniture (Chair, Kitchen, Dining or Wooden)
2018-02-28	Wood	0.574	Alloa Community Enterprises	82 items of Furniture (Chair, Kitchen, Dining or Wooden)
2019-02-28	Textiles and Footwear	0.147	Alloa Community Enterprises	7 items of Soft Furniture (Mattress, single)
2019-02-28	Textiles and Footwear	0.168	Alloa Community Enterprises	8 items of Soft Furniture (Mattress, single)

1.5.2. Stirling Community Food - redistributed food



Stirling Community Food is a project that helps to reduce food waste in Stirling by collecting (from supermarkets & aggregators) excess and near-sell-by-date food, routing it away from waste bins, and making it available (for free) to the community.

Stirling Community Food is in the process of publishing its data as open data. The following

examples of measurement data have been taken from a **draft** of that work.

Here's a snipped image of Stirling Community Food's measurement data:

	А	В	С	D	G	K	L	М	N	0
2					Recei					Donated to
3	Da	te	Neighbourly	Fareshare	Cooperative	Other	Grand Total	Waste (Kg)	Composted (Kg)	animal sanctuary (Kg)
311	Thursday	28-Jan	52.7		17	18.39	88.1	2.85	0.48	54.65
312	Friday	29-Jan	32	255.52		0.6	288.1	12.5	8	

The main steps in processing Stirling Community Food's measurement data (the snippet shown above) are:

- a. Parse the essential values out of columns A-0
- b. Derive and standardise values ...
 - i. We map the 'outcomes' (i.e. how the food material got used) to a scottish-carbon-metric-material value. We build this mapping table as a CSV file: stirling-community-food-to-carbon-metric.csv. Here's the data from that file:

outcome	material
human-food, bio-etc & sanctuary	Food and Drink Waste (wet AD)
compost-indiv	Food and Drink Waste (Composting)

- ii. Calaculate the total amounts of food materials for each of the outcome (converting from kgs to tonnes), to calculate the tonnes-weight value.
- iii. Interpret column B to establish the occurrence-date value.
- iv. Capture outcome information in the enabler-specific value.
- c. Output the end result as the CSV rows:

occurrence-date	scottish-carbon- metric-material	tonnes-weight	enabler	enabler-specific
2021-01-28	Food and Drink Waste (wet AD)	0.08761	Stirling Comunity Food	Used for human- food, bio-etc & sanctuary
2021-01-28	Food and Drink Waste (Composting)	0.00048	Stirling Comunity Food	Used for compost- indiv
2021-01-29	Food and Drink Waste (wet AD)	0.28012	Stirling Comunity Food	Used for human- food, bio-etc & sanctuary
2021-01-29	Food and Drink Waste (Composting)	0.008	Stirling Comunity Food	Used for compost- indiv

1.5.3. Stirling Council - household recycling bins



Stirling Council set a precedent by being the first (and still only) Scottish local authority to have published open data about their bin collection of household waste.

The following examples of measurement data have been taken from that published data.

Here's a snipped image of Stirling Council's measurement data:

	4	В		С		E		F		Н	
1	L	Date	•	Route	•	Waste Collected	~	Quantity	•	Category	₽ T
7	7	03/01/2021 12:1	4	Balfron		72 Mixed Glass			3.5	Recycling	
3	5	04/01/2021 12:1	6	Bridge of Allan		107 Containers Stream	า	1.	86	Recycling	
24	19	10/01/2021 09:5	0	Killearn		106 Fibre (Paper & Ca	rd)	0.	24	Recycling	

The main steps in processing Stirling Council's measurement data (the snippet shown above) are:

- a. Parse the essential values out of columns B C, E F & H
- b. Derive and standardise values ...
 - i. We map column E to to a scottish-carbon-metric-material value. We build this mapping table as a CSV file: stirling-council-to-carbon-metric.csv. Here's a snippet of data from that file:

Waste Collected	material
106 Fibre (Paper & Card)	Mixed paper and board
107 Containers Stream	Average Plastics
72 Mixed Glass	Glass (mixed colours)

- ii. Use column F as the tonnes-weight value.
- iii. Interpret column B to establish the occurrence-date value.
- iv. Note the Route information in the enabler-specific value.
- c. Output the end result as the CSV rows:

occurrence-date	scottish-carbon- metric-material	tonnes-weight	enabler	enabler-specific
2021-01-03	Glass (mixed colours)	3.5	Stirling Council	Balfron
2021-01-04	Average Plastics	1.86	Stirling Council	Bridge of Allan
2021-01-10	Mixed paper and board	0.24	Stirling Council	Killearn

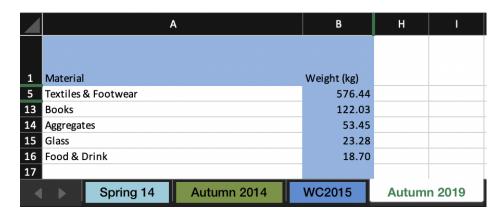
1.5.4. The Fair Share - reused, student-oriented items



The Fair Share is a university based, reuse store. It accepts donations of second-hand books, clothes, kitchenware, electricals, etc. and sells these to students.

The Fair Share is in the process of publishing its data as open data. The following examples of measurement data have been taken from a **draft** of that work.

Here's a snipped image of The fair Share's measurement data:



The main steps in processing The Fair Share's measurement data (the snippet shown above) are:

- a. Parse the essential values out of columns A B, & from the worksheet name
- b. Derive and standardise values ...
 - i. We map column B to to a scottish-carbon-metric-material value. We build this mapping table as a CSV file: the-fair-share-to-carbon-metric.csv. Here's a snippet of data from that file:

(The Fair Share's) Material	material
Textiles & Footwear	Textiles & Footwear
Books	Books
Aggregates	Aggregates (Rubble)
Glass	Glass (mixed colours)
Food & Drink	Food and Drink Waste (wet AD)

- ii. Use column B divided by 1000, as the tonnes-weight value.
- iii. Map the worksheet's name to establish the occurrence-date value.
- iv. Note the university's semester in the enabler-specific value.
- c. Output the end result as the CSV rows:

occurrence-date	scottish-carbon- metric-material	tonnes-weight	enabler	enabler-specific
2019-12-31	Textiles & Footwear	0.57644	The Fair Share	autumn semester
2019-12-31	Books	0.12203	The Fair Share	autumn semester
2019-12-31	Aggregates (Rubble)	0.05345	The Fair Share	autumn semester
2019-12-31	Glass (mixed colours)	0.02328	The Fair Share	autumn semester
2019-12-31	Food and Drink Waste (wet AD)	0.0187	The Fair Share	autumn semester

2. Reference data → CSVs

In this section we walk-through building our two reference data CSVs.

2.1. Carbon metric

The *carbon impact* is a measure devised by Zero Waste Scotland (ZWS), to convey the whole-life carbon impact of waste, from resource extraction and manufacturing emissions, right through to waste management emissions. Its unit-of-measure is *tonnes of carbon dioxide equivalent* (CO₂eT).

This is a reasonable basis for measuring carbon savings so we will use it as reference data.

It is defined in The Scottish Carbon Metric document. For our purpose, its key data is in table 6.2. This contains per-material weight-multipliers that can be used to calaculate CO_2eT amounts.

Here's a snipped image of that table:

Waste Stream	Carbon Weighting
Textiles	100.00
Textiles and Footwear	84.70
Aluminium cans and foil	65.87
Footwear	31.17
Mixed Cans	27.80
Scrap Metal	16.07
Steel Cans	12.25



This table's data has been published as linked open data by our project, in previous work. See the co2e-multiplier files in this repo. But we will redo that work here so that that we can provide an explanatory walk-through.

We copy the table's data from its original PDF format into a more tractable, CSV file: carbon-metric.csv. Here's a snippet of data from that file:

material	multiplier
Textiles	100.00
Textiles and Footwear	84.70
Aluminium cans and foil	65.87
Footwear	31.17
Mixed Cans	27.80
Scrap Metal	16.07
Steel Cans	12.25

So, for our carbon savings data:

- carbon-metric.csv will be referenced as the basis for calculations.
- CO₂eT will be used as the primary unit-of-measure

For example, consider 1 tonne of (used) shoes...

If these were landfilled at a waste site then the *carbon impact* would be:

```
(1 tonne) x (the 'multiplier' value for 'Footwear' from 'carbon-metric.csv')
= (1 tonne) x (31.17 CO2e)
= 31.17 CO2eT
```

Instead, if a reuse store sells them to its customers, it has made a carbon saving of 31.17 CO2_eT.



The term *carbon saving* is a little misleading. A better name for it might be *carbon impact deferrals* since all objects are eventually disposed of. But we won't pursue that philosphical totality in this document.

2.2. Enabler organisations

We describe the enabler organisations in enablers.csv. Here's the data from that file:

name	latitude	longitude
Alloa Community Enterprises	56.122913	-3.781621
Stirling Comunity Food	56.115672	-3.936217
Stirling Council	56.113345	-3.936807
The Fair Share	56.146389	-3.919833

3. CSVs → linked data

3.1. CSVs may be good enough

In sections 2 & 1 we *re-worked* the source data into CSV files with inconsistences rectified, text values mapped to cannocical code lists, and numeric & date values standardised. The resulting CSVs may not be as detailed or have all the nuances as their source data - but for the purpose of understanding *carbon savings*, they are **easy to use**, understand, consume and parse. Indeed, for many purposes and for use by non data experts, such CSVs will be good enough for publication as **open data** without further augmentation.

3.2. Describing our data using linked data vocabularies

Our CSVs have *implicit* meaning and linking. E.g.

- an occurrence-date value, *implicity* has the semantics of being a *date*
- a carbon-metric-material value in the measurements CSV, *implicity* is linked to the same material value in the reference data CSV.

For a standalone case study, this may be good enough but, for our data to become part of the global linked data graph, we need to define its semantics *explicitly*, in standardised way. I.e. we need to describe our data using standard *linked data vocabularies*, to explain to everyone how to interpret our data and how it is linked to other data.

3.3. Using CSVW

CSV on the Web (CSVW) is a standardised mechanism for associating linked data

semantics/metadata with CSV files.

It is nice because it allows us to keep our existing CSVs simple: unadulterated by linked data complications. (Although the CSVs must be of the tidy kind, as output from sections 2 & 1.)

Let's create CSVW files to give our CSVs linked data semantics.

3.3.1. CSVW for the carbon-metric reference data

carbon-metric-metadata.json contains the CSVW that gives linked data semantics to the carbon-metric.csy reference data.

Its features of interest are described below.

```
{
  "@context": "http://www.w3.org/ns/csvw",
  "tableSchema": {
    "columns": [{
      "name": "material",
      "titles": "material",
      "datatype": "string", ①
      "propertyUrl": "http://datacommonsscotland.org/linked-data/property/hasMaterial"
2
   }, {
      "name": "multiplier",
      "titles": "multiplier",
      "datatype": "decimal", ①
      "required": true, ③
      "propertyUrl": "http://datacommonsscotland.org/linked-
data/property/hasMultiplier" ②
    }, {
      "virtual": true, 4
      "propertyUrl": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
      "valueUrl": "http://datacommonsscotland.org/linked-data/class/carbon-metric"
    "primaryKey": "material", 5
    "aboutUrl": "http://datacommonsscotland.org/linked-data/entity/carbon-
metric/{material}" 6
 }
}
```

- ① Assign standard (XMLSchema) data types to each of the CSV's columns.
- ② Define a new RDF predicate for each of the CSV's columns.
- 3 Declare the multiplier value to be mandatory.
- ④ Define the virtual, standard predicate rdf:type to say what class these rows are.
- ⑤ Declare the material value to be the primary key (and to be mandatory).
- **6** Define how to construct the URI (unique identifier) for each of the CSV's rows.

3.3.2. CSVW for the enablers reference data

enablers-metadata.json contains the CSVW that gives linked data semantics to the enablers.csv reference data.

Its new (not previously discussed) features of interest are described below.

```
{
  "@context": "http://www.w3.org/ns/csvw",
  "tableSchema": {
    "columns": [{
      "name": "name",
      "titles": "name",
      "datatype": "string",
      "propertyUrl": "http://datacommonsscotland.org/linked-data/property/hasName"
    }, {
      "name": "latitude", ①
      "titles": "latitude",
      "datatype": "decimal",
      "propertyUrl": "http://datacommonsscotland.org/linked-data/property/hasLatitude"
    }, {
      "name": "longitude", ①
      "titles": "longitude",
      "datatype": "decimal",
      "propertyUrl": "http://datacommonsscotland.org/linked-
data/property/hasLongitude"
   }, {
      "virtual": true,
      "propertyUrl": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
      "valueUrl": "http://datacommonsscotland.org/linked-data/class/enablers"
    }],
    "primaryKey": "name",
    "aboutUrl": "http://datacommonsscotland.org/linked-data/entity/enablers/{name}"
 }
}
```

1 The latitude and longitude values are optional, by CSVW default.

3.3.3. CSVW for the carbon-savings measurement data

carbon-savings-metadata.json contains the CSVW that gives linked data semantics to the carbon-savings.csv measurement data.

Its new (not previously discussed) features of interest are described below.

```
{
    "@context": "http://www.w3.org/ns/csvw",
    "tableSchema": {
        "columns": [{
            "name": "occurrenceDate",
```

```
"titles": "occurrence-date",
      "datatype": "date",
      "propertyUrl": "http://datacommonsscotland.org/linked-
data/property/hasOccurrenceDate"
    }, {
      "name": "material",
      "titles": "scottish-carbon-metric-material",
      "datatype": "string",
      "propertyUrl": "http://datacommonsscotland.org/linked-
data/property/hasMaterial",
      "valueUrl": "http://datacommonsscotland.org/linked-data/entity/carbon-
metric/{material}" ①
    }, {
      "name": "tonnesWeight",
      "titles": "tonnes-weight",
      "datatype": "decimal",
      "required": true,
      "propertyUrl": "http://datacommonsscotland.org/linked-
data/property/hasTonnesWeight"
    }, {
      "name": "enabler",
      "titles": "enabler",
      "datatype": "string",
      "propertyUrl": "http://datacommonsscotland.org/linked-data/property/hasEnabler",
      "valueUrl": "http://datacommonsscotland.org/linked-
data/entity/enablers/{enabler}" (1)
    }, {
      "name": "enablerSpecific",
      "titles": "enabler-specific",
      "datatype": "string",
      "propertyUrl": "http://datacommonsscotland.org/linked-
data/property/hasEnablerSpecific"
    }, {
      "virtual": true,
      "propertyUrl": "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
      "valueUrl": "http://datacommonsscotland.org/linked-data/class/carbon-savings"
    "primaryKey": ["occurrenceDate", "material", "enabler", "enablerSpecific"], ②
    "foreignKeys": [{
      "columnReference": "material",
      "reference": {
        "resource": "carbon-metric.csv", ①
        "columnReference": "material"
      }
    }, {
      "columnReference": "enabler",
      "reference": {
        "resource": "enablers.csv", ①
        "columnReference": "name"
      }
    }],
```

```
"aboutUrl": "http://datacommonsscotland.org/linked-data/entity/carbon-
savings/{occurrenceDate}/{material}/{enabler}/{enablerSpecific}"
}
}
```

- ① Declare the material & enabler values to be, in essence, links into the the carbon-metric & enablers data.
- ② The primary key is a composite. Including enablerSpecific as a component of the primary key, allows the possibility for the enabler to report measurements at an additional level of granularity.

3.4. Generating an RDF graph

We use Swirrl's useful csv2rdf tool to generate a linked data/RDF graph of our data:

```
$ ls data/
carbon-metric.csv carbon-savings.csv enablers.csv ①

$ ls csvw/
carbon-metric-metadata.json carbon-savings-metadata.json enablers-metadata.json ①

$ ./csv2rdf-0.4.6 -m minimal -t data/carbon-savings.csv -u csvw/carbon-savings-
metadata.json -o rdf/carbon-savings.ttl ②

$ ./csv2rdf-0.4.6 -m minimal -t data/carbon-metric.csv -u csvw/carbon-metric-
metadata.json -o rdf/carbon-metric.ttl ②

$ ./csv2rdf-0.4.6 -m minimal -t data/enablers.csv -u csvw/enablers-metadata.json -o
rdf/enablers.ttl ②

$ ls rdf/
carbon-metric.ttl carbon-savings.ttl enablers.ttl ③
```

1 The inputs:

the CSV files containing the actual data, and the CSVW files providing the linked data semantics/metadata.

- ② Run the csv2rdf tool against each CSV+CSVW input pairing.
- ③ The output: Turtle (.ttl) files which define an RDF graph of our data:
 carbon-savings.ttl
 carbon-metric.ttl
 enablers.ttl

4. Using the linked data

The linked data/RDF graph is defined by the Turtle files that we generated in section 3.4. To help us examine and query the RDF graph, we use the Blazegraph tool.

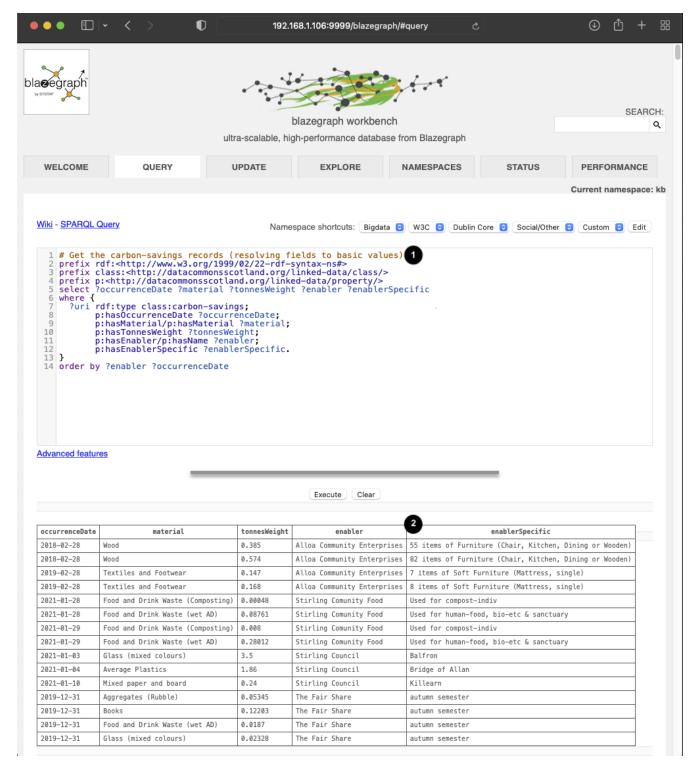
First we load our Turtle files into Blazegraph's data store:

```
$ java -cp blazegraph.jar com.bigdata.rdf.store.DataLoader journal.properties
rdfdir/*.ttl
...
Reading properties: journal.properties
Will load from: rdf/carbon-metric.ttl
Will load from: rdf/carbon-savings.ttl
Will load from: rdf/enablers.ttl
...
Load: 223 stmts added in 0.251 secs, rate= 888, commitLatency=0ms,
{failSet=0,goodSet=3}
```

Then we run Blazegraph's graph engine and SPARQL query service:

```
$ java -server -Xmx4g -jar blazegraph.jar
...
Welcome to the Blazegraph(tm) Database.
Go to http://192.168.1.106:9999/blazegraph/ to get started.
```

Opening that URL in a web browser, gives us a UI allowing us to exaime and query our RDF graph:



- ① Enter a SPARQL query. This example finds carbon-savings records then returns the specified field values.
- 2 See the results of that query.

5. Next?

TODOs

- Add other dimensions such as geographic region (sourced from? given to?)
- Adopt other vocabs? (statistical vocabs?)

- Annotations like dc:title?
- say something about naming and structure... important. Should carbon-metric have been co2e-multipler with its hasMaterial property being rdfs:label?
- this is allows us to very-roughly compare the carbon-saving aspect of the responsible-entities
- hmmm, like any tool, it can be/drive/build good and bad. Hopefully good in today's context. But for now, this is just an exploratory exercise.
- publishing it as open data (e.g. on GitHub)
- data revision? fact retraction and upsert
- cardinality, extra granularity, appending to amounts as the data comes in
- be careful to calculate comparable (say, per-day) amounts from records that report on different time intervals.
- Support upload by enablers (upload → validate → add to the published LoD)
- some aspects of 'provenance' storer id and timestamps might be best left to the comms & storage environment (e.g. Git, Fluree, etc.).
- · Context: COP26 etc.
- a "proper" bibliography?

References

- [carbon-metric] Zero Waste Scotland. The Scottish Carbon Metric. 2011.
- [linked-data] W3C. RDF/XML Syntax Specification. 2004.
- [tidy-data] Hadley Wickham. Tidy data. 2014.
- [CSVW] W3C. CSV on the Web: A Primer. 2016.