

<b>Title</b>	Publishing local 5 star data.  A Technical Companion to the DCLG ‘Code of Recommended Practice for Local Authorities on Data Transparency’.
<b>Revision</b>	4
<b>Status</b>	Draft
<b>Date</b>	24 <sup>th</sup> June 2013
<b>Author</b>	Paul Davidson, LeGSB Director of Standards, and CIO Sedgemoor District Council
<b>Purpose</b>	To provide technical guidance to Local Authorities in England implementing the DCLG code of practice.
<b>Type</b>	LeGSB Guide

Revision	Date	Author	Notes
1	28/03/2013	Paul Davidson, for LeGSB	
2	12/04/2013	Paul Davidson	Take on feedback to revision1 Describe the 5* scheme Restructure headings and layout
3	30/04/2013	Paul Davidson	Minor extra explanations suggested by DCLG
4	24/06/2013	Paul Davidson	Forward from Roger Hampson

## 1. Forward

### 1.1. Roger Hampson, Chief Executive of the London Borough of Redbridge

In September 2011 the DCLG published the ‘Code of Recommended Practice for Local Authorities on Data Transparency’. This outlined the minimum datasets which were to be published by local authorities along with the 5 step process to producing these datasets and others as open linked (5 star) data.

This technical companion to the DCLG code of practice gives real world examples of local authorities which are publishing open 5 star linked data along with the processes followed to achieve this.

It details the practicalities of producing 5 star data by giving step-by-step guides to these real world examples. These examples show how open linked data can enhance local residents’ understanding of data as well as enabling 3<sup>rd</sup> parties to access and make use of your data. It also references useful web resources which can assist this process.

This is a technical document and knowledge of linked data and data formats is required to get the most from this document, and it will allow your authority to:

- understand the fundamentals of linked data;
- recognise the benefits of linked data and when it is of value and when it is not;
- be able to produce linked data or procure linked data services;
- to collaborate with other local authorities to propose and reuse various definitions and vocabularies.

It also highlights the continued work needed to enlarge the definitions and descriptions of data so that datasets can be viewed, compared and linked on a both a national and local basis.

Understanding this document will allow your local authority to be at the forefront of the open linked data agenda, contributing to it and in the process making your data useful to your citizens and your organisation.

Roger Hampson, Chief Executive, London Borough of Redbridge, June 2013

## 2. Introduction

- 2.1. This LeGSB<sup>1</sup> guide is a technical companion to the Department for Communities and Local Government's (DCLG<sup>2</sup>) 'Code of Recommended Practice for Local Authorities on Data Transparency'<sup>3</sup>, which was first issued in September 2011, and relates to English local authorities.
- 2.2. The DCLG 'code' promotes the publication of local data in "*open and machine-readable formats*", and recommends a 5 step journey towards a "*fully open format*". In particular, it makes certain recommendations such as
- Publishing to the 5\* model ( explained later in the guide )
  - Publishing lists of data holdings that have, or could, be published.
- 2.3. Feedback from local authorities tells us that, while many wish to publish data in a smarter, joined up way, that will benefit their residents, they have not been able to find relevant material that explains the practicalities of 5\* data publishing. At LGA events across England in March 2013, LeGSB presented the material in this guide to roomfuls of local authority practitioners, who then fed back that they would value it being turned into a proper guide, and were likely to take part in supporting web collaborations.
- 1.4. This guide sets out to provide sufficient information, for local authority practitioners to
- understand the fundamentals of technologies such as 'linked data';
  - recognise when 'linked data' adds value, and when it doesn't;
  - be able to prepare data so that it could be converted to a linked data form, by themselves, or by others;
  - be able to create and use 'linked data' themselves, or be able to procure 'linked data' services;
  - be able to consider the costs, risks, and benefits of the technologies;
  - challenge what the experts say;
  - collaborate across local authorities to propose, and re-use, vocabularies, definitions, master data, and so on;
  - contribute to the vision of a 'data ecosystem' of joined up information about local services.
- 1.5. The guide focuses on factual information about data publishing techniques; and informs a wider debate about the case for publishing data about local services across the 5\* model.
- 1.6. This is a long document that will take at least half a day to work through and digest. We hope that the reader is rewarded with a confident command of how 5\* data publishing can be applied to local data, towards a data ecosystem that supports efficiency, transformation and transparency of local services.
- 1.7. The guide walks through real examples of local authorities publishing 5\* data, to introduce terms such as '*triples*', '*uri*', '*rdf*', and so on, and describes the steps that takes existing 'traditional' data, and repurposes it as linked data that can be queried over the web. Each example builds on the one before, so do read them in the order that they are presented.
- 1.8. Whilst this guide is primarily aimed at people with a technical background in public sector data, it will also be of interest to:
- non-technical people who wish to understand the potential for efficiency, transformation, and transparency, when publishing data using these techniques;
  - anyone looking for a practical understanding of the technologies involved.

<sup>1</sup> <https://knowledgehub.local.gov.uk/group/localegovernmentstandardsbody>

<sup>2</sup> <https://www.gov.uk/government/organisations/departments-for-communities-and-local-government>

<sup>3</sup> <https://www.gov.uk/government/publications/local-authority-data-transparency-code>

1.9. We have supported this document with a web site at

<http://legsb.i-network.org.uk/resources/publishinglocal5stardata>

, where we take each of the themes raised here and provide more detailed 'how to' information. The web site is where we invite you to collaborate and to add your experiences and recommendations for how certain types of data should be treated.

1.10. With your support, the web site will build into a rich resource base of practical proven solutions to common data publishing scenarios.

1.11. This document is recommended by

- The Department for Communities and Local Government
- The Local Government Association (LGA)
- The Uk Government Linked Data Working Group

## 2. The 5\* Rating System for Open Data

2.1. In 2010, Sir Tim Berners-Lee proposed a 5\* rating system<sup>4</sup> to describe the format in which data can be published, in which, "... *you get more stars as you make it progressively more powerful, easier for people to use.*"

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

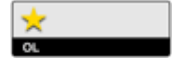

★★★★★ All the above, plus: Link your data to other people's data to provide context

<http://www.w3.org/DesignIssues/LinkedData.html>

3.2. You can find a good walk through of the 5\* rating system at <http://5stardata.info/> which also has examples, and benefits of each level.

3.3. This guide is focused on how to create 5\* data, and explains related terms such as RDF and SPARQL.

3.4. To consider how the 5\* rating system applies to data published by local authorities ...




Star Rating <sup>5</sup>	In practice	Pros	Cons
	This is typically a static document, perhaps containing data as tables.  Often this would be a web page as html <sup>6</sup> , a document as pdf <sup>7</sup> , or an image.	<ul style="list-style-type: none"> <li>• easy to produce;</li> <li>• retains presentation and layout;</li> </ul>	<ul style="list-style-type: none"> <li>• cannot further manipulate the data such as sorting, filtering, summing etc;</li> <li>• cannot join or compare to other data, or earlier versions;</li> </ul>
	This is typically data, such as a spreadsheet, published in the format	<ul style="list-style-type: none"> <li>• no new tools or skills necessary;</li> </ul>	<ul style="list-style-type: none"> <li>• assumes that the consumer has the same</li> </ul>

<sup>4</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

<sup>5</sup> Badges from <http://lab.linkeddata.deri.ie/2010/lod-badges/>

<sup>6</sup> <http://en.wikipedia.org/wiki/HTML>

<sup>7</sup> <http://en.wikipedia.org/wiki/Pdf>

	of the tool that was used to extract it, such as xls <sup>8</sup> .	<ul style="list-style-type: none"> <li>data can be downloaded and further processed and analysed.</li> </ul>	<p>tool as the producer, or is able to use the format.</p> <ul style="list-style-type: none"> <li>very large data sets might not be attractive to download.</li> </ul>
	This is typically 2* type data published using an open format, such as csv <sup>9</sup> or xml <sup>10</sup> .	<ul style="list-style-type: none"> <li>there is typically a choice of open source tools available for each open format</li> </ul>	<ul style="list-style-type: none"> <li>Some conversion necessary.</li> <li>Where data has greater structure than a simple 2-dimensional table, many files may be necessary, which may become disconnected;</li> <li>The meaning and scope of columns and rows can be hard to express.</li> </ul>
	This is explicitly about publishing data to the RDF <sup>11</sup> data model, and providing a query service using the SPARQL <sup>12</sup> language.	<ul style="list-style-type: none"> <li>enables others to make statements over the web about individual lines of data;</li> <li>can be queried over a data service so that a complete data set does not have to be downloaded;</li> <li>gives precise definitions to the meaning of the data.</li> </ul>	<ul style="list-style-type: none"> <li>requires a skill set that most local authorities do not have yet;</li> <li>not suitable to provide directly to residents.</li> <li>Requires some infrastructure</li> </ul>
	This is Linked Data in the Rdf model, that contains links to external datasets to describe the 'things' that the data refers to.	<ul style="list-style-type: none"> <li>adds context;</li> <li>can build into a 'data ecosystem'.</li> <li>can enable 3<sup>rd</sup> parties to join data from many sources together to provide new targeted information services and insight.</li> </ul>	<ul style="list-style-type: none"> <li>requires a 'spine' of core reference data to make links to.</li> </ul>

2.5. Linked Data (5\*) raises a particular opportunity to describe local public services where many providers serve the same communities.

<sup>8</sup> [http://en.wikipedia.org/wiki/Microsoft\\_Excel\\_file\\_format#File\\_formats](http://en.wikipedia.org/wiki/Microsoft_Excel_file_format#File_formats)

<sup>9</sup> [http://en.wikipedia.org/wiki/Comma-separated\\_values](http://en.wikipedia.org/wiki/Comma-separated_values)

<sup>10</sup> <http://en.wikipedia.org/wiki/Xml>

<sup>11</sup> <http://en.wikipedia.org/wiki/Rdf>

<sup>12</sup> <http://en.wikipedia.org/wiki/SPARQL>

2.5.1. In the past, a lead authority would have tried to warehouse this type of information from many sources and then perhaps show it at its website. However, copying and keeping that information up-to-date takes effort, and often goes wrong, leading to poor information being passed on.

2.5.2. Linked data provides an opportunity where each organisation can express just its own information, which will therefore always be accurate and up-to-date, which can be combined later by anyone about any location, theme, community, audience, etc.

2.5.3. This application is described in the third example that follows, and in more detail at the web site.

### 3. Using this Guide

3.1. A good way to understand the fundamentals of 'linked data' is to walk through some examples. We have selected real examples from English local authorities.



If the example illustrates a fundamental point about linked data, and how it applies to local public data, we will highlight the point with a box like this.

3.2. These examples also illustrate some of the design decisions that we need to collectively address, so that our 'linked data' stands a good chance of linking to other 'linked data', and therefore becoming more useful than in its 3\* form.



If the example draws out a design consideration for local public data, we will highlight that with a box like this.

3.3. Occasionally, a point will need more space for explanation than we have space for here, or is well described in other documents. Often these will be deeper technical issues, or techniques that the linked data community is still debating over. You can still get a good understanding of the potential for publishing in this way, but to actually do it, you will need this extra detail.



If a point needs further explanation than we have space for here, we will provide more at the companion website, and highlight it with a box like this.

3.4. For each example, the guide runs through the following steps:

1	Start with some traditional data.	Often this is a spreadsheet, but could also be <ul style="list-style-type: none"> <li>• an XML file</li> <li>• a result set from a database</li> </ul> ... and so on.
2	Model the data.	Create a simple diagram that shows the 'things' that are in the data, and how they are related to other 'things'.
3	Transform the data into a 'linked data' format (RDF).	There are a number of formats that can be used. If you know what the linked data will contain, this can simply be a script to create a new text file.
4	Publish and query	Publishing can simply be about making a text file available

		at a web site, and promoting it. To enable data to be searched though, will need a service.
5	Present and Visualise	If you can get results from a search of linked data, then finally, that data can be presented at a web site, or app, etc.

#### 4. Example 1 - Publishing Air Quality data at Bristol City Council.

##### 4.1. About Air Quality

4.1.1. Councils regularly monitor air pollution to support their activities and duties under the Local Air Quality Management (LAQM) regime. LAQM was introduced to protect citizens against the harmful effects of poor air quality.

4.1.2. Poor air quality causes illness and premature death to people; injury to ecosystems; and physical damage to buildings and structures. It also affects perceptions of the local area and quality of life for residents.

4.1.3. Bristol City Council has used Air Quality data as a pilot for linked data.

##### 4.2. Step 1: Some 'traditional' data.

4.2.1. Bristol City Council are able to create spreadsheets listing the measurements that they have taken

- at various locations around the city
- at various times, at 15 minute intervals
- for particular pollutants

... and provide both 'raw' and 'ratified' readings.

	A	B	C	D	E	F	G	H	I	J	K
1	SiteID	Date	Time	NO2raw_ppb	NOxraw_ppb	NOraw_ppb	NO2rat_ppb	NOxrat_ppb	NOrat_ppb	CAQI	
2	203	01/01/2001	00:30:00	13.5	65	50.5	4.73	19.44	14.71	2	
3	203	01/01/2001	00:45:00	13	55.5	41.5	4.58	16.57	11.99	2	
4	203	01/01/2001	01:00:00	-10	-10	-10				2	
5	203	01/01/2001	01:15:00	-10	-10	-10				2	
6	203	01/01/2001	01:30:00	12.5	46	32.5	4.43	13.7	9.27	2	
7	203	01/01/2001	01:45:00	12.5	36.5	23.5	4.28	10.83	6.55	2	
8	203	01/01/2001	02:00:00	11	54.5	43	3.82	16.26	12.44	2	
9	203	01/01/2001	02:15:00	9	44	35.5	2.92	13.09	10.17	2	
10	203	01/01/2001	02:30:00	10.5	35.5	24	3.82	10.52	6.7	2	
11	203	01/01/2001	02:45:00	12	34	22	3.97	10.07	6.09	2	
12	203	01/01/2001	03:00:00	11.5	32	20	3.97	9.47	5.49	2	
13	203	01/01/2001	03:15:00	11.5	40.5	28	4.13	12.03	7.91	2	
14	203	01/01/2001	03:30:00	13	32	18	4.58	9.47	4.89	2	
15	203	01/01/2001	03:45:00	11	42.5	31.5	3.67	12.64	8.97	2	
16	203	01/01/2001	04:00:00	12.5	31.5	18	4.43	9.31	4.89	2	
17	203	01/01/2001	04:15:00	11	23	11.5	3.82	6.75	2.92	2	
18	203	01/01/2001	04:30:00	11	17.5	6	3.82	5.09	1.26	2	
19	203	01/01/2001	04:45:00	11	19	7	3.97	5.54	1.56	2	
20	203	01/01/2001	05:00:00	10	12.5	2.5	3.37	3.58	0.2	1	
21	203	01/01/2001	05:15:00	8.5	10	1.5	2.92	3.07	0.15	1	
22	203	01/01/2001	05:30:00	10	12.5	2.5	3.37	3.82	0.45	1	
23	203	01/01/2001	05:45:00	8.5	11	2.5	2.92	3.37	0.45	1	

4.2.2. Bristol City Council can provide data for over 8 years. Downloading all of the spreadsheets would take considerable time and space, and then be unwieldy to run calculations and find trends on.



Very large volumes of data may not be practical or attractive for people to download and process locally

- 4.2.3. Each reading is contained in a column where the heading, combines the two ideas of
- The pollutant being measured
  - The process that has been applied to the reading.

Just looking at the spreadsheet does not give any certainty about what the pollutants are, or what process has been used to arrive at the reading. The units of the readings are in 'parts per billion', but the spreadsheet does not explicitly say that. These items are defined elsewhere by the council, but if the spreadsheet were downloaded, these definitions would become disconnected.



Spreadsheet type data often contains column headings which are not sufficient to give confidence about the meaning of the data. Even if these are defined elsewhere, they can become disconnected from the data when downloaded.

Combining or comparing data from different sources which use the same column heading is unsafe.

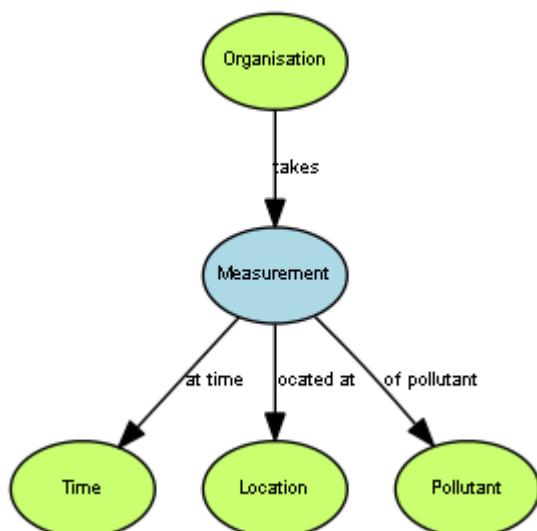
- 4.2.4. The spreadsheet refers to site 203, and the council describes the location of site 203 in another spreadsheet. However, without the other spreadsheet, the code is meaningless, so these would have to be downloaded together and kept aligned, as new versions were created. If the data were to be combined with similar data from another source, we might find another site 203, which refers to a different location.



Spreadsheets often use codes and identifiers but it is sometimes not obvious what they refer to, or which organisation looks after them.

### 4.3. Step 2:Modelling the Data

- 4.3.1. Modelling is about finding the 'types of things' that the data is about, and how each 'thing' is related to other 'things'.





4.3.2. The model is colour coded so that the green ‘types of things’ are those that we might expect to find more about, in other data, from other sources. In this case

Organisation	<p>Bristol City Council.</p> <ul style="list-style-type: none"> <li>• What type of organisation is that?</li> <li>• What geographic area are they responsible for?</li> <li>• What is their air quality strategy?</li> <li>• How much do they spend on monitoring air quality?</li> </ul>
Time and Location	<p>What other events have also occurred around that time and location?</p> <ul style="list-style-type: none"> <li>• traffic counts</li> <li>• health incidents?</li> </ul> <p>Who lives at that location?</p> <ul style="list-style-type: none"> <li>• population</li> <li>• demographics</li> <li>• deprivation</li> </ul>
Pollutant	<p>What do we know about the pollutant?</p> <ul style="list-style-type: none"> <li>• Safe levels?</li> <li>• Health implications</li> </ul>

4.3.3. None of that information is in the original spreadsheet so we will need to find a way to link our data, to where that data can be found. Similarly, our data may contain information that enriches data from other sources, so it will be helpful to publish ours in a way that others can link to.



Modelling the data seeks to find the ‘types of things’ that it refers to, and how they are related to other ‘things’.

In particular, we want to find those ‘things’ for which there may be other, otherwise unrelated, data.

4.3.4. Modelling can start simple and have extra detail added later. There are no strict standards for this process; linked data guidance often talks about using a white board to work with a team of practitioners to then agree that the model captures the necessary ‘things’ and ‘relationships’.

4.3.5. In this case, the ‘organisation’ taking the ‘measurement’ could easily have been missed from the model, as the data we have is all from one source. However, if this data were to be combined with data from other sources, we would have lost that important link.

4.3.6. We have used a ‘directed graph’<sup>13</sup> to capture the ‘types of things’ and relationships. The direction of the arrows is simply to explain the label given to the relationship. Relationships are mostly ‘many-2-many’.



There are a number of tools that can be used to create diagrams like this. Here we have used ‘GraphViz’, but there are others. We shall list some on the companion web site.

<sup>13</sup> [http://en.wikipedia.org/wiki/Directed\\_graph](http://en.wikipedia.org/wiki/Directed_graph)



4.3.7. It is not vital to get this exactly right before moving to the next stage, as new concepts and relationships can be easily added later, without a dramatic impact.



Many 'types of thing' occur over and over again in public service data. e.g. 'location', 'organisation', 'service'.

The companion web site will invite you to collaborate to create a useful list of these, which can be re-used consistently.

We can then work together to find the best definitions, vocabularies, reference data and so on, which can form the basis of links.

4.3.8. From the diagram, you can start to see, what linked data calls, 'triples'. Whilst there are more formal definitions of a 'triple', for the moment, you can think of a triple as a statement about how a 'thing' is related to another 'thing'.



A triple – you can think of it as a statement about how a 'thing' is related to another 'thing'.

4.3.9. From the diagram, you can see that the data could include triples that say

'thing'	'relationship to'	'another thing'
Organisation	takes	Measurement
Measurement	of	Pollutant
Measurement	at	Location

4.3.10. We will see in the next section, that creating linked data, involves breaking the original data into a series of statements, where each statement is a triple.



Creating linked data is about breaking traditional data into a series of triples.

#### 4.4. **Step 3:**Transform the data

4.4.1. An extract from the linked data version of the spreadsheet looks like ...

```
<http://bristol-data-epimorphics.dyndns.org/data/environment/air-quality/observation/conc/location/00203/date/1999-01-01/time/00:15:00/duration/PT15M/substance/NO>
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/month> "01" ;
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/interval> <http://reference.data.gov.uk/id/gregorian-interval/1999-01-01T00:15:00/PT15M>;
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/AirQualityMeasurement>;
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://purl.org/linked-data/cube#Observation>;
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/monitoringPoint> <http://bristol-loc-epimorphics.dyndns.org/so/ef/MonitoringPoint/aqmp.bcc/00203>;
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/year> "1999";
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/ratifiedReading> "10";
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/pollutant> <http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NO>;
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/dateTime> "1999-01-01T00:15:00";
.
```

#### 4.4.2. Without looking closely at the contents, a few points can be made.

<ul style="list-style-type: none"> <li>linked data is for machines, not humans!</li> </ul>	<p>When you looked at the original spreadsheet, it would not have taken long to get comfortable with it, and find, for example</p> <ul style="list-style-type: none"> <li>Each row is a measurement</li> <li>There are columns for time, location, and concentration of certain pollutants.</li> </ul> <p>Raw linked data on the other hand, is not suitable for humans to make sense of; machines need to act on it for us.</p>
<ul style="list-style-type: none"> <li>linked data can be contained in a simple text file</li> </ul>	<p>It is possible ( although not advisable ) to type linked data directly into a text file.</p>
<ul style="list-style-type: none"> <li>RDF is the data model for Linked Data.</li> </ul>	<p>RDF<sup>14</sup> ( Resource Description Format) describes a method of expressing triples, but there is more than one format that can be used, including</p> <ul style="list-style-type: none"> <li>RDF/XML<sup>15</sup></li> <li>N3<sup>16</sup></li> <li>Turtle<sup>17</sup></li> <li>N-Triples<sup>18</sup></li> </ul> <p>This example uses the ‘Turtle’ format, as it is perhaps the most readable; we shall continue with that format throughout the rest of the guide.</p> <p>Other formats are equally valid; some being more suited to being read by other machines.</p>

<sup>14</sup> [http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework)

<sup>15</sup> <http://en.wikipedia.org/wiki/RDF/XML>

<sup>16</sup> [http://en.wikipedia.org/wiki/Notation\\_3](http://en.wikipedia.org/wiki/Notation_3)

<sup>17</sup> [http://en.wikipedia.org/wiki/Turtle\\_\(syntax\)](http://en.wikipedia.org/wiki/Turtle_(syntax))

<sup>18</sup> <http://en.wikipedia.org/wiki/N-Triples>



Linked data is for machines, not humans.



RDF is a method for containing triples, but there are many formats that can be used.

#### 4.4.3. Looking at a triple in more detail

```
<http://bristol-data-epimorphics.dyndns.org/data/environment/air-quality/observation/conc/location/00203/date/1999-01-01/time/00:00:00/duration/PT15M/substance/NOX>
```

```
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/ratifiedReading>
```

"6"

... we can see the three parts that make up a statement. In this case, this statement says that "a particular measurement, had a ratified reading, of 6 parts per billion", **BUT**, at this stage, you cannot tell any of that just by looking at this triple! We will see later how to extract the meaning from a triple.

You can have some confidence about the meaning of the literal "6"!; but just looking at the rest of the triple tells you nothing more. All you can say at the moment is that, some 'thing' is related in some way, to "6".



Just looking at a single triple tells you very little about the statement it is making.

#### 4.4.4. Looking at the first part of that triple

```
<http://bristol-data-epimorphics.dyndns.org/data/environment/air-quality/observation/conc/location/00203/date/1999-01-01/time/00:00:00/duration/PT15M/substance/NOX>
```

... at first glance, looks incomprehensible, and long. This is an example of a URI.

4.4.5. A URI<sup>19</sup> ( Uniform Resource Identifier ) is a string of characters, used to identify a 'thing' ( there are more formal definitions but this will do for now ).



A URI is a string of characters, used to identify a 'thing'.

4.4.6. You might start to recognise some information by looking at the URI string, which gives you an idea what 'thing' it is referring to. For example, it mentions a location, a date, and a substance. HOWEVER, you cannot tell any of that just by looking at the URI itself. In fact, you should not pick out and use bits of information from a URI; we will see soon, that you have to look a URI up, to have any certainty about the 'thing' that it is referring to.

<sup>19</sup> <http://en.wikipedia.org/wiki/URI>



Just looking at a URI tells you very little about the ‘thing’ that it refers to, even if it seems to contain concepts and values.

4.4.7. A URI looks like the more familiar URL, but it does not have to, in fact, ABC123, would also be a valid, if unhelpful, URI that could have been used to identify this measurement.

4.4.8. For a URI to be useful, we need to have a way to find out what it means. URIs for linked data typically start with ‘http://’ so that they can be looked up over the web, to provide information about the ‘thing’ that they identify. We shall ‘look up’ some of these URIs later in the example, but for now, we still don’t know much about what this particular URI is referring to.



URIs for linked data typically start with ‘http://’ so that they can be looked up over the web to find out what they mean.

4.4.9. One thing that we do know about this URI, is who has control over the information that is returned when it is looked up. In this case, it is the people who control the content at <http://bristol-data-epimorphics.dyndns.org>. This is important because other data is only likely to re-use URIs to mean the same thing, when they are defined by organisations that they know and trust. Using the same URI across many datasets is the basis of linking.



Links are made between otherwise separate datasets when they use the same URIs.

4.4.10. So, all we know about this URI so far, is that it refers to a ‘thing’, and that the people at <http://bristol-data-epimorphics.dyndns.org> have control over what that thing is. We won’t know what that is until we look it up, which we will do later in the example.

4.4.11. Looking at the second part of that triple ...

```
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/ratifiedReading>
```

... this is a URI that states a relationship between the ‘thing’ that the first URI identifies, to, in this case, the literal “6”.

4.4.12. As before, despite seeing the word ‘ratifiedReading’, we can have no idea what this relationship is until we look it up, and it is the people at <http://bristol-data-epimorphics.dyndns.org> who have control over what the URI means. We will look it up soon.

4.4.13. Here is another triple from the air quality data

```
<http://bristol-data-epimorphics.dyndns.org/data/environment/air-quality/observation/conc/location/00203/date/1999-01-01/time/00:00:00/duration/PT15M/substance/NOX>
```

```
<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/pollutant>
```

```
http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NOX
```

Again, we can have no idea what the statement means, but, notice that the first URI, is the same as the first URI from the previous triple. So now, we can see that this second statement is about the same ‘thing’ that the first statement was about ( although we still don’t know what

that ‘thing’ is ).



Statements that use the same URI are referring to the same ‘thing’.

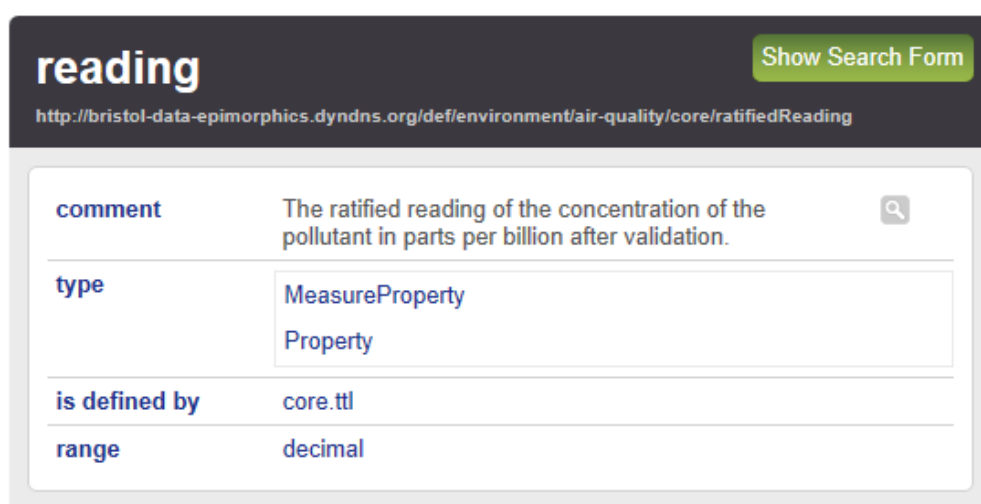
4.4.14. Note also that the third part of the triple is not a literal as in the first example, but another URI.

4.4.15. We will find out soon that this statement says that the measurement was of ‘Nitrogen Oxide’, and putting the two triples together tells us that the measurement of Nitrogen Oxide was 6 parts per billion. Other triples about the same measurement tell us about where and when the measurement was taken. In this way, a single row from a spreadsheet is described using many triples.

4.4.16. We can look up some of these URIs by typing them into a web browser. If we look up

`<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/ratifiedReading>`

we get a web page returned by the people at <http://bristol-data-epimorphics.dyndns.org/>



reading	
<a href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/ratifiedReading">http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/ratifiedReading</a>	
comment	The ratified reading of the concentration of the pollutant in parts per billion after validation.
type	MeasureProperty Property
is defined by	core.ttl
range	decimal



Looking up a URI is also known as

- dereferencing
- resolving

... and now we find a precise definition of what the URI means. In this case

- it identifies it as a ‘property’, which is a ‘type of thing’ that can be used to state a relationship
- it explains that this is a ‘ratified reading’ in parts per billion.
- it defines that the value of the property must be a decimal number.

4.4.17. There will be more about defining ‘properties’ and ‘types of things’ in a later example.

4.4.18. We said earlier that linked data is not for humans, it is for machines. If you are a machine looking up a URI, you need a machine-readable response; in this case

```
<?xml version="1.0" encoding="UTF-8"?>
- <result version="0.2" href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-
quality/core/ratifiedReading.xml" format="linked-data-api">
  <definition href="http://bristol-data-epimorphics.dyndns.org/meta/def/environment/air-
quality/_vocab/_term.xml"/>
  <extendedMetadataVersion href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-
quality/core/ratifiedReading.xml?_metadata=all"/>
  - <primaryTopic href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-
quality/core/ratifiedReading">
    <comment lang="en">The ratified reading of the concentration of the pollutant in parts per billion
    after validation.</comment>
    <isDefinedBy href="http://bristol-data-epimorphics.dyndns.org/sources/def/environment/air-
quality/core.ttl"/>
    <label lang="en">reading</label>
    <range href="http://www.w3.org/2001/XMLSchema#decimal"/>
    - <type>
      <item href="http://purl.org/linked-data/cube#MeasureProperty"/>
      <item href="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    </type>
  </primaryTopic>
</result>
```

- 4.4.19. Getting the right format of data when looking up a URI, will depend up whether a person, or a machine, is asking for it. When looking up a URI, the response you get back will be dependent upon the 'content type' that you ask for in the http header of the request. For people, normal web browsers look after that for us and ask for an html web page. When programs, running on computers, need to look up a URI, and expect a machine-readable response, they need to explicitly ask for the right format. The companion web site will point to guidance about how to do that.



How to use 'content negotiation' to offer a range of formats, when your URIs are dereferenced, and when dereferencing URIs from other sources.

- 4.4.20. When publishing linked data, you will re-use URIs that are defined elsewhere. However, where a suitable URI does not exist, you will create your own. We will shortly give some advice on how to construct the URI string. For URIs that you create, you will need to provide a response when others dereference your URI. That response will contain data that defines what the URI means, and gives further useful information and links. You will probably respond with html for humans, and with one or more machine-readable formats too.



How to respond when your URIs are dereferenced.



The companion web site will invite you to propose what formats we should be responding with when URIs are dereferenced for common scenarios.

- 4.4.21. There is a subtle, yet vital distinction in linked data, between a URI, and the URL that you get back when looking one up. Note that the URIs in the examples do not have an extension ( such as .html ) as you would often see in a URL. The URI identifies an actual 'real world thing', such as an individual school, or a local authority objective, whereas the URL that is the response to looking it up, locates web information about that 'thing'. Triples use URIs to make statements about the 'real world things', but when you look a URI up, they get redirected to

URLs that locate a description of that ‘thing’. The companion web site will point to guidance about how to do that.



How to redirect an http request for a URI, to a URL that gives information about it.



A stricter definition of a URI and the further classifications of URL and URN can be found at <http://www.ietf.org/rfc/rfc3986.txt> .

4.4.22. To continue, we can look up

`<http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NOX>`

and we get a web page returned by the people at <http://bristol-data-epimorphics.dyndns.org/>



type	Pollutant Concept
is defined by	air-pollutant.ttl

... or, depending on the format that is asked for in the request, a machine-readable version

```
<?xml version="1.0" encoding="UTF-8"?>
- <result version="0.2" href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NOX.xml" format="linked-data-api">
  <definition href="http://bristol-data-epimorphics.dyndns.org/meta/def/environment/air-quality/_vocab/_term.xml"/>
  <extendedMetadataVersion href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NOX.xml?_metadata=all"/>
- <primaryTopic href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NOX">
  <isDefinedBy href="http://bristol-data-epimorphics.dyndns.org/sources/def/environment/air-quality/air-pollutant.ttl"/>
  <label lang="en">NOx - Nitrogen oxide (mono or di)</label>
  - <type>
    <item href="http://www.w3.org/2004/02/skos/core#Concept"/>
    <item href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/Pollutant"/>
  </type>
</primaryTopic>
</result>
```

... both of which reveal that the URI identifies Nitrogen Oxide.



4.4.23. If another statement, from another dataset, perhaps from another publisher, also used this same URI to refer to Nitrogen Oxide, then, we will see later, that we could link the two statements.



Statements that use the same URI, are about the same ‘thing’.

4.4.24. It is therefore important that the most appropriate organisation for a particular ‘type of thing’, provides sets of URIs to define those things. This then is the ‘core reference data’ that underpins the links. In this case, perhaps the Department for Food, Environment & Rural Affairs (DEFRA) would be the most appropriate organisation to provide a list of ‘pollutants’ that are defined in the National Air Quality Strategy, and publish them as a ‘URI Set’, or recommend an existing set.



The Cabinet Office provide guidance about how to design URI Sets for this purpose.

4.4.25. A URI set from a trusted source is most likely to be re-used in linked data from many separate publishers. This is the difference between 4\* and 5\* data. You can only achieve 5\* data, if there are quality URIs from a credible source, that describe the ‘things’ that are in your data.



4\* data uses URIs created by the publisher of the data. 5\* data makes use of URIs created elsewhere, which are likely to be re-used in other data from other sources.

4.4.26. In this example, there was not a useable URI Set for pollutants, and therefore, the people at <http://bristol-data-epimorphics.dyndns.org/> created their own URIs for the pollutants that were in the data. However, it is unlikely, and indeed inadvisable, that other councils would use these URIs, and therefore, the immediate linking opportunity is lost. Whilst it is possible to add further triples later to say that a URI means the ‘same as’ another URI, we will get better value from linked data when we can link to quality sources of these URI Sets.

4.4.27. To make a success of 5\* data publishing, we need to discover and promote the existing URI Sets that describe the ‘things’ that we often find in local public data. Where those URIs do not exist, we need to work with the owners of that content to set them up.

4.4.28. There are already some good reliable sources of URI Sets that can be confidently used to make links across local public linked data, for example

Publisher	URI Sets
The esd toolkit <sup>20</sup>	Types of ‘service’, ‘function’, ‘metric’ etc
DCLG at their data cabinet <sup>21</sup>	Local Authorities in England
The Ordnance Survey <sup>22</sup>	Administrative geographies
Companies House <sup>23</sup>	Companies in the UK
CIPFA <sup>24</sup>	Financial Expenditure Categories

<sup>20</sup> <http://standards.esd.org.uk/>

<sup>21</sup> <http://opendatacommunities.org/data>

<sup>22</sup> <http://data.ordnancesurvey.co.uk>

<sup>23</sup> <http://www.companieshouse.gov.uk/about/miscellaneous/URI.shtml>

<sup>24</sup> <http://www.cipfastats.net/sercop/>



The companion web site will invite you to propose what URI Sets we should be using in local public data.

4.4.29. However, there is plenty more ‘core reference data’ that needs to be published as URI Sets to enable local public data to be linked.



The companion web site will invite you to propose and prioritise which core reference data is necessary to make your data link. We can then take that to the various groups that are working to publish public data.

4.4.30. Returning to the first URI in the triples that we have looked at ...

```
<http://bristol-data-epimorphics.dyndns.org/data/environment/air-quality/observation/conc/location/00203/date/1999-01-01/time/00:00:00/duration/PT15M/substance/NOX>
```

This URI identifies the event of a single measurement of a particular pollutant, at a location and a time. It effectively identifies a row in the original spreadsheet. This is NOT ‘core reference data’ and there is no chance that this URI will have already been created by another organisation, so It will have to be created by the people that are publishing the data.

4.4.31. When creating a local URI, there are some points to consider.

- The URI must be unique

The URI must not be re-used to mean something else.

A good way of ensuring that is to

- a) root the URI at a web domain, that means that the rest of the string will be unique within that domain.
- b) use ‘natural’ keys that refer to the ‘thing’ that the URI identifies.

The use of natural keys in the URI string, can make it a simple job to create URIs from spreadsheets or databases etc. This is why URIs strings often seem to be sufficient to define the ‘thing’, but as you can see, it is not safe to assume that.

The Cabinet Office provide guidance about using ‘concept/value’ pairs when creating URIs.

- The URI should dereference to give information useful information.

Sir Tim Berners-Lee described Cool URIs at <http://www.w3.org/TR/cooluris/>



The Cabinet Office provides guidance about using 'concept/value' pairs when creating URIs.

4.4.32. So, the task of 'transforming the data' is about taking each row from the spreadsheet, and creating a series of statements, in 'triple' form, that use URIs or literals. Where quality URIs do not already exist, create your own at a domain that you control, using natural keys to keep them unique and provide a response when someone ( or some machine ), looks them up.

4.4.33. The result of that can be a simple ( but long ) text file.

#### 4.5. **Step 4:**Publish and query

4.5.1. The simplest way to publish linked data, is to host the text file at a web site, and promote it to be downloaded by others.

4.5.2. However, to get value from having the data in this form, it is useful to load it into a linked data database, and allow that to be queried by anyone over the web. A linked data database store triples, and is called a triplestore.



A linked data database is called a triplestore.



There are a number of products and services that provide a 'triplestore'; some opensource. The companion web site will give more information about options for 'triplestores'.

4.5.3. Unlike traditional relational databases, triplestores do NOT need to be configured to know about the subject or structure of the data that they will contain. A triple store can be 'turned on' and then just loaded with files containing triples.



Triplestores do not need to be configured to know about the structure of the data that they will contain.

4.5.4. If a new field, or relationship, or index, is needed to a relational database, a considerable amount of development and testing is needed. In this circumstance, no change is necessary to a triplestore, as it just takes on more triples.

4.5.5. Linked Data can be loaded into any number of triplestores, alongside other linked data from other sources, so that queries can be run over them.



Linked data can be loaded into any number of triplestores.

4.5.6. It does not follow that each organisation that has linked data to publish, needs its own triplestore. Some might be setup regionally; some might be set up to contain data about a

particular theme. Some might be hosted by the public sector, and others may be run commercially, or by volunteers. Local authorities will want to consider if they will operate their own triplestores, and/or, encourage others to load the local authority's data into other triplestores.



There are a number of products and services that provide a triplestore; some opensource. The companion web site will give more information about options for triplestores.

4.5.7. Similarly, local authorities will want to consider which triplestores they wish to consume data from.

4.5.8. Many technical practitioners at local authorities will be comfortable with SQL<sup>25</sup>, which is a language for querying traditional relational databases. Linked data, in a triplestore, can be queried using a language called SPARQL<sup>26</sup>. Triplestores offer what is known as a 'SPARQL Endpoint', which is effectively a web location where SPARQL queries can be directed to a triplestore.



Linked data can be queried using a language called SPARQL.



There are good books, and tutorials that explain how to use SPARQL. The companion web site will list some, and give some examples.

4.5.9. To continue with the Bristol Air Quality example, the linked data has been loaded into a triplestore, which has a SPARQL endpoint.

<sup>25</sup> <http://en.wikipedia.org/wiki/Sql>

<sup>26</sup> <http://en.wikipedia.org/wiki/Spqrql>

## SPARQLer - General purpose processor

General SPARQL query : input query, set any options and press "Get Results"

```
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX sr: <http://data.ordnancesurvey.co.uk/ontology/spatialrelations/>
PREFIX def-aq: <http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/core/>
PREFIX def-bcc: <http://bristol-loc-epimorphics.dyndns.org/def/ef/MonitoringPoint/AirQuality/>
PREFIX def-poll: <http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/>
PREFIX def-caqi: <http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/caqi/>

SELECT DISTINCT ?measurement ?reading ?pointname ?easting ?northing WHERE {
  ?measurement a def-aq:AirQualityMeasurement .
  ?measurement def-aq:pollutant def-poll:NO .
  ?measurement def-aq:ratiifiedReading ?reading .
  ?measurement def-aq:monitoringPoint ?point .
  ?point rdfs:label ?pointname .
  ?point sr:easting ?easting .
  ?point sr:northing ?northing .
  FILTER (?reading > 100)
}
LIMIT 20
```

Target graph URI (or use FROM in the query)

Output XML: ☒ with XSLT style sheet (leave blank for none):

or JSON output: ☐

or text output: ☐

or CSV output: ☐

or TSV output: ☐

Force the accept header to text/plain regardless ☐

### 4.5.10. This query asks for measurement information where

• The reading was over 100 ppb	FILTER (?reading > 100)
• The pollutant was Nitrogen Oxide.	?measurement def-aq:pollutant def-poll:NO .

### 4.5.11. We don't have the space here to explain how SPARQL works, but to help you recognise what is going on

• PREFIX	<p>Enables us to shorten the references to URIs that are used later in the query. We don't need to have these PREFIXes, but the query itself would then be harder to read.</p> <p>For example, the PREFIX def-poll: equates to <a href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant">http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant</a> so that URI <a href="http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NO">http://bristol-data-epimorphics.dyndns.org/def/environment/air-quality/air-pollutant/NO</a> can be expressed as def-poll:NO in the query.</p> <p>The name given to the prefix is given by whoever is making the query.</p>
• SELECT DISTINCT	As with SQL, the DISTINCT keyword requests that duplicate results are not returned.
• ?variable	This is a list of the variables that should be returned in the results. The names given to the variables are given by whoever is making the query, and don't have to be meaningful, or relate to the expected contents. This is different to SQL where the names of fields in the query needs to be the same as those in the database.
• Pattern	<p>These are sets of triple patterns that need to be matched to find solutions to the query.</p> <p>For example</p> <p>?measurement def-aq:pollutant def-poll:NO .</p>

	<p>... means find any triple that has a pollutant of Nitrogen Oxide, and bind the first part of the triple to the variable ?measurement.</p> <p>Other lines that also use the same ?measurement variable must also be true for that URI for it to appear in the results.</p>
• FILTER	These are expressions that can be used to further filter the results.

4.5.12. SPARQL endpoints typically offer results in a range of formats. In this case, as html

reading	pointname	easting	northing
"122.95" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"brislington" @en	"361178" ^^<http://www.w3.org/2001/XMLSchema#double>	"171566" ^^<http://www.w3.org/2001/XMLSchema#double>
"122.95" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Brislington Depot" @en	"361178" ^^<http://www.w3.org/2001/XMLSchema#double>	"171566" ^^<http://www.w3.org/2001/XMLSchema#double>
"104.07" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"brislington" @en	"361178" ^^<http://www.w3.org/2001/XMLSchema#double>	"171566" ^^<http://www.w3.org/2001/XMLSchema#double>
"104.07" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Brislington Depot" @en	"361178" ^^<http://www.w3.org/2001/XMLSchema#double>	"171566" ^^<http://www.w3.org/2001/XMLSchema#double>
"104.92" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"brislington" @en	"361178" ^^<http://www.w3.org/2001/XMLSchema#double>	"171566" ^^<http://www.w3.org/2001/XMLSchema#double>

... and as XML

```
<?xml version="1.0"?>
- <sparql xmlns="http://www.w3.org/2005/sparql-results#">
  - <head>
    <variable name="measurement"/>
    <variable name="reading"/>
    <variable name="pointname"/>
    <variable name="easting"/>
    <variable name="northing"/>
  </head>
  - <results>
    - <result>
      - <binding name="measurement">
        <uri>http://bristol-data-epimorphics.dyndns.org/data/environment/air-quality/observation/conc/location/00203/date/2002-01-01/time/18:00:00/duration/PT15M/substance/NO</uri>
      </binding>
      - <binding name="reading">
        <literal datatype="http://www.w3.org/2001/XMLSchema#decimal">122.95</literal>
      </binding>
      - <binding name="pointname">
        <literal xml:lang="en">brislington</literal>
      </binding>
      - <binding name="easting">
        <literal datatype="http://www.w3.org/2001/XMLSchema#double">361178</literal>
      </binding>
      - <binding name="northing">
        <literal datatype="http://www.w3.org/2001/XMLSchema#double">171566</literal>
      </binding>
    </result>
  </results>
</sparql>
```

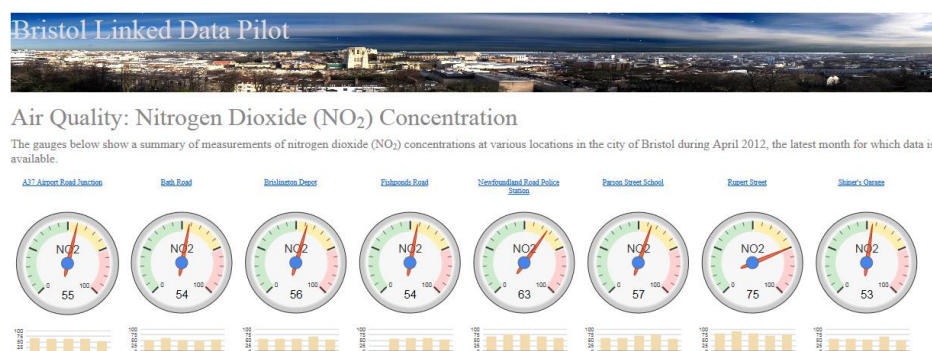
4.5.13. In this way, a small results set can be obtained from a large set of data. SPARQL also has functions that enable data to be aggregated, totaled, averaged and so on, without having to first download all of the data.

4.5.14. SPARQL is powerful, but can also be a specialism to learn, and also requires some knowledge of the URIs and shape of the data in a triplestore. As an alternative, some publishers also provide an ‘api’<sup>27</sup> onto a triplestore which provides a much easier way of querying data for the most common types of question. For example, we may have been able to ask this question as

<http://bristol-data-epimorphics.dyndns.org/air-quality/pollutant/NO/readingmorethan/100>

#### 4.6. **Step 5:Present and Visualise**

4.6.1. Once a machine-readable result set can be obtained from a query, we are now back to ‘known ground’ when it comes to presenting it and visualising it. At Bristol, they created a web site to show the data, using some google dials and charts.



#### 4.7. More about this pilot from Bristol City Council

4.7.1. Bristol City Council worked with Epimorphics<sup>28</sup> to produce this pilot of applying linked data to local public data.

4.7.2. You can find this pilot at

<http://bristol-data-epimorphics.dyndns.org/>

the resulting presentation of the data;

<http://bristol-data-epimorphics.dyndns.org/sparql.html>

the SPARQL endpoint;

<http://bristol-data-epimorphics.dyndns.org/dataOverview.html>

for the story of how it was done.

### 5. Example 2 - Publishing Land Supply Data at Hampshire County Council.

#### 5.1. About Land Supply Data

5.1.1. Housing land supply data is collected by councils to help them monitor the rate at which development is occurring, with surveys of sites occurring at the end of the financial year

5.1.2. Alongside the ability to monitor development rates the data collected forms an essential part in ‘Small Area Population Forecasting’; allowing the council to project future population and incorporate this data into planning for future services.

5.1.3. The data in this pilot shows the projected phasing of the sites over the coming five years.

5.1.4. This example starts to show the potential for 5\* data.

<sup>27</sup> <http://en.wikipedia.org/wiki/Api>

<sup>28</sup> <http://www.epimorphics.com/web/>



## 5.2. Step 1:Some ‘traditional’ data

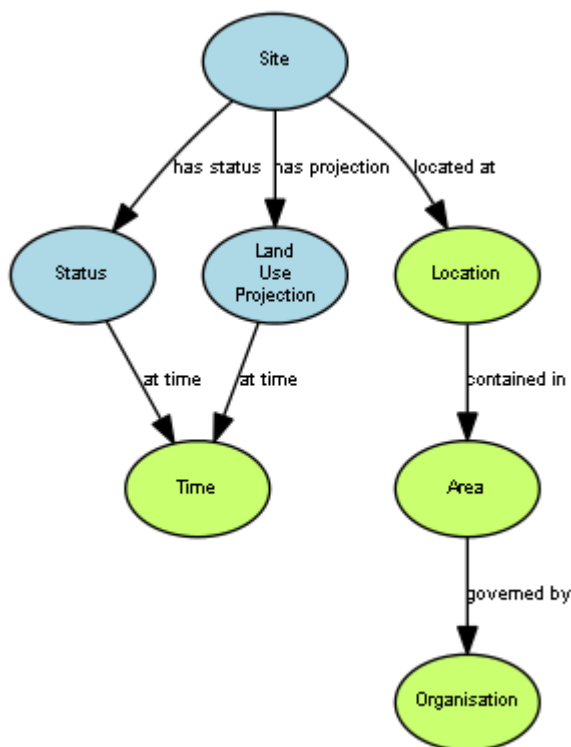
5.2.1. Hampshire collated spreadsheets from information from each of the Districts, and Unitary Authorities in their administrative area, about the intended number of properties to be built on defined sites over the next few years.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	DISTRICT	DIST	SITE	SCH OID	ADDRESS	PREVIOUSLY DEVELOPED LAND	STATUS	2011- 12_Gains	2011- 12_Losses	NET_DWLS_ AVAILABLE	NET_PERMITTED _DWLS	2012_13	2013_14	2014_15	2015_16
1															
2	BASINGSTOKE & DEANE	01	0300	1360	PARK PREVIEW	PART BROWNFIELD	UNDER CONSTRUCTI	33	0	132	132	65	37	30	0
3	BASINGSTOKE & DEANE	01	0317	1629	LAND NORTH	GREENFIELD	UNDER CONSTRUCTI	0	0	784	784	0	50	150	150
4	BASINGSTOKE & DEANE	01	0318	2047	LAND AT	GREENFIELD	UNDER CONSTRUCTI	141	0	257	257	87	80	80	10
5	BASINGSTOKE & DEANE	01	0337	1127	BOUNDARY H	BROWNFIELD	NOT STARTED	0	0	115	115	40	75	0	0
6	BASINGSTOKE & DEANE	01	0339	2415	LAND AT	GREENFIELD	DELETED 2012. TOTA	44	0	0	0	0	0	0	0
7	BASINGSTOKE & DEANE	01	0390	2361	LAND AT	PART BROWNFIELD	PHASE 1 COMPLETE,	54	0	164	164	60	60	44	0
8	BASINGSTOKE & DEANE	01	0391	2337	LAND TO THE	GREENFIELD	NOT STARTED	0	0	50	0	0	0	0	0
9	BASINGSTOKE & DEANE	01	0392	1170	ALDERMASTC	GREENFIELD	NOT STARTED	0	0	300	0	0	0	0	0
10	BASINGSTOKE & DEANE	01	0393	1877	LAND AT	GREENFIELD	NOT STARTED	0	0	130	0	0	0	0	50
11	BASINGSTOKE & DEANE	01	0394	389	FORMER ALL	GREENFIELD	DELETED 2012. 291 C	32	0	0	0	0	0	0	0
12	BASINGSTOKE & DEANE	01	0395	961	LAND NORTH	GREENFIELD	NOT STARTED	0	0	750	0	0	0	65	85
13	BASINGSTOKE & DEANE	01	0396	534	LAND AT	PART BROWNFIELD	NOT STARTED	0	0	300	0	0	0	0	0
14	BASINGSTOKE & DEANE	01	0398	393	LAND SOUTH	BROWNFIELD	NOT STARTED	0	0	40	0	0	0	0	0
15	BASINGSTOKE & DEANE	01	0399	525	OLD INDUSTRI	BROWNFIELD	NOT STARTED	0	0	30	0	0	0	0	0
16	BASINGSTOKE & DEANE	01	0408	545	HARROW GAF	BROWNFIELD	NOT STARTED	0	0	27	27	0	0	10	17
17	BASINGSTOKE & DEANE	01	0413	2146	CLIFTON HOU	BROWNFIELD	NOT STARTED	0	0	205	205	0	0	0	0
18	BASINGSTOKE & DEANE	01	0426	661	COUNTY INFA	BROWNFIELD	DELETED 2012. TOTA	3	0	0	0	0	0	0	0
19	BASINGSTOKE & DEANE	01	0428	1314	GREYTOWN H	BROWNFIELD	NOT STARTED	0	0	24	24	0	0	24	0
20	BASINGSTOKE & DEANE	01	0429	1591	LAND AT	BROWNFIELD	DELETED 2012. TOTA	63	0	0	0	0	0	0	0
21	BASINGSTOKE & DEANE	01	0440	1332	19-21	BROWNFIELD	DELETED 2012. TOTA	11	0	0	0	0	0	0	0
22	BASINGSTOKE & DEANE	01	0444	2098	CHEAM HAWT	BROWNFIELD	6 DWELLINGS COMPL	0	0	6	6	0	0	0	0
23	BASINGSTOKE & DEANE	01	0449	414	LAND AT	BROWNFIELD	STAFF FLATS COMPI	0	0	65	65	0	0	0	0
24	BASINGSTOKE & DEANE	01	0459	1464	151-157	BROWNFIELD	DELETED 2012. ALTE	0	0	0	0	0	0	0	0
25	BASINGSTOKE & DEANE	01	0460	2141	AND 1 NEW S'	BROWNFIELD	UNDER CONSTRUCTI	0	0	12	12	12	0	0	0
26	BASINGSTOKE & DEANE	01	0461	1690	BROOK HOUS	BROWNFIELD	NOT STARTED	0	0	30	30	0	0	0	30
27	BASINGSTOKE & DEANE	01	0462	2561	CHAILEY COU	BROWNFIELD	NOT STARTED	0	0	24	24	0	0	0	0
28	BASINGSTOKE & DEANE	01	0463	2007	WILKINS	BROWNFIELD	DELETED 2012. PERM	0	0	0	0	0	0	0	0

## 5.3. Step 2:Modelling the Data

5.3.1. Looking for ‘types of things’ and relationships in this data, draws the opportunities to link the data via

Organisation	<p>The councils that govern each geographic area that contain each ‘site.</p> <ul style="list-style-type: none"> <li>The District Councils and Unitary Councils in Hampshire</li> <li>The County Council also fits this definition</li> </ul> <p>This example actually also relates each ‘Site’ to the Planning Authority, but that introduces the idea of ‘planning’ as a ‘public service’ provided by an ‘organisation’, and we will get to that in a later example. For now, we keep the model simple.</p>
Location	<p>The location of a site, as described by coordinates, can be used to find other information at or around that location.</p>



## 5.4. Step 3: Transform the data

### 5.4.1. An extract from the linked data version of the spreadsheet looks like .

<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://linkeddata.hants.gov.uk/def/land-supply/LandSupplySite>
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://www.ordnancesurvey.co.uk/ontology/BuildingsAndPlaces/v1.1/BuildingsAndPlaces.owl#hasAddress>	"URBAN PRIORITY AREA GOLDSMITH AVENUE/FANCETT ROAD PORTSMOUTH"
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://linkeddata.hants.gov.uk/def/land-supply/siteStatus>	"NOT STARTED"
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://linkeddata.hants.gov.uk/def/land-supply/previouslyBuilt>	<http://linkeddata.hants.gov.uk/def/land-supply/concept/previously-bi>
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://linkeddata.hants.gov.uk/def/land-supply/planningAuthority>	<http://opendatacommunities.org/id/unitary-authority/portsmouth>
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://data.ordnancesurvey.co.uk/ontology/Admingeo/district>	<http://data.ordnancesurvey.co.uk/id/7000000000037254>
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://data.ordnancesurvey.co.uk/ontology/Admingeo/district>	"Land supply site: Portsmouth, 0383"
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://www.w3.org/2000/01/rdf-schema#label>	"portsmouth-0383"
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://www.w3.org/2004/02/skos/core#notation>	S0.795714
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	-1.075231
<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	<http://linkeddata.hants.gov.uk/def/land-supply/LandSupplySite>
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	"BOUNDARY HALL MULFORDS HILL TADLEY"
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://www.ordnancesurvey.co.uk/ontology/BuildingsAndPlaces/v1.1/BuildingsAndPlaces.owl#hasAddress>	"NOT STARTED"
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://linkeddata.hants.gov.uk/def/land-supply/siteStatus>	<http://linkeddata.hants.gov.uk/def/land-supply/concept/previously-bi>
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://linkeddata.hants.gov.uk/def/land-supply/previouslyBuilt>	<http://opendatacommunities.org/id/district-council/basingstoke>
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://linkeddata.hants.gov.uk/def/land-supply/planningAuthority>	<http://data.ordnancesurvey.co.uk/id/7000000000017653>
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://data.ordnancesurvey.co.uk/ontology/Admingeo/district>	"Land supply site: Basingstoke And Deane, 0337"
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://www.w3.org/2000/01/rdf-schema#label>	"Basingstoke-0337"
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://www.w3.org/2004/02/skos/core#notation>	S1.358336
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	-1.146852
<http://linkeddata.hants.gov.uk/id/land-supply/site/basingstoke/0337>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	

### 5.4.2. Picking out some triples

<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>

<http://linkeddata.hants.gov.uk/def/land-supply/siteStatus>

"NOT STARTED"

5.4.3. Notice that the domain for these URIs is at "hants.gov.uk". URIs from a ".gov.uk" domain, or other trusted and well known domains, gives more 'believability' to the URIs. A person considering if they should use a data set, or a URI is likely to assume that there will be some quality control being applied to data published by such a site, as opposed to the same data from, say "fly-by-night.com". This is the first step in understanding the provenance of data, i.e. who is making this statement.



Establishing the 'provenance' of data starts with knowing who is making each statement.

5.4.4. So far in our examples, we can tell which organisation defines the meaning of each URI, but that does not tell us who is making a statement which might use those URIs. Linked Data statements can be made by anyone, about anything, using anyone's URIs, so it becomes important to link provenance, and quality information to each statement. This information can then be used to judge if a set of data is fit for a purpose.



There is much more to checking the provenance and quality of linked data. We explain more at the companion web site, and highlight activity in the linked data community.

5.4.5. As we have seen, a file containing linked data, can be hosted anywhere on the web, or be contained in many triplestores, so it does not follow that the statements in a set of linked data, are being made by the people at the web site that it is published at.



It does not follow that the statements in a set of linked data, are being made by the people at the web site that it is published at.



The companion web site will invite you to propose what provenance information you expect to be linked to local linked data so that others can judge if it is suitable for their purpose.

5.4.6. The last part of the triple, is the literal "NOT STARTED", but there are only a limited number of values that are contained in the spreadsheet for this. This is effectively, one possible value from a 'controlled vocabulary'. It is often useful to publish these 'controlled vocabularies' separately, so that each entry has its own URI. The terms can then be used in many data sets as URIs instead of the literal, knowing that the term has the same meaning.

5.4.7. Controlled Vocabularies are typically published in SKOS<sup>29</sup> format. That format gives a URI to a list ( known as a 'concept scheme' ) , and a URI to each entry in the list.



Lists of controlled values are typically published using the SKOS format.

5.4.8. The esdToolkit<sup>30</sup> publish all of their lists for types of things like 'service', 'function', 'circumstance' etc as SKOS lists, so that any of their terms can be used in linked data, and can be the basis of links such as

- Who else provides the same type of service?
- What type of service addresses a type of circumstance?

5.4.9. If you were to look up the URIs in this example triple, you would find that this statement says that a site has a status of "NOT STARTED".

5.4.10. Here is another triple

<sup>29</sup> <http://en.wikipedia.org/wiki/SKOS>

<sup>30</sup> <http://standards.esd.org.uk/>

`<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>`

`<http://www.w3.org/2003/01/geo/wgs84_pos#lat>`

`"50.795714"`

5.4.11. Notice that this statement is about the same ‘thing’ as the previous statement. If we looked it up, we would find that it relates to a ‘site’ that is within the area governed by Portsmouth Council. This is site ‘0383’ on Portsmouth’s list, and there could also be another site ‘0383’ on another council’s list. As it is Hampshire County Council that have created this URI, they have ensured that it is unique by including both the name of the council, and the site number into the URI string.

5.4.12. If Portsmouth Council had created the URI, they may have used

`<http://data.portsmouth.gov.uk/id/land-supply/site/0383>`

... as the URI string, which Hampshire ( and anyone else ) could then have used in their linked data.

5.4.13. Notice also that the second part of the triple is NOT created by the people at hants.gov.uk. This URI is defined by [www.w3.org](http://www.w3.org), which is a well-known domain, with processes to ensure the quality of their content.

5.4.14. Looking that URI up gives

```
<rdf:Property rdf:about="http://www.w3.org/2003/01/geo/wgs84_pos#lat">
  <rdfs:domain rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing" />
  <rdfs:label>latitude</rdfs:label>
  <rdfs:comment>The WGS84 latitude of a SpatialThing (decimal degrees).</rdfs:comment>
</rdf:Property>
```

... which gives a precise definition of the meaning of the property and the type of ‘thing’ that it can apply to. Crucially, this is a URI to a ‘property’ that is likely to appear in other linked data, from other sources.

5.4.15. If you were to find another linked data statement that also used this URI to define latitude, and which had the same value, then ( assuming you did the same for longitude ) you can be confident that it is referring to the same location. Whereas a statement with the same value, that used a URI of say

<http://make-it-up-as-you-go-along.com/lat>

or

<http://unclear.com/y>

... may not be referring to the same place



We said before: Links are made between otherwise separate datasets when they use the same URIs.

5.4.16. This takes us closer to 5\* data, in that it is using a URI that was not created from within the original dataset.



The companion web site will invite you to propose the

- ‘properties’,
- URI Sets,
- controlled vocabularies

that should be used for the ‘types of thing’ that we commonly describe in local public data.

5.4.17. Here is another triple

`<http://linkeddata.hants.gov.uk/id/land-supply/site/portsmouth/0383>`

`<http://linkeddata.hants.gov.uk/def/land-supply/planningAuthority>`

`<http://opendatacommunities.org/id/unitary-authority/portsmouth>`

5.4.18. If we ( or more likely a machine ) looked up these URIs, then we would find that the statement is saying that the ‘site’ has a ‘planning authority’ of Portsmouth Council. However, rather than create their own URI for Portsmouth Council, ( or worse, use a literal of “Portsmouth Council” ), they have re-used a URI that is defined by the people at <http://opendatacommunities.org>.

5.4.19. <http://opendatacommunities.org> is the linked data publishing point for the Department of Communities and Local Government (DCLG), and so is a trusted source of quality URIs. They provide a URI Set that identifies each local authority in England.

5.4.20. Looking up <http://opendatacommunities.org/id/unitary-authority/portsmouth> gives

#### OPEN DATA COMMUNITIES

Our data Theme: Geography Dataset: Local a... Resource

resource

### Portsmouth

uri

<http://opendatacommunities.org/id/unitary-authority/portsmouth>

predicate	object
<a href="#">billing authority code</a>	E1701
<a href="#">government office region</a>	South East
<a href="#">governs</a>	<a href="#">The City of Portsmouth</a>
<a href="#">GSS</a>	E06000044
<a href="#">has census code</a>	00MR
<a href="#">label</a>	Portsmouth
<a href="#">Openly Local URL</a>	<a href="http://openlylocal.com/councils/353-Portsmouth-City-Council">http://openlylocal.com/councils/353-Portsmouth-City-Council</a>
<a href="#">page</a>	<a href="http://www.portsmouth.gov.uk/">http://www.portsmouth.gov.uk/</a>

...which therefore gives much more information about that council.

## 5.5. Step 4:Publish and query

5.5.1. Hampshire County Council provide a SPARQL endpoint to this data.

## SPARQL query

```
SELECT DISTINCT ?address ?status WHERE {
  ?site
  <http://www.ordnancesurvey.co.uk/ontology/BuildingsAndPlaces/v1.1/BuildingsAndPlaces.owl#hasAddress> ?address .
  ?site <http://linkeddata.hants.gov.uk/def/land-supply/siteStatus> ?status
  FILTER ( ?status = "UNDER CONSTRUCTION" )
}
```

5.5.2. This query asks for the addresses of sites that have a status of “UNDER CONSTRUCTION”, which gives results of

### results

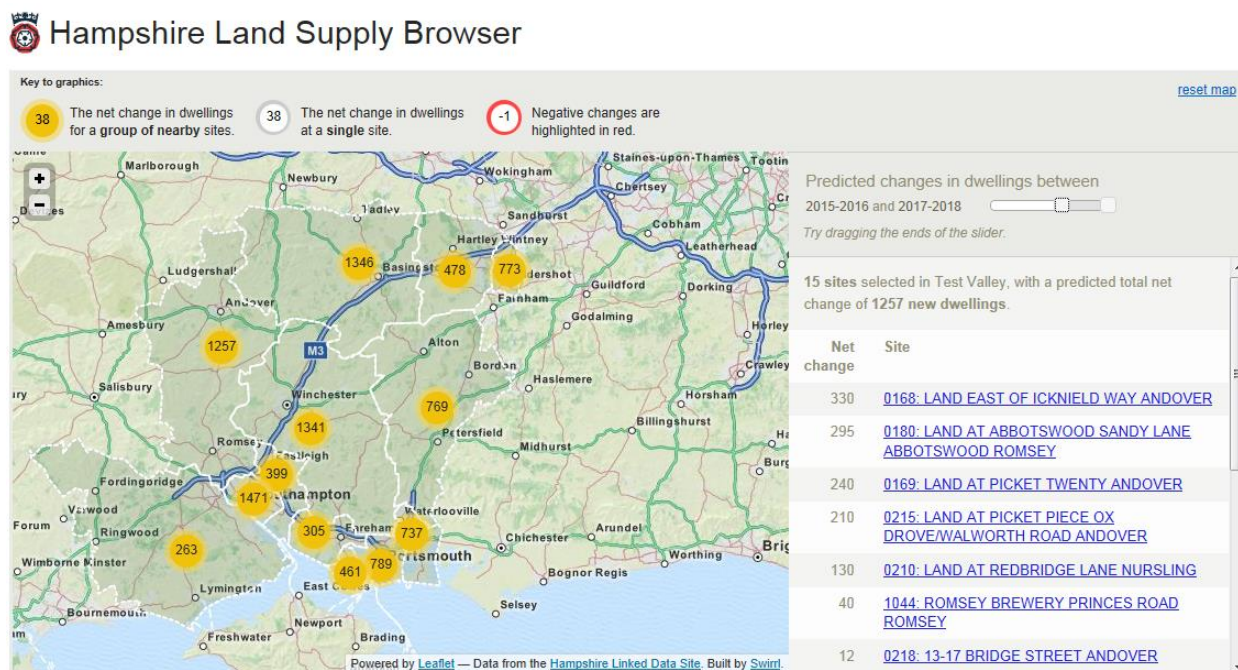
address	status
"LAND OPPOSITE RECREATION GROUND MARLANDS LANE WEST MEON"	"UNDER CONSTRUCTION"
"271-273 FARNBOROUGH ROAD FARNBOROUGH"	"UNDER CONSTRUCTION"
"20-26 COLLEGE STREET AND 29-35 RICHMOND STREET SOUTHAMPTON"	"UNDER CONSTRUCTION"
"THORNERS HOMES REGENTS PARK ROAD SOUTHAMPTON"	"UNDER CONSTRUCTION"
"CAR PARK AND REDUNDANT HIGHWAY WELLINGTON STREET SOUTHSEA PORTSMOUTH"	"UNDER CONSTRUCTION"
"49 THOROLD ROAD SOUTHAMPTON"	"UNDER CONSTRUCTION"
"ST FRANCIS OF ASSISI CHURCH MONTAGUE AVENUE SOUTHAMPTON"	"UNDER CONSTRUCTION"
"REAR OF FORMER BAPTIST CHURCH & THE LA HEATHFIELD ROAD PORTSMOUTH"	"UNDER CONSTRUCTION"
"AREAS 4 AND 7 HUNTS POND ROAD FAREHAM"	"UNDER CONSTRUCTION"
"THE BOURNE BOURNE FIELDS TWYFORD WINCHESTER"	"UNDER CONSTRUCTION"
"REGENERATION SCHEME FAROE/MALDIVE CLOSE BASINGSTOKE"	"UNDER CONSTRUCTION"
"LAND SOUTH OF WINCHESTER ROAD FOUR MARKS ALTON"	"UNDER CONSTRUCTION"
"FORMER RENAULT GARAGE 128 MILTON ROAD PORTSMOUTH"	"UNDER CONSTRUCTION"
"LINDEN 43 ANSTEY LANE ALTON"	"UNDER CONSTRUCTION"
"59 PROJECT HOUSE AVENUE ROAD PORTSWOOD SOUTHAMPTON"	"UNDER CONSTRUCTION"
"SAVOY BUILDINGS SOUTH PARADE SOUTHSEA PORTSMOUTH"	"UNDER CONSTRUCTION"
"LAND AT HAZEL COTTAGE CHASE ROAD LINDFORD BORDON"	"UNDER CONSTRUCTION"
"COLDEAST HOSPITAL COLDEAST WAY SARISBURY SOUTHAMPTON"	"UNDER CONSTRUCTION"
"121 VINE ROAD & 195-207 COXFORD ROAD SOUTHAMPTON"	"UNDER CONSTRUCTION"
"160-162 FAIR OAK ROAD BISHOPSTOKE"	"UNDER CONSTRUCTION"

... and options for a range of machine-readable formats.



## 5.6. Step 5:Present and Visualise

### 5.6.1. Hampshire County Council provide a web site that presents and visualises the data.



5.6.2. The site allows you to zoom into the map to get more detail, and a slider for the year, to show projections.

5.6.3. Taking these first two examples together, you might ask, “where do we plan to build new houses, where there is already poor air quality”? This would require that you bring together these two sets of linked data, which have never previously ‘met’, and run queries over them. You would want to rely on consistent use of URIs for ‘location’ and ‘time’.



New conclusions can be reached by querying unrelated sets of data, that use consistent URIs to refer to the same thing.

## 5.7. More about this pilot from Hampshire County Council

5.7.1. Hampshire County Council worked with Swirrl<sup>31</sup> to produce this pilot of applying Linked Data to local public data.

5.7.4. You can find this pilot at

<http://linkeddata.hants.gov.uk/land-supply-mapper/map.html>

the resulting presentation of the data;

<http://linkeddata.hants.gov.uk/land-supply-mapper/map.hhttp://linkeddata.hants.gov.uk/sparql>

the SPARQL endpoint;

<sup>31</sup> <http://www.swirrl.com/>



## 6. Example 3 - Open Public Services at Sedgemoor District Council and Devon County Council.

### 6.1. About Open Public Services

- 6.1.1. The Open Public Services White Paper<sup>32</sup>, first issued in July 2011, gives an expectation that people will have easy access to timely information about what services are available and how good they are. This has prompted both Sedgemoor District Council, and Devon County Council to consider how linked data can be used to publish information about objectives, plans, resources, decisions, services, and what we know about the communities that use them.
- 6.1.2. Additionally, Devon County Council are in a local partnership to pilot community neighbourhood budgeting at one of their towns, which will require that information from many sources is brought together so that residents can make informed decisions about the use of resources.
- 6.1.3. These councils are experimenting to learn how/ linked data can provide the data ecosystem necessary to support these aims.

### 6.2. Step 1: Some 'traditional' data

- 6.2.1. Sedgemoor, like all councils, manage a series of programs and projects which are designed to impact on local communities, businesses, and the environment. Sedgemoor use a product called Covalent<sup>33</sup> for this; there are a number of products in use for this purpose across the public sector. Progress on these projects is regularly reported to committees and is published in document form along with committee decisions.
- 6.2.2. The data is not two-dimensional data which would naturally fit a spreadsheet, as it contains milestones, status updates, links to supporting documentation and so on. However, in this case, it is straightforward to extract the data in an XML form. Here is an entry that describes a project to build a new swimming pool.

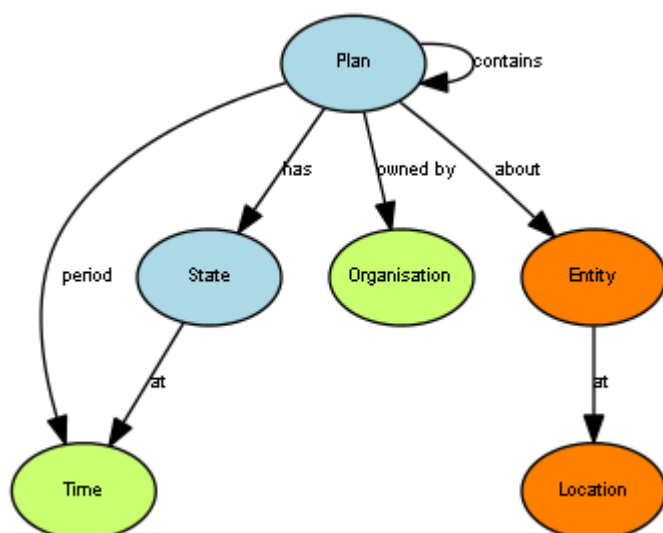
```
<actn_action
  actn_id="2256087"
  actn_parent_id="2255957"
  actn_code="02 Supp Action"
  actn_title="1) To open a new pool in Bridgwater by Spring 2013"
  actn_description="&lt;html&gt;  &lt;head&gt;    &lt;/head&gt;
  actn_due_date="31/03/2013"
  actn_start_date="12/07/2012"
  actn_progress="76"
  actn_active="true"
  actn_subactions="0"
  actn_status="I"
  actn_worst_child_status="U"
  actn_priority="3"
  actn_parent_due_date="31/03/2014"
  actn_parent_start_date="12/07/2012"
  actn_original_due_date="31/03/2013"
  actn_next_update_due="31/03/2013"
  actn_originator_name="Sedgemoor"/>
```

### 6.2. Step 2: Modelling the Data

- 6.2.3. Looking for 'types of things' and relationships in this data, draws out a number of opportunities for linking.

<sup>32</sup> <http://files.openpublicservices.cabinetoffice.gov.uk/OpenPublicServices-WhitePaper.pdf>

<sup>33</sup> <http://www.covalentsoftware.com/>



A Plan	A plan is often contained in another plan, and will have an expected period of time over which it will run. .
State	From time to time, a snapshot of the progress of a plan is taken and reported, which can build into a progress history.
Organisation	A plan will be owned by an Organisation.
Entity and Location.	<p>In this case, the ‘things’ that each plan is about, are NOT contained in the data. In this example, the plan to build a swimming pool, does not identify the swimming pool itself. We will need to add extra triples later to link to URIs that define the things ( entities ) that the plans are about.</p> <p>Often those ‘entities’ will have a location, which again can be a useful link and aid to visualisation.</p> <p>We have colour code these ‘types of thing’ in orange to show that the data does not yet exist.</p>

6.2.4. If our model defines new ‘types of things’ and relationships, that have not been defined before, then we will also need to create a machine-readable view of the model.

6.2.5. The machine-readable view of a model is itself expressed as a series of triples. This example uses URIs from the RDF Schema<sup>34</sup> (rdfs) vocabulary.

6.2.6. Where in the guide we have referred to a ‘type of thing’, rdfs uses the term ‘Class’. You can also see that the rdf vocabulary defines the idea of a ‘Property’ that can be associated with a Class to express values or links.

6.2.7. The rdfs vocabulary can be used to define the classes that a property can be related to as

- the domain ( for the first part of a triple )
- the range ( for the third part of a triple )

```
:Plan
  a      rdfs:Class ;
  rdfs:subClassOf  pscm:Plan ;
  rdfs:label      "Plan" ;
  rdfs:comment    "" ;
  .

:plnContainedInPlan
  a      rdf:Property;
  rdfs:domain  :Plan;
  rdfs:range   :Plan;
  rdfs:label   "contained in"@en;
  rdfs:comment "The parent plan of a plan" ;
  .

:plnOwner
  a      rdf:Property;
  rdfs:domain  :Plan;
  rdfs:range   :Party;
  rdfs:label   "owned by"@en;
  rdfs:comment "The Party that owns the plan" ;
  .

:plnPeriod
  a      rdf:Property;
  rdfs:domain  :Plan;
  rdfs:range   :Time;
  rdfs:label   "on"@en;
  rdfs:comment "The Period over which the plan applies" ;
  .
```



A machine-readable view of classes and properties is known as a

- Linked Data Vocabulary
- Ontology



Linked data vocabularies are themselves defined in linked data using vocabularies such as

- rdf
- rdfs
- owl ( the web ontology language )

6.2.8. In general, we will get more linking value from local data when we use linked data vocabularies that are commonly used at a national or international scale. Consequently, we should avoid creating new vocabularies for a single use by a single local authority, but work together to either

- Agree on the best vocabulary to use to describe a common concept in local data

<sup>34</sup> <http://en.wikipedia.org/wiki/RDFS>

- Create new vocabularies collaboratively and promote them for use across local public data



There are good materials and tutorials about how to create linked data vocabularies.

6.2.9. There are a few sites on the web that list linked data vocabularies, such as

- <http://lov.okfn.org/>
- <http://www.schema.org/>



The companion web site will invite you to propose the linked data vocabularies that we should consistently use, or that should be created, so that we can describe local public data.

## 6.3. Step 3: Transform the data

6.3.1. An extract from the linked data version of the XML file looks like .

```
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/Plan> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/InSet> <http://data.pauldcavidson.com/id/ops/set/4702> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/plnOwner> <http://data.pauldcavidson.com/id/ops/organisation/19> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/identifier> "2353773" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/title> "PR 41 - Bridgewater Swimming Pool" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/description> <p> In progress </p> <p> </p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/plnPeriod> <http://data.pauldcavidson.com/id/ops/organisation/19/plan/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/State> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/time/1> <http://data.pauldcavidson.com/def/ops/timeStart> "2012-10-02" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/time/1> <http://data.pauldcavidson.com/def/ops/timeEnd> "2013-03-31" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/plnContainedInPlan> <http://data.pauldcavidson.com/id/ops/organisation/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/plnState> <http://data.pauldcavidson.com/id/ops/organisation/19/plan/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/699304> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/State> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/699304> <http://data.pauldcavidson.com/def/ops/staPeriod> <http://data.pauldcavidson.com/id/ops/organ/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/699304/time/1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/T>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/699304/time/1> <http://data.pauldcavidson.com/def/ops/timeInstant> "2012-07-31 14:48:04" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/699304> <http://data.pauldcavidson.com/def/ops/description> <html> <head> </head> <body> <p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/699304> <http://data.pauldcavidson.com/def/ops/staProgress> ""0%"" </body> </p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/plnState> <http://data.pauldcavidson.com/id/ops/organisation/19/plan/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/728136> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/State> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/728136> <http://data.pauldcavidson.com/def/ops/staPeriod> <http://data.pauldcavidson.com/id/ops/organ/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/728136/time/1> <http://data.pauldcavidson.com/def/ops/timeInstant> "2012-10-02 09:09:39" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/728136> <http://data.pauldcavidson.com/def/ops/description> <html> <head> </head> <body> <p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/728136> <http://data.pauldcavidson.com/def/ops/staProgress> ""0%"" </body> </p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/id/ops/organisation/19/plan/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529> <http://data.pauldcavidson.com/def/ops/plnState> <http://data.pauldcavidson.com/id/ops/organ/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529> <http://data.pauldcavidson.com/def/ops/staPeriod> <http://data.pauldcavidson.com/id/ops/organ/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529/time/1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/T>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529/time/1> <http://data.pauldcavidson.com/def/ops/timeInstant> "2012-11-08 10:17:20" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529> <http://data.pauldcavidson.com/def/ops/description> <html> <head> </head> <body> <p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/7448529> <http://data.pauldcavidson.com/def/ops/staProgress> ""0%"" </body> </p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcavidson.com/def/ops/plnState> <http://data.pauldcavidson.com/id/ops/organisation/19/plan/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/762731> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/State> .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/762731> <http://data.pauldcavidson.com/def/ops/staPeriod> <http://data.pauldcavidson.com/id/ops/organ/>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/762731/time/1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://data.pauldcavidson.com/def/ops/T>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/762731/time/1> <http://data.pauldcavidson.com/def/ops/timeInstant> "2013-01-14 11:30:51" .
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/762731> <http://data.pauldcavidson.com/def/ops/description> <html> <head> </head> <body> <p>
<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765/state/762731> <http://data.pauldcavidson.com/def/ops/staProgress> ""0%"" </body> </p>
```

6.3.2. Picking out some triples

<http://data.pauldcavidson.com/id/ops/organisation/19/plan/3765>

<http://data.pauldcavidson.com/def/ops/plnOwner>

<http://data.pauldcavidson.com/id/ops/organisation/19>

6.3.3. Notice that the domain for these URIs is at “pauldcavidson.com”; definitely not a trustworthy source. We have kept these URIs away from the .gov.uk domains as we experiment. We would not want other linked data to use them, yet.

6.3.4. The property that says who owns this ‘plan’ is taken from the linked data vocabulary that we created earlier. There will be better URIs to defines things like ‘plans’, perhaps in this case from StratML, but we can make that sort of decision later in the experiment. The intension is

that the companion web site will build into a list of good linked data vocabularies to use for purposes like this.

6.3.5. The owner of the plan is described using a local URI, ( if you look it up you find that it is referring to Sedgemoor District Council ) whereas it would be better to use a URI from a national URI Set.

6.3.6. Here is another triple

```
<http://data.pauldcdavidson.com/id/ops/organisation/19/plan/3765 >
```

```
<http://data.pauldcdavidson.com/def/ops/title>
```

```
"PR 41 - Bridgwater Swimming Pool"
```

6.3.7. The 'title' property would perhaps be better if it came from the 'dublin-core'<sup>35</sup> linked data vocabulary, or perhaps the rdfs vocabulary for its 'label property, or perhaps from skos. These again show the opportunities that we have to use URIs and vocabularies consistently across local data to increase the chances of useful links being made.

6.3.8. Here are two triples ...

```
<http://data.pauldcdavidson.com/id/ops/organisation/19/plan/3765>
```

```
<http://data.pauldcdavidson.com/def/ops/plnState>
```

```
<http://data.pauldcdavidson.com/id/ops/organisation/19/plan/3765/state/7627317>
```

... and ...

```
<http://data.pauldcdavidson.com/id/ops/organisation/19/plan/3765/state/7627317>
```

```
<http://data.pauldcdavidson.com/def/ops/description>
```

```
"Sedgemoor District Council have announced that the new swimming pool and cafe  
at Trinity Sports and Leisure will open its doors to the public from: 11:30am on  
Friday 1st March 2013."
```

6.3.9. The first triple links a plan to a status update for the plan. The URI for the plan, and for the status update, both use natural keys from the database, to ensure that they are unique. Giving a new URI to a status update enables

- further triples to make statements about the status such as dates, %completion, description;
- data from other sources to make statements about the progress on the plan.

#### 6.4. Step 4: Publish and query

6.4.1. This data has been loaded to a triplestore and can be queried using SPARQL

<sup>35</sup> <http://dublincore.org/documents/dc-rdf/>

```
@PREFIX ops-def: <http://data.pauldcdavidson.com/def/ops/>

SELECT DISTINCT ?plantitle ?statusdesc ?statusdate WHERE {

    ?plan rdf:type ops-def:Plan .
    ?plan ops-def:title ?plantitle .
    ?plan ops-def:plnState ?status .
    ?status ops-def:description ?statusdesc .
    ?status ops-def:staPeriod ?statusperiod .
    ?statusperiod ops-def:timeInstant ?statusdate .

}
```

Change HTTP method: [GET](#) [POST](#)

Send Query

Reset

6.4.2. This query asks for the plans and their statuses, giving a machine readable response including

```
<?xml version="1.0"?>
<result>
  - <binding name="plantitle">
    <literal>PR 68 - West Quay Wall
      Repairs</literal>
  </binding>
  - <binding name="statusdesc">
    <literal><html> <head> </head> <body>
      <p> Works progressing within budget.
        Major reconstruction complete, facings
        etc still to be done. awaiting final results
        re pedestrianisation. (A Griffin 16-01-13)
      </p> </body> </html></literal>
  </binding>
  - <binding name="statusdate">
    <literal>2013-01-16 14:19:50</literal>
  </binding>
</result>
```

## 6.5. Step 5:Present and Visualise

6.5.1. We have created a simple view of that data which also shows the SPARQL queries that are being used to get the data.



## Progress Updates

Date	Description
14/01/2013	Sedgemoor District Council and 1610 have announced that the new swimming pool and café at Trinity Sports and Leisure will open its doors to the public from: 11:30am on Friday 1st March 2013.
08/11/2012	Still on target for completion and handover early Feb 2013
02/10/2012	Swimming on target for handover in early Spring.
31/07/2012	Next steps identified at Member briefing: <ul style="list-style-type: none"> <li>•Increase publicity in run-up to opening and continue to liaise and involve Pool Task and Finish group</li> <li>•Make final design decisions</li> <li>•Opening ceremony (and first public swims) are held by 4th March 2013.</li> </ul>

**Get the data for this page**

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?uri ?p7 ?p9 ?p11 WHERE {

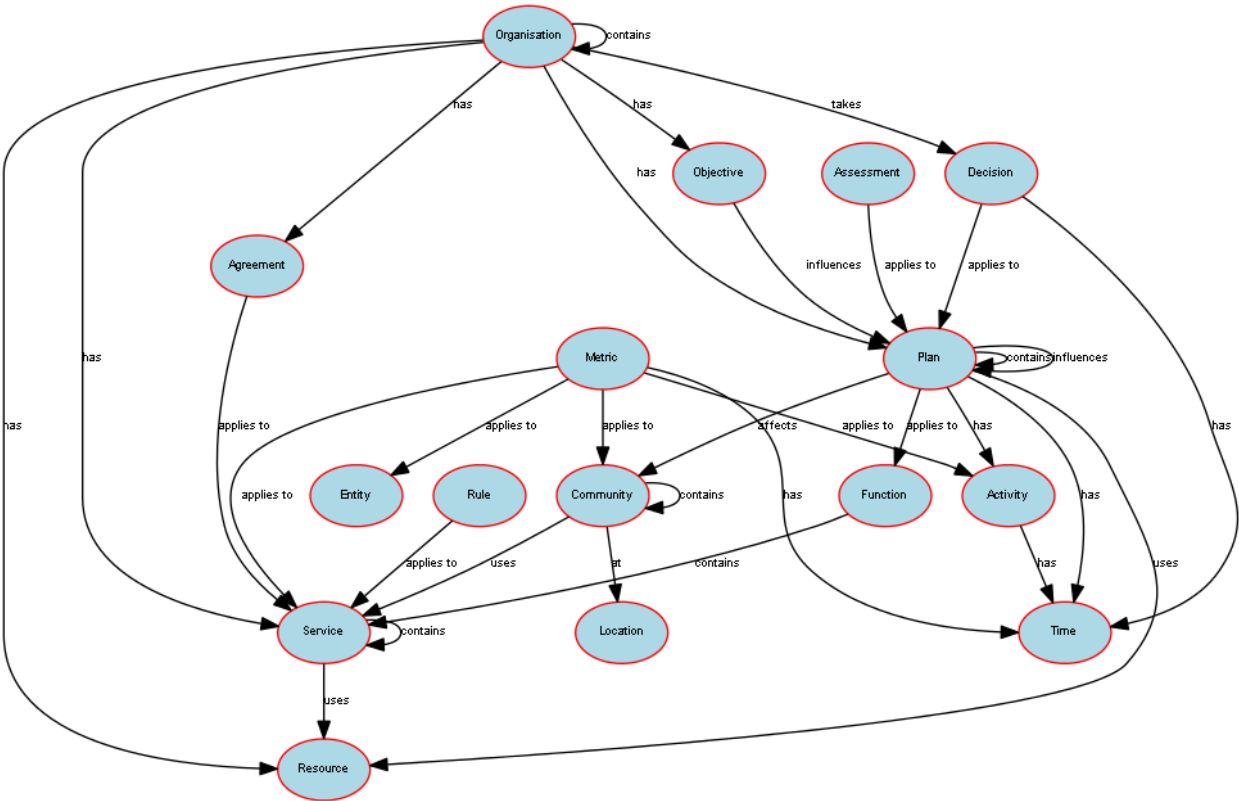
<http://data.pauldcdaavidson.com/id/ops/organisation/19/plan/3765> <http://data.pauldcdaavidson.com/def/ops/plnState> ?uri .
?uri rdf:type <http://data.pauldcdaavidson.com/def/ops/State> .
OPTIONAL { ?uri <http://data.pauldcdaavidson.com/def/ops/staPeriod> ?p7 .
OPTIONAL { ?p7 <http://data.pauldcdaavidson.com/def/ops/timeInstant> ?p9 . } }
OPTIONAL { ?uri <http://data.pauldcdaavidson.com/def/ops/description> ?p11 . }
}

order by DESC(?p9)

```

## 6.6. More about this pilot from Sedgemoor and Devon

6.6.1. At Devon County Council, they have thought through the information that could come from each of the many organisations that serve their communities.



### 6.6.2. The example above just showed the ‘Organisation has Plan’ link.

6.6.3. In the past, a lead authority would have tried to warehouse this type of information from many sources and then perhaps showed it at its website. However, copying and keeping that



information up-to-date takes effort, and often goes wrong, leading to poor information being passed on.

6.6.4. Linked data provides an opportunity where each organisation can express just its own information, which will therefore always be accurate and up-to-date, which can be combined later by anyone about any location, theme, community, audience, etc.

6.6.5. The model, allows organisations to state and link their objectives, plans, decisions, and so on. They can say and share what they know about the communities that they serve. Those communities can simply be the residents of a town, but more usefully could be segmentations such as ...

- the long term unemployed
- carers
- commuters
- families with young children
- the tourist industry

... and so on.

6.6.6. The LGINform<sup>36</sup> service from the Local Government Association (LGA) is soon to provide an 'api' onto 'metrics' data for each local authority.

6.6.7. Typically, public sector organisations have computer systems or content management facilities that record this type of information, for example

Class	System
Resources	Financial Management
Decisions	Committee Minutes
Plans	Performance Management
Services	Customer Relationship Management (CRM)
Agreements	Contracts Register
Community Metrics	Local Information Systems

6.6.8. Local public organisations allocate their resources over their services and their improvement plans, which are designed to have an observable impact on their communities.

6.6.9. All of this could usefully be expressed as linked data, and become the basis of how we transform local services to meet the needs of local people.



The data and endpoint from this example will be directly available from the companion web site.

## 7. Next Steps

7.1. We hope that you have enjoyed this LeGSB guide and perhaps have a better understanding of the technical implications, and potential for transformation, of publishing local data to 4\* and 5\* standards.

<sup>36</sup> <http://www.local.gov.uk/about-lginform>

7.2. We now need to work together across local government to fill in some of the gaps that we have highlighted, and to make some collective design decisions to get the best from our data, and to make the case for moving forward.

7.3. The web site picks up the themes raised here, and invites you to contribute to

- what 'concepts' commonly recur in local public service data?
- for those concepts, which properties/URI Sets/controlled vocabularies should we consistently use?
- what existing linked data is already available that could be linked to?
- what 'core reference data' is missing, which we will need to be able to make links in linked data?

7.4. Similarly, the web site will build further information on technical issues that are emerging from the linked data community, including

- options for modelling;
- data formats to respond with when dereferencing URIs;
- options for triplestores;
- describing quality and provenance;
- patterns for common scenarios such as statistics;

7.5. Much of our data is contained in computer systems, so LeGSB will continue to work with suppliers of solutions to local authorities to embrace data publishing, and linked data in particular; to embed these facilities into their products.