Publishing local 5 star data.

- A Technical Companion to the DCLG ‘Code of Recommended Practice for Local Authorities on Data Transparency

- A ‘sandbox’ for Open Public Services

- Collaborating via Profiles for Linked Data

_Paul Davidson, The Local e-Government Standards Body (LeGSB), and CIO of Sedgemoor District Council_

To: Socitm South West, 21st June 2013, Exeter
LeGSB’s aim

To promote eStandards for Efficiency, Transformation, and Transparency of Local Services

- A Standard that supports all three, is particularly attractive
  - Predominantly about data standards
  - … and in particular, Open Data Standards

- Standards that can be re-used across many …
  - Disciplines / Sectors – e.g. Welfare, Health, Justice
  - Information Uses – e.g. Interoperability, public open data, evidencing policy

- Standards that local public services need …
  - As directed by our sponsors

- Standards that build into an architecture
  - e.g. a data ecosystem
• Publication should be in open and machine-readable formats. The recommended 5 step journey to a fully open format is:
  – * Available on the web (whatever format) but with an open license
  – ** As for one star plus available as machine-readable structured data (e.g. Excel instead of image scan of a table)
  – *** As for two star plus use a non-proprietary format (e.g. CSV and XML)
  – **** All the above plus use open standards from the World Wide Web Consortium (such as RDF and SPARQL) and
  – ***** All the above plus link your data to other people’s data to provide context.
<table>
<thead>
<tr>
<th>Star Rating</th>
<th>In practice</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>This is typically a static document, perhaps containing data as tables. Often this would be web page as html, a document as pdf, or an image.</td>
<td>• easy to produce; • retains presentation and layout;</td>
<td>• cannot further manipulate the data such as sorting, filtering, summing etc; • cannot join or compare to other data, or earlier versions;</td>
</tr>
<tr>
<td>2</td>
<td>This is typically data, such as a spreadsheet, published in the format of the tool that was used to extract it, such as xls</td>
<td>• no new tools or skills necessary; • data can be downloaded and further processed and analysed.</td>
<td>• assumes that the consumer has the same tool as the producer, or is able to use the format. • very large data sets might not be attractive to download.</td>
</tr>
<tr>
<td>3</td>
<td>This is typically 2* type data published using an open format, such as csv or xml.</td>
<td>• there is typically a choice of open source tools available for each open format</td>
<td>• Some conversion necessary. • Where data has greater structure than a simple 2-dimensional table, many files may be necessary, which may become disconnected; • The meaning and scope of columns and rows can be hard to express.</td>
</tr>
<tr>
<td><strong>Publishing up to step 5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This is explicitly about publishing data to the RDF data model, and providing a query service using the SPARQL language.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• enables others to make statements over the web about individual lines of data;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• can be queried over a data service so that a complete data set does not have to be downloaded;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• gives precise definitions to the meaning of the data.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• requires a skill set that most local authorities do not have yet;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• not suitable to provide directly to residents.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requires some infrastructure</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| This is Linked Data in the RDF model, that contains links to external datasets to describe the ‘things’ that the data refers to. |
| • adds context; |
| • can build into a ‘data ecosystem’. |
| • can enable 3rd parties to join data from many sources together to provide new targeted information services and insight. |
| requires a ‘spine’ of core reference data to make links to. |
Why do we need this guide?

- 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.

  “Understanding this document will allow your organisation to be at the forefront of the open linked data agenda melding it to be of use to your citizens and your organisation.”
  
  Roger Hampson, Chief Executive LB Redbridge

a really good piece of work - simple to read, comprehensive and helpful.

Jos Creese, CIO Hampshire C.C.
Get the guide

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

• Explanations of Linked Data concepts and techniques

• Examples of local authorities publishing 5* data
  – Bristol City Council – Air Quality Data
  – Hampshire County Council – Land Use Projections
  – Devon County Council – Community Neighbourhood Budgeting

• Step by step walk through of
  – Modelling
  – Transformation into RDF
  – Querying
  – Building applications

• Draws out ‘gaps’
  – which ‘concepts’ commonly recur in local public service data?
  – 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?
# Bristol City Council – Air Quality Data

- **Get 3* Data**

<table>
<thead>
<tr>
<th>SiteID</th>
<th>Date</th>
<th>Time</th>
<th>NO2raw_ppb</th>
<th>NOxraw_ppb</th>
<th>NOraw_ppb</th>
<th>NO2rat_ppb</th>
<th>NOXrat_ppb</th>
<th>NOrat_ppb</th>
<th>CAQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>01/01/2001</td>
<td>00:30:00</td>
<td>13.5</td>
<td>65</td>
<td>50.5</td>
<td>4.73</td>
<td>19.44</td>
<td>14.71</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>01/01/2001</td>
<td>00:45:00</td>
<td>13</td>
<td>55.5</td>
<td>41.5</td>
<td>4.58</td>
<td>16.57</td>
<td>11.99</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>01/01/2001</td>
<td>01:00:00</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>01/01/2001</td>
<td>01:15:00</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>01/01/2001</td>
<td>01:30:00</td>
<td>12.5</td>
<td>46</td>
<td>32.5</td>
<td>4.43</td>
<td>13.7</td>
<td>9.27</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>01/01/2001</td>
<td>01:45:00</td>
<td>12.5</td>
<td>36.5</td>
<td>23.5</td>
<td>4.28</td>
<td>10.83</td>
<td>6.55</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>01/01/2001</td>
<td>02:00:00</td>
<td>11</td>
<td>54.5</td>
<td>43</td>
<td>3.82</td>
<td>16.26</td>
<td>12.44</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>01/01/2001</td>
<td>02:15:00</td>
<td>9</td>
<td>44</td>
<td>35.5</td>
<td>2.92</td>
<td>13.09</td>
<td>10.17</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>01/01/2001</td>
<td>02:30:00</td>
<td>10.5</td>
<td>35.5</td>
<td>24</td>
<td>3.82</td>
<td>10.52</td>
<td>6.7</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>01/01/2001</td>
<td>02:45:00</td>
<td>12</td>
<td>34</td>
<td>22</td>
<td>3.97</td>
<td>10.07</td>
<td>6.09</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>01/01/2001</td>
<td>03:00:00</td>
<td>11.5</td>
<td>32</td>
<td>20</td>
<td>3.97</td>
<td>9.47</td>
<td>5.49</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>01/01/2001</td>
<td>03:15:00</td>
<td>11.5</td>
<td>40.5</td>
<td>28</td>
<td>4.13</td>
<td>12.03</td>
<td>7.91</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>01/01/2001</td>
<td>03:30:00</td>
<td>13</td>
<td>32</td>
<td>18</td>
<td>4.58</td>
<td>9.47</td>
<td>4.89</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>01/01/2001</td>
<td>03:45:00</td>
<td>11</td>
<td>42.5</td>
<td>31.5</td>
<td>3.67</td>
<td>12.64</td>
<td>8.97</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>01/01/2001</td>
<td>04:00:00</td>
<td>12.5</td>
<td>31.5</td>
<td>18</td>
<td>4.43</td>
<td>9.31</td>
<td>4.89</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>01/01/2001</td>
<td>04:15:00</td>
<td>11</td>
<td>23</td>
<td>11.5</td>
<td>3.82</td>
<td>6.75</td>
<td>2.92</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>01/01/2001</td>
<td>04:30:00</td>
<td>11</td>
<td>17.5</td>
<td>6</td>
<td>3.82</td>
<td>5.09</td>
<td>1.26</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>01/01/2001</td>
<td>04:45:00</td>
<td>11</td>
<td>19</td>
<td>7</td>
<td>3.97</td>
<td>5.54</td>
<td>1.56</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>01/01/2001</td>
<td>05:00:00</td>
<td>10</td>
<td>12.5</td>
<td>2.5</td>
<td>3.37</td>
<td>3.58</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>01/01/2001</td>
<td>05:15:00</td>
<td>8.5</td>
<td>10</td>
<td>1.5</td>
<td>2.92</td>
<td>3.07</td>
<td>0.15</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>01/01/2001</td>
<td>05:30:00</td>
<td>10</td>
<td>12.5</td>
<td>2.5</td>
<td>3.37</td>
<td>3.82</td>
<td>0.45</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>01/01/2001</td>
<td>05:45:00</td>
<td>8.5</td>
<td>11</td>
<td>2.5</td>
<td>2.92</td>
<td>3.37</td>
<td>0.45</td>
<td>1</td>
</tr>
</tbody>
</table>
Bristol City Council – Air Quality Data

- Model it

| Organisation | Bristol City Council.  
|--------------|----------------------  
|              | - What type of organisation is that?  
|              | - What geographic area are they responsible for?  
|              | - What is their air quality strategy?  
|              | - How much do they spend on monitoring air quality?  

| Time and Location | What other events have also occurred around that time and location?  
|------------------|-------------------------------------------------  
|                  | - traffic counts  
|                  | - health incidents?  

<table>
<thead>
<tr>
<th>Who lives at that location?</th>
</tr>
</thead>
</table>
|                             | - population  
|                             | - demographics  
|                             | - deprivation  

| Pollutant | What do we know about the pollutant?  
|-----------|-----------------------------------  
|           | - Safe levels?  
|           | - Health implications  

**Diagram:**

- Organisation
- Measurement
- Pollutant
- Time
- Location
- Pollutant
Bristol City Council – Air Quality Data

- Transform it
Bristol City Council – Air Quality Data

• Triples

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"
Bristol City Council – Air Quality Data

• Triples

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
Help us to improve the standards and techniques

The LeGSB guide raises topics where we provide more information, invite local authorities and linked data expertise to collaborate on

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

The guide demonstrates that 5* data publishing requires external data to link to – made up of definitions of concepts, and identifiers for ‘things’. To create a data ecosystem, we need to collaborate to determine

- which ‘concepts’ commonly recur in local public service data?
- 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?

[http://legsb.i-network.org.uk/resources/publishinglocal5stardata/](http://legsb.i-network.org.uk/resources/publishinglocal5stardata/)
Making local links

- **OPERATIONAL** - Data about real people and places, with real needs and circumstances, using real services, i.e. case work

- **STATISTICAL** - Aggregated operational data – organised using common classifications and segmentations

- **ANALYTICAL** - The conclusions drawn from an analysis of statistical data

- **POLITICAL** - The decisions taken to shape services, e.g. budgets, strategies, priorities, targets etc.
Joining it up?
Sand Box

- http://www.openpublicdata.com

Open Public Data - edit Item

ops:Agreement

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>23</td>
</tr>
<tr>
<td>name</td>
<td></td>
</tr>
<tr>
<td>type</td>
<td></td>
</tr>
<tr>
<td>category</td>
<td></td>
</tr>
<tr>
<td>description</td>
<td></td>
</tr>
<tr>
<td>estimated total contract value</td>
<td></td>
</tr>
<tr>
<td>start date</td>
<td></td>
</tr>
<tr>
<td>end date</td>
<td></td>
</tr>
</tbody>
</table>

```
<http://data.sedgemoor.gov.uk/id/contracts/agreement/23> a ops:Agreement ;
    skos:prefLabel """"Electricity Supply"""" ;
    ops:classification """"0931000 - Electricity"""" ;
    dcterms:type """"Contract"""" ;
    ops:estimatedtotalcontractvalue """"255000"""" ;


<http://data.sedgemoor.gov.uk/id/contracts/period/23>
    a ops:Period ;
    ops:periodStart """"2012-12-01"""" ;
    ops:periodEnd """"2015-11-30"""" ;


<http://data.sedgemoor.gov.uk/id/contracts/organisation/23>
    a org:Organization ;
    skos:prefLabel """"Eon UK plc,"""" ;
```
Profiles?

- Working with W3C – RDF Validation
  
  - To define how a dataset has used classes, properties, uri-sets etc, from many ontologies/sources.
  
  - To invite others with similar data to use the same profile, and therefore find that their data can be queried together.
  
  - To be able to register a dataset so that it can be discovered by its ‘Profile’.
  
  - To be able to ‘query’, and present a dataset (or across many datasets) by referring to its ‘profile’
  
  - To be able to validate that a dataset does conform to a ‘profile’
  
  - To enable others to create a new profile by adapting an existing one.
  
  - To be able to create a register of good practice ‘patterns’ for commonly recurring data structures, that can then be re-used.
Legsb
The Local e-Government Standards Body

paul.davidson@sedgemoor.gov.uk

legsb.i-network.org.uk