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: UkGovLd, 19th June 2013, Manchester
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.
# Publishing up to step 3

<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(^{11}) data model, and providing a query service using the SPARQL(^{12}) 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 3\(^{rd}\) 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.
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?
Collaborating

Publishing Local 5* Data

This LeGSB guide is currently a draft. Various Linked Data experts have checked it for us and we have been able to improve it as a result. The final version will be available soon when we get endorsements and recommendations from other organisations, but for now, here is the draft.

LeGSB-PublishingLocal5StarData-Rev3

Please give us feedback to this document using the form at the bottom of the page. You will see that we have also provided further pages to discuss some of the issues and ‘gaps’ that the guide highlights.

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.
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?
http://www.openpublicdata.com
Profiles?

- 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

www.legsb.gov.uk