

Title	Publishing Linked Data for UK Local Government
Creator	Paul Davidson, CIO Sedgemoor District Council, for the Local e-Government Standards Body (LeGSB)
Date.issued	2009-11-10
Description	A discussion paper to consider the activities and roles necessary to progress Linked Data for Local Government.
Status	Revision 5
Subject	Information management
Type	Discussion Paper

Change Control

Revision	Date	Author	Notes
4	10 th November 2009	Paul Davidson	For LGA Linked Data Meeting 10 th November 2009.
5	7 th December	Paul Davidson	Incorporates amendments from Office of Public Sector Information (OPSI)

Purpose

This paper captures the thinking of the Local e-Government Standards Body (LeGSB) Executive Steering Board meeting on the 28th October 2009, in the role that it can take in making progress with Linked Data for Local Government.

Background to Linked Data and Making Public Data Public

Linked Data¹ (sometimes called ‘Linked Open Data’) refers to a method of publishing ‘raw’ information over the web in a style that enables and encourages others to join it up with other related information, to create new views and services. It is at the heart of the Semantic Web which proposes a step change from a web of documents, to a web of data. The step change is the ability to ‘link’ data rather than just discover related documents.

Typically, the UK public sector makes a wealth of information available via the web, however, that information is

- scattered across many government web sites
- contained in documents so that each individual piece of information is not independently available
- published in a range of formats such as PDFs and spreadsheets

A Linked Data approach would publish each separate statement from a document or list, in a standard format that machines can interpret. Where a statement refers to a ‘thing’ that a separate source of information might also refer to, common reference data and definitions of terms can be used to ‘link’ the data.

In June 2009, Sir Tim Berners-Lee was appointed to head a panel to advise how to make non-personal public data as widely available as possible. This project is now known as ‘Making Public Data Public’, and is focused on using Linked Data techniques.

The Department for Communities and Local Government have been asked to declare which data sets they are going to make available by December 2009, which includes Local Authority performance data.

The Data that Local Authorities can make available

The Communities and Local Government Secretary, John Denham has recently said ...

¹ Linked Data – See <http://www.w3.org/DesignIssues/LinkedData.html>

"Public data can ensure consistency - so that residents in Hackney can compare their services with Kent and Lancashire as easily as with their neighbours in Haringey. Making public data public should not be optional for local authorities - it is essential to their role as accountable leaders in modern, democratic institutions".

"Digitalising council records could mean that citizens can find out everything from the council accounts to the number of streetlights and community wardens, to when the rubbish is collected and the hedges trimmed,"

The 'Non-Personal' information that Local Authorities hold that might be of interest to the public and businesses falls into categories of: ...

Democratic

For example

- Which councillors, from which party, represent which ward?
- What committees have what responsibilities?
- Which councillors are members of which committees?

Accounts

For example

- What is the budget and actual expenditure on a service, or project etc
- What does the Council spend its money on and where does it come from?
- How much money is raised from parking fees and fines?

Service

For example

- When is the rubbish collected for my property?
- When will the grass in the village square be cut next?
- What services are provided by which organisation to meet a need or circumstance?

Performance

For example

- How quickly and accurately does the council handle applications for Housing Benefit?

Context and Place

For example

- What is the population of the area?
- What are the demographics of the area?
- How many anti-social-behaviour orders were issued in the last 12 months?
- How many single people are on the Housing Waiting list?

Operational

For example

- Where are the potholes that the council is aware of?
- Where is there an available place in a care home?

Policy and Strategy

For example

- What targets have been set to reduce teenage pregnancies?
- What plans are there to protect biodiversity?

The role that the Local e-Government Standards Body can take

LeGSB can

- facilitate the production of a model of the Data Cloud necessary for Local Government Linked Data, e.g. Services List, Administrative Boundaries
- demonstrate the level of acceptance from the Local Authority Community for candidate entries into the Data Cloud.
- already takes a leading role with the CTO Council and Public Sector standards bodies to determine the design of Linked Data for the Public Sector.

Other roles

Other stakeholders will need to be found to

- own the business case for Linked Data for Local Government.
- engage with the Making Public Data Public project to establish which data Local Government will make available when.
- advise Local Authorities on how to expose the data.
- support the professionalism, skills and tools, to enable Local Authorities to publish their data in this way.
- provide and/or define infrastructure for publishing services
- consider quality assurance of the content for consistency

Other Data Handlers

There are other data handlers, including government departments, who already handle and publish lots of Local Government data; in particular

- financial
- performance
- context

A shared vision is required across these stakeholders of the Local Authority data that can be released in this way.

The esd-Toolkit could create and maintain much of the reference data, and definitions necessary, by re-purposing the vocabularies and logical models that have been devised in conjunction with their subscribing Local Authorities.

The components of Linked Data

Linked Data is most powerful when it can be linked to other Linked Data! To that end, it is founded on 'Common Reference Data' and 'Shared Meaning'.

Linked Data is published on the web for machines to read rather than humans, often using the RDF² data model. RDF breaks a statement down into three parts (so that an RDF statement is known as a 'triple').

- Subject
- Predicate
- Object

² Resource Description Framework - <http://www.w3.org/RDF/>

So for example:

Subject	Predicate	Object
Curry Mallet Primary School	is located within	the District of South Somerset
South Somerset District Council	is located within	the County of Somerset
Curry Mallet Primary School	is a type of	Primary School
Primary Schools in Somerset	achieved	an average score of 4 for Key Stage 2 Mathematics for the School Year 2008/2009
Jim Smith	achieved	a score of 5 for Key Stage 2 Mathematics for the School Year 2008/2009.

The examples above are ‘human-readable’ but the actual RDF representation of this data is in a form that machines can interpret and inference from. Each part of a ‘triple’ is therefore represented by a ‘universal identifier’ rather than the actual text that you see above.

A universal identifier for a school might use the reference provide from Edubase³ via the Department for Children, Schools, and Families. In this way, if two separate sets of data both refer to the same school using the same identifier, the linked data approach could join them.

Universal Identifiers are also used for the predicate part of the statement, so for example if ‘is located within’ is used consistently, data can be joined based on the relationship as well as the subject/object.

In Linked Data, Universal Identifiers are created as URI’s⁴ (Uniform Resource Identifiers). For Linked Data to be effective, a series of sets of reference data needs to be established so that data can be linked to it. The Local e-Government Standards Body (LeGSB) has led work for the Cabinet Office to propose how URI Sets for the Public Sector should be designed so as to give data owners the confidence to reference their data against them.

URI Sets are the central pillars that support linked data. With only a handful of URI Sets, we can suddenly link all sorts of different types of data across the public sector.

The URI for a school might look like ...

<http://education.data.gov.uk/school/12345>

and looking that up (known in the Linked Data world as ‘dereferencing’) might tell you

- The school name
- The address of its prime location
- The name of the head teacher
- The type of school
- The age range
- The religious character
- The capacity

... and it might provide machine-readable links to other URIs such as

- the Local Authority
- the curriculum

³ edubase - <http://www.edubase.gov.uk>

⁴ URI – Uniform Resource Identifier – Defined at <http://tools.ietf.org/html/rfc3986>

This design guidance is available on the Cabinet Office web site, as a part of a wider Public Sector Information Architecture, at

http://www.cabinetoffice.gov.uk/cio/chief_technology_officer/public_sector_ia.aspx

Examples of the Reference Data that Local Authority data will need to link to

Concept	Current source of Identifiers	Readiness for Linked Data
Local Authorities	<p>The SNAC code issued by ONS has been the most popular reference to identify a Local Authority, and other public bodies that cover a geographic region.</p> <p>The Ordnance Survey have issued an RDF 'triple' store of Local Authorities linked to SNAC codes, containing official names, and 'known as' names.</p> <p>The London Gazette, as the official journal of Her Majesty's Government, are the first to publish notices of the creation, deletion, or alteration of public bodies.</p>	<p>The SNAC codes are being made available as URIs by the Making Public Data Public project under the 'statistics' sub domain of data.gov.uk.</p> <p>These would be more properly defined in a 'public-administration' sub domain.</p>
Administrative Boundaries	<p>The Ordnance Survey have issued an RDF triple store of Administrative Boundaries of Public Sector Bodies as a research project. This store contains links to the other Authorities that they border or that they contain, but does NOT contain the actual boundary defined as a polygon, so it is possible to infer which County a District is contained in, but not what Authority a post code is in.</p>	<p>The Making Public Data Public program has set this up a a candidate URI set.</p> <p>Boundary data will be made freely available from April 2010, as part of changes to OS.</p>
Output Areas and Super Output Areas	<p>From ONS. I am not aware that these are available as an online reference list.</p>	
Services	<p>The esd-toolkit have developed a comprehensive set of identifiers for Local Government Services, linked through to the appropriate Powers and Duties that they arise from. These are then linked to the Legislation URIs provided by The Office of Pubic Sector Information (OPSI).</p> <p>There is no Public Sector wide list of services.</p>	<p>Could use the esd-toolkit URIs.</p>
National Performance Indicators	<p>CLG provide this data as spreadsheets.</p>	<p>Department for Communities and Local Government are considering providing this data as a URI set.</p>
Political Parties	<p>The Electoral Commission's register defines Political Parties.</p>	

Demographic segmentations	These might come from SDMX Dimensions, as used by the statistics community, or from 'Customer Insight' categories.	
Locations of Spatial Objects	The Ordnance Survey provide identifiers (TOIDs) to spatial objects. The EU INSPIRE standard is also considering the relevance of URIs.	

Examples of the Reference Data that Local Authority data will want to publish themselves

Councillors
Projects
Partnerships
Committees

Links to the Linked Data Cloud

There are a number of well established sources of reference data that form the basis of the connections that mash-up style sites use. The latest view of these is given at

<http://linkeddata.org/>

The reference data that Local Government links its data to should have corresponding entries in this cloud so that our data can be discovered and linked with data that is not sourced from within the Public Sector.

The most obvious sources of this reference data comes from

- DBPedia – to define terms
- GeoNames – to define places

Paul Davidson
CIO Sedgemoor District Council
And Director, Local e-Government Standards Body (LeGSB)