

Title	Meaningful, Comparable, Linkable – Notes from LG Group workshop 2nd February 2011.
Revision	1
Date	3 rd February 2011
Author	Paul Davidson, LeGSB Director of Standards
Purpose	A discussion paper to raise action points against. To update the LeGSB Executive on progress.
Subject (IPSV)	1526-Information management

Revision	Date	Author
1	03/02/2011	Paul Davidson, for LeGSB

This document

This document captures the views and potential action points that arose at the workshop, 2nd February 2011.

Present at the workshop

Paul Davidson	CIO Sedgemoor D.C., for LeGSB
Stuart Harrison	Lichfield District Council
Gerry McMullen	Birmingham City Council
Gesche Schmid	LGA Analysis and Research
Ingrid Khoeler	LGA
Tim Allen	LGA

Objectives

The objective of the workshop was to

To explore issues and opportunities for the 'local' sector when publishing public open data, and the potential role of common open standards and common infrastructure in making open local data meaningful, comparable and linkable in empowering innovation in local public services and in engaging local people and communities in local decision making.

The group added to that scope

- The ability express data quality and accuracy
- Improving data quality and accuracy by exposing the data
- Confidently and responsibly mixing crowd-sourced data with public open data

Meaningful, comparable, linkable

Understanding what is meant by “meaningful, comparable, linkable”

Meaningful (purposeful)	Data that can be put to good use. Data that can help people to ‘do it for themselves’. Data that can be innovated around.
Meaningful (Contextualised)	Data that has sufficient context to reach useful conclusions. E.g. Cost to maintain roads + length of roads, gives cost per mile of maintaining roads.
Comparable	Capable of being compared. Common Formats. Common Definitions of terms. Ability to combine similar data from various sources.
Linkable	Contains common references to the ‘real-world’ instances of things, or common

vocabulary for segmentations.

Example of local public data

Bristol City Council, are among a number of councils who publish air quality data, and register it at data.gov.uk

air-quality-nox.csv												
	A	B	C	D	E	F	G	H	I	J	K	L
1	Mobile Pie NOx data											
2	-----											
3												
4	Data period : 20/01/2011 11:00 - 21/01/2011 11:00											
5												
6												
7	Date	Time	NO2 Brislington	NOx Brislington	NO Brislington	NO2 NWPS	NOx NWPS	NO NWPS	NO2 Parson St	NOx Parson Stri	NO Parson Stri	NO2 Ruper
8												
9			[ppb]	[ppb]	[ppb]	[ppb]	[ppb]	[ppb]	[ppb]	[ppb]	[ppb]	[ppb]
10	20/01/2011	11:00:00	19.3	43	23.7	33.2	102.9	69.7	38.3	103.5	65.2	62.9
11	20/01/2011	11:15:00	19.5	41.4	21.9	36	95.3	59.3	26.9	65	38.1	61.6
12	20/01/2011	11:30:00	17	37.5	20.4	32.6	90.3	57.8	22.8	58.3	35.5	51
13	20/01/2011	11:45:00	17.9	42.3	24.4	27.9	71.2	43.4	24.2	66.4	42.2	49.1
14	20/01/2011	12:00:00	17.5	37.6	20.1	36.2	103.7	67.6	35.2	104.8	69.7	60.2
15	20/01/2011	12:15:00	16.4	33.2	16.8							45.1
16	20/01/2011	12:30:00	20	47.4	27.4	30.2	82.8	52.6	25.8	66.6	40.8	54.6
17	20/01/2011	12:45:00	14.2	27.7	13.5	34.3	98.7	64.3	17.5	28.7	11.2	55
18	20/01/2011	13:00:00	13.3	28.3	15	43.3	111.9	68.6	18.8	41	22.2	54.6
19	20/01/2011	13:15:00	15.4	30.4	15	35.7	105.9	70.2	23.7	59	35.3	65.1
20	20/01/2011	13:30:00	18.2	37	18.8	35.1	101.3	66.2	21.7	43.6	22	42.3
21	20/01/2011	13:45:00	17.6	35	17.4	29.7	87.9	58.2	24.2	59.3	35.1	54.3
22	20/01/2011	14:00:00	17.6	34.4	16.8	35.4	108.5	73.1	26	61	35	75.3
23	20/01/2011	14:15:00	19	39.2	20.2	46.5	139.5	93	30.6	89.5	58.9	63.2

Above is an extract from one of Bristol City Council's air quality data files in csv format.

A separate csv file, which is also registered at data.gov.uk provides further information about the location of the monitoring stations.

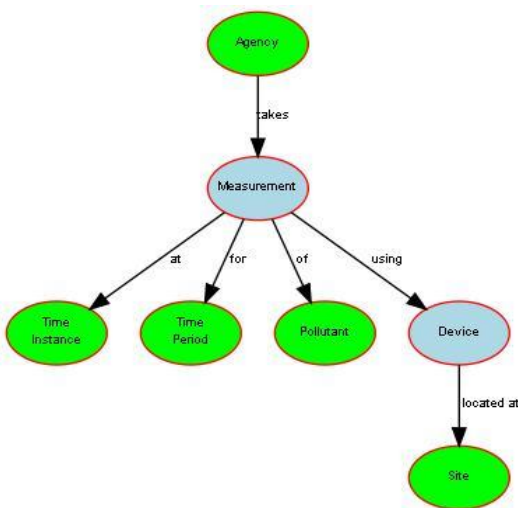
air-quality-monitoring-site-information.csv							
	A	B	C	D	E	F	G
1	Station Name (AQD)	Pollutant	Location Type	station ID (as station filename	Easting	Northing	
2	Bath Road	NO2	Roadside	1 arena_C	360382	171659	
3	CREATE Centre	O3	Background	2 create_C	356838	172209	
4	Brislington Depot	NO2	Background	3 brisl_C	361178	171556	
5	Newfoundland Road Pol	NO2	Roadside	4 Broadmead_C	359644	173681	
6	Parson Street School	NO2	Roadside	5 parsons_C	358053	170593	
7	Rupert Street	NO2	Roadside	6 rupert_C	358667	173108	
8	Shiner's Garage	NO2	Roadside	7 Shiners_C	361000	173352	
9	Cheltenham Road	NO2	Roadside	8 chelt_C	358925	174608	
10	Wells Road	NO2	Roadside	9 wells_C	360903	170024	
11							

LeGSB has established a repeatable checklist approach to proposing linked data standards for a given data scenario (such as Air Quality).

- Find a lead Local Authority for a group of data scenarios
- Find the associations and professional bodies that represent the discipline
- Find other councils that want to participate
- Find other stakeholders from the public sector
- Find suppliers of vertical applications of the data, who want to contribute capacity and expertise
- Find existing standards for the data scenarios
- Work with linked data expertise associated with data.gov.uk
- Model the data scenarios as a simple directed graphs
- Propose an ontology for the scenarios re-using existing linked data vocabularies
- Find sources of reference data that can be used as the basis of links within the data and therefore become URI Sets

- Pilot the approach
- Establish tools and services for converting traditional data formats, i.e. Helper App
- Propose the Potential Uses for the data in a linked data form
- Publish Ontology and Guidance

One of the outputs of this process is an ‘Ontology’ for the data scenario, which defines the concepts in the data and how they relate to each-other. This can be graphically represented as a directed graph.



In the example, the concepts shown in green illustrate where further information could be linked to, which would enhance meaning and comparison.

If the data were published in this form, then correlations may be found to otherwise unrelated data sets such as the demographics of the population of an area, leading to useful services that may relate health issues to air quality.

Potential Action Point

Work up the ‘Air Quality’ example with Bristol City Council and present to the Local Data Panel, to illustrate

- The steps necessary to re-purpose data into this form
- The innovation that is possible when the data is provided in this form.
- Correlations to otherwise unrelated data such as health.

Making links to common reference data

Making data linkable includes using common identifiers within concepts. Some of these will be specific to a data scenario (e.g. Pollutants for Air Quality), but others will recur across a wide range of types of data, such as:

- Location
- Addresses

Statistical Geographies
Localities

Organisations

- Public Sector Bodies
 - Local Authorities
 - Departments
 - Local Agencies
- Businesses
- Voluntary Sector

Services

- Service Types
- Actual Services

People

- Customers
- Segmentation
- Circumstances

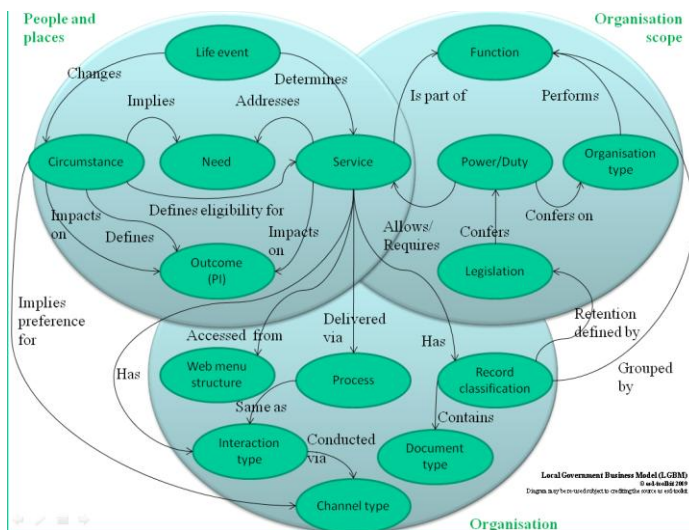
Some programmes, such as INSPIRE are already setting up sources of definitive reference data, and some of those are doing that in a 'linked data friendly' way.

Potential Action Point

Prioritise a manageable set of reference data types and work to influence the owners of those identifiers to make them available in a linked data friendly way, which includes

- License free
- Properly governed
- Web addressable
- Useful onward links

The esd-Toolkit has devised a 'Local Government Business Model' that defines a set of commonly recurring concepts and a set of terms for each.



Potential Action Point

Promote the use of the Local Government Business Model as a source of controlled vocabularies that can be used to make links in local public open data.

The ‘roles’ that enable data to be meaningful, comparable, linkable

Linked Data ‘Role’	
Many Data Publishers.	<p>Types of Data</p> <ul style="list-style-type: none"> • Transactional Data e.g. Spending, Pothole Incidents • Reference Data, e.g. Councils • Controlled Lists , e.g. Types of Councils, Types of Local Services <p>A Data Publisher can publish their data in a linked data form as static text files in a flavour of RDF, that can be ‘dereferenced’ over the web.</p> <p>May register data with one or more registries.</p>
Many Registries	<p>Enables a ‘Data Publisher’ to register that they have published data of a particular type, using an ontology, and with a set of metadata.</p> <p>Provides query and subscriber services so that data can be discovered, and gives alerts to new versions.</p>
Many Aggregation Points	<p>Operates a ‘Triplestore’</p> <ul style="list-style-type: none"> • Pulls in selected data sets • Provide Query and ‘api’ services • Has ‘Service Levels’ <p>Draws on Linked Data referenced by one or more Registries.</p> <p>Will be motivated by one or more of</p> <ul style="list-style-type: none"> • Commercial • Lobbying – single issue • Public Service • Hobbyist
Many Consumers	<p>Providing web based facilities such as</p> <ul style="list-style-type: none"> • Web Sites • Mobile Apps <p>Queries one or more Aggregation Points</p>
Many End-Users	<p>Uses services provided by a Linked Data Consumer</p>

Stakeholders who may take on those roles – who do we need to engage with?

Who is providing advice and guidance to the ‘local sector’?

- The LG Group led by the LGA
- Esd-Toolkit
- SOCITM
- LeGSB
- Transparency Board
- Local Data Panel
- CTO Council / Cabinet Office
- UK Location Council
- Knowledge Council

- European Commission
- Public Sector Bodies who lead on certain data themes
 - Office of National Statistics
 - Ordnance Survey

Potential Action Point

Propose a vision and ambition for local public data, and align those providing guidance to the local sector, to it. This may be a local slant on the public data principles produced by the Transparency Board and published at data.gov.uk (see <http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles>)

Who has the operational Data?

- Local Authorities
- Local Sector partners
- Government Departments
- Other agencies
- The public

Who has controlled vocabularies?

- Professional Associations (e.g. CIPFA)
- Esd-Toolkit
- Other Domain standards Boards

Who has reference data / contextual data?

- Public Sector Bodies who lead on certain data themes
 - Office of National Statistics
 - Ordnance Survey
 - Department for Transport
 - ... and so on
- The Linked Data Cloud e.g. Wikipedia

Who operates Registries (a place to find out where all of the data of a particular type is)?

- Data.gov.uk
- Local-direct.gov

Potential Action Point

Consider how to provide one or more facilities on the web where a single query can return the locations of all data of a prescribed type, from all data publishers. For local data, that would mean being able to find the locations of all the data each local authority based on a query of a set metadata profile, which might include:

- Type of data (e.g. Air Quality)
- Publisher (e.g. London Borough of Walford)

- Published date (e.g. 2011-04-01)
- Is about concept/identifier (e.g. school/1234)

Who may provide aggregated query services?

- Knowledge Hub
- Private Sector
- Lobbying Groups

Potential Action Point

Consider if the local sector should offer aggregated query services, and how they would be differentiated from those which might be created organically e.g.

- Service Levels
- Non-Commercial
- Useful data
- Reliable and Sustainable

Investing in Skills and Technologies

Each Local Authority may make their own investment in data skills and technologies. However, given that meaningful, comparable, linkable data requires coordination of formats, definitions and identifiers, a focus for local sector data services may be attractive.

Potential Action Point

Consider the local sector may be supported by data services that may include:

- Transformation of tabular data into Linked Data
- Registries
- Aggregation and querying services
- Support for consuming linked data