

Publishing Planning Application Data from across the authorities in Hampshire

Paul Davidson, The Local e-Government Standards Body (LeGSB)

April 2014

- The workshop will be two sessions,
- Standards and Schemas
 - To define a common target data format and vocabulary to be used across the contributors
 - harmonising data over the contributors, looking for standard definitions, lists, etc.
 - re-use, or create, standards that can scale to be used Nationally
 - stakeholders for data and standards
- Publishing and Visualisations
 - the means by which users (public, business, members, techies) interact and access it.
 - how you think a user may wish to make the most of the data
 - how to see it
 - what questions they may want to ask of it etc
 - Sort of like 'planningalerts', but because it's made by the authorities themselves it is much richer than a screenscape and can be linked to relevant datasets.

Introduction to LeGSB

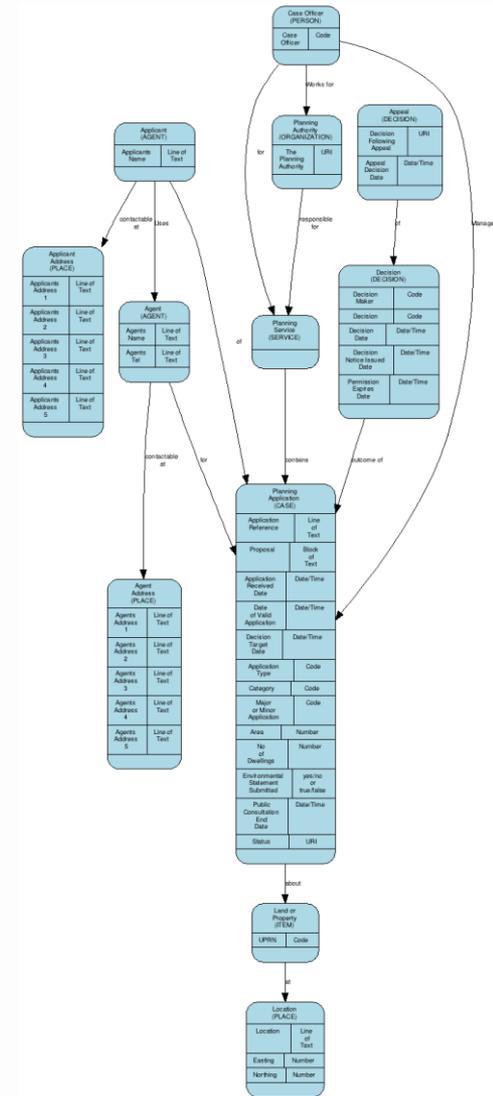
- Operating since 2006
- Funded by Central Government Departments – currently
 - Department for Communities and Local Government (*DCLG*)
 - Department for Work and Pensions (*DWP*)
 - Department for Business, Innovation and Skills (*BIS*).
- Mission
 - To promote Standards for Efficiency, Transformation, and Transparency of Local Services
- People
 - All personnel are employees of local authorities, and central government departments.
- Web Site
 - <http://www.legsb.gov.uk>

Harmonising Data

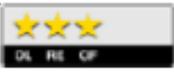
- What Data?
 - Scope
 - Fields, Concepts
 - Where is it defined?
 - How is it 'codified'?
 - How is it already used as data
 - Returns
- Who Cares?
 - Stakeholders
- How might it link to other data, now or in the future?
 - When do applications become real/completed outputs (particularly for new housing)? Challenge here includes hooking in to Council Tax registers?
 - What do statutory consultees say about individual applications? I'm especially interested in any capturing comments/objections from Environment Agency given the flooding app I published yesterday.

What Data?

- **Scope**
 - A scoping statement?
- **Fields**
 - Concept, Entity, Class
 - Datatype
 - Property
 - encoding
 - URI Set
- **Material**
 - Planning Advisory Service
 - Planning Data Collection forms
 - Data Fields identified by Hampshire
 - Redbridge
 - Leeds Data Mill



Syntax - Publishing up to step 3

Star Rating ⁵	In practice	Pros	Cons
 <p>DL</p>	<p>This is typically a static document, perhaps containing data as tables.</p> <p>Often this would be web page as html⁶, a document as pdf⁷, or an image.</p>	<ul style="list-style-type: none"> • easy to produce; • retains presentation and layout; 	<ul style="list-style-type: none"> • cannot further manipulate the data such as sorting, filtering, summing etc; • cannot join or compare to other data, or earlier versions;
 <p>DL RE</p>	<p>This is typically data, such as a spreadsheet, published in the format of the tool that was used to extract it, such as xls⁸.</p>	<ul style="list-style-type: none"> • no new tools or skills necessary; • data can be downloaded and further processed and analysed. 	<ul style="list-style-type: none"> • 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.
 <p>DL RE CP</p>	<p>This is typically 2* type data published using an open format, such as csv⁹ or xml¹⁰.</p>	<ul style="list-style-type: none"> • there is typically a choice of open source tools available for each open format 	<ul style="list-style-type: none"> • 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.

Syntax - Publishing up to step 5

	<p>This is explicitly about publishing data to the RDF¹¹ data model, and providing a query service using the SPARQL¹² language.</p>	<ul style="list-style-type: none"> enables others to make statements over the web about individual lines of data; can be queried over a data service so that a complete data set does not have to be downloaded; gives precise definitions to the meaning of the data. 	<ul style="list-style-type: none"> requires a skill set that most local authorities do not have yet; not suitable to provide directly to residents. Requires some infrastructure
	<p>This is Linked Data in the Rdf model, that contains links to external datasets to describe the 'things' that the data refers to.</p>	<ul style="list-style-type: none"> 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. 	<ul style="list-style-type: none"> requires a 'spine' of core reference data to make links to.

CSV for Download

- 'csv' format is effectively representing the data as a spreadsheet with rows and columns which would be made available for download at the council's web site.
- Advantages are:
 - Easy to achieve.
 - Easy to read by the public using widely available and open source software
- Some drawbacks are:
 - Downloading creates a copy of the spreadsheet, which naturally goes out-of-date when updates are applied to the definitive dataset held by the council.
 - A user would have to return periodically to the council's web site to see if a later version was available.
 - A 2-dimensional spreadsheet can struggle to represent complex data, particularly where structures repeat, link.
 - The meaning of the data in cells is provided by the column label, which is easily misinterpreted.

XML for download

- 'xml' provides a richer structure in which to represent data, and clarify the meaning of elements within the data.
- Advantages are:
 - Good for other computer systems to import from, regularly, as the rigid structure can be used to pick data items out to update another database.
 - Can cope with repeating structures
 - Data in XML can be further transformed into other formats, or other XML definitions.
- Drawback are:
 - Requires that an XML Schema (xsd) is developed that defines the structure of the XML data.
 - Changes to the XML schema need to be handled, and communicated under change control.
 - Cannot carry additional data that is not defined in the schema.
 - Not useable directly by the public.

xml as the payload responding to an api querying the data

- An 'api' (application programming interface) can be developed to provide a means of querying the definitive data, in situ, and responding with a result set in xml format.
- Advantages are:
 - Data is not copied, so the results are always up to date.
 - The call to the api can be made by another computer rather than by a person, and therefore is good to enable other portals to gather information from one or more sources and then re-present it.
- Drawbacks are:
 - Requires infrastructure to run a data service.
 - The api itself needs to be described so that potential users know how to use it.
 - Can only query the data using the search facilities provided by the 'api'.
 - Changing the api would 'break' portals that relied on it.

rdf for download

- 'rdf' is a way of representing data as a series of statements. The statements are then linked to each other, and potentially to other statements about the same things in other, otherwise unrelated, datasets.
- Advantages are:
 - There are some existing published structures for representing common components as rdf.
 - the statements can be defined using well known 'properties' from various sources.
 - Extra fields can be added without breaking existing implementations.
 - Links can be made to other data about the services described such as performance, ratings, the provider, what is known about the target user group.
 - Once downloaded, anyone can load the data into a 'triplestore' with no prior knowledge of the structure of the data.
- Drawbacks are:
 - Not useable directly by the public
 - Lack of skills to create and consume data in this form.

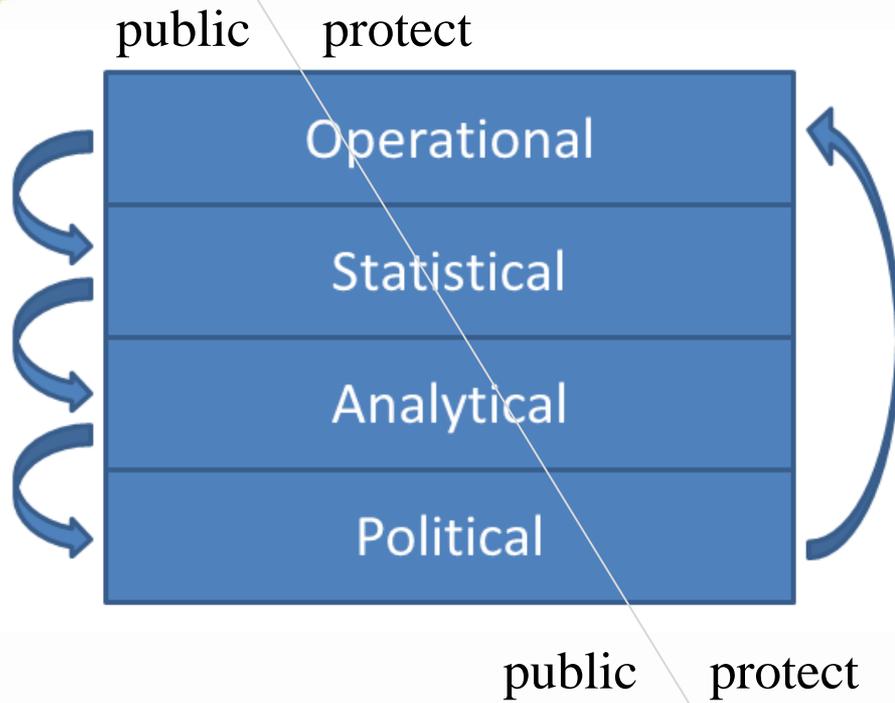
rdf as the means of storing the data to support an 'api'.

- When rdf data is loaded into a triplestore, it provides a generic 'api' (SPARQL) that can be used to bring back results from a query.
- Advantages are:
 - Can search against any term, and patterns in the data.
 - Can return results in many formats, including csv, xml, json.
 - There are open source triplestores, which can run on opensource architectures, minimising the cost.
- Drawbacks
 - The SPARQL query language can be hard to use. This can be mitigated by setting up popular queries in advance.

A Local Data Ecosystem?

- Where organisations can contribute data and also consume it, sharing at a case level, through to insight and joined up policy making.
- An ecosystem needs to have some common terms, so that data about the same thing can be linked and enriched. Those common terms can come from the Smart City Concept Model.

Context

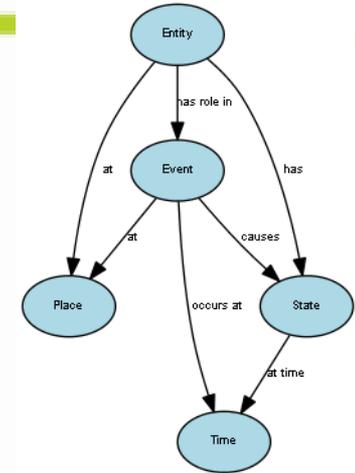
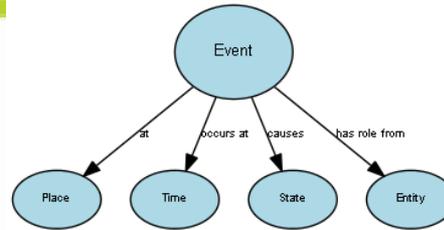


- OPERATIONAL - Data about real people and places, with real needs and circumstances, using real services, i.e. case work
- STATISTICAL - Aggregated operational data – organised using common classifications and segmentations
- ANALYTICAL - The conclusions drawn from an analysis of statistical data
- POLITICAL - The decisions taken to shape services, e.g. budgets, strategies, priorities, targets etc.

The Smart City Concept Model

Model

Name Smart City Concept Model
Version 0.0.2



Views

[Items](#)

[Events](#)

[Services](#)

[Cases](#)

[Objectives](#)

[All](#)

Prime Concepts

Account	A container of information, held by an AGENT, in which to record data arising from EVENTS that relate to an ENTITY in a specific ROLE.
Agreement	A negotiated arrangement between AGENTS as to a course of action
Building	A man-made OBJECT with roof and walls, with a fixed PLACE, intended for sheltering PERSONS or other OBJECTS.
Case	A container for information recording the history of EVENTS initiated by a SERVICE Demand.
Community	A group of PERSONS and/or ORGANISATIONS that share common characteristics such as location, circumstance etc.
Decision	A conclusion or resolution reached after consideration
Event	An occurrence that has or may happen over a period of TIME.
Metric	A measure of demography, characteristics, activity or performance
Object	A physical ENTITY
Objective	A set of STATES desired by an AGENT.
Organisation	A group of PERSONS with a collective goal.
Person	An individual human being, while alive.
Place	A geographic position, area, or volume.
Plan	A list of steps with TIME and RESOURCE, used to achieve an OBJECTIVE.
Resource	The capacity that can be derived from an ENTITY to produce a benefit.
Rule	an explicit or understood regulation or principle governing conduct or procedure within a particular area of activity
Service	The capacity to carry out a Procedure.
State	A circumstance or condition of an ENTITY at a TIME.
Time	TIME is a dimension in which EVENTS can be ordered from the past through the present into the future, and also the measure of durations of EVENTSs and the intervals between them.

Group Concepts

- [Agent](#) An ENTITY, but most often a PERSON, or ORGANISATION, providing a SERVICE or taking a ROLE in an EVENT
- [Entity](#) Something that has a distinct, separate existence.

Identifiers for 'things'

- Companies
 - All UK companies are registered at *Companies House*
 - Each is given a Registration Number
 - Builds into a URI Set
- For example
<http://business.data.gov.uk/id/company/02050399>
- *URI*
 - Uniform Resource Identifier
 - a component of the World Wide Web
 - provide a means of uniquely naming a 'Thing'
 - A web address for a fact
 - Can be 'looked up' over the web to give
 - *html*
 - *data*
- Guidance
 - Designing *URI* Sets for the UK Public Sector
 - <https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector>

Categorisations from LG Inform

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

Local Government Association | LG Inform | LG Inform Plus | API | Standards

Search | Help | About | Sign in | Register



esd standards
powering LG Inform
[link to current tab]

items

search

linked data

help

A to Z list

Glossary of services

Channels list

Means of interacting between...

Circumstances list

Characteristics of people and p...

Functions list

Functional hierarchy of local g...

Interactions list

Ways in which people interact t...

Issues and benefits list

Issues addressed and benefits...

Life events list

Happenings, planned or otherw...

Metrics group of lists

Measures of demography, char...

Needs list

Requirements that, if met, resul...

Powers and duties list

Legislation conferring the right...

Processes list

Sequences of activities perfor...

Services group of lists

Services provided by local gove...

Subjects list

IPSV, the Integrated Public Sec...

Website navigation group of lists

Hierarchy of local government...

Legsb
The Local e-Government Standards Body

paul.davidson@sedgemoor.gov.uk

www.legsb.gov.uk