Opening up public data to improve transparency and efficiency

Ghent eGovernment Conference, December 2010

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

Points of view ...

- Sedgemoor District Council
  - We have your data, how would you like it?
  - http://en.wikipedia.org/wiki/Sedgemoor
  - http://statistics.data.gov.uk/doc/local-authority/40UC

- Local e-Government Standards Body (LeGSB)
  - Mission to promote eStandards that support Efficiency, Transformation, and Transparency of Local Services
  - Semantics, Syntax, Data Quality, Rights, Authentication, Transport, Governance

- Cabinet-Office CTO Council
  - Cross Government Enterprise Architecture, and Domains
  - UK eGIF (now Standards and Architecture Framework), eGMS, PSIA
  - UK Govt. Infrastructure Strategy

- making public data open and accessible by establishing a well maintained repository of definitions and taxonomies that makes data consistent throughout Europe.

- Incorporate EU best practices into a common service delivery architecture (information, process and application-layer) that provides one common language and helps local governments share services

- Areas could include: public service lists, standard information on citizens, standard information on businesses, transport timetables, environmental information and geodata (GIS-data).

Public Data

- … meaning …

  – "Public Data" is the objective, factual, non-personal data on which public services run and are assessed, and on which policy decisions are based, or which is collected or generated in the course of public service delivery."

  > – Defined by the Transparency Board at data.gov.uk

LeGSB
Local e-Government Standards Body

- Public data policy and practice will be clearly driven by the public and businesses who want and use the data, including what data is released when and in what form – and in addition to the legal Right To Data itself this overriding principle should apply to the implementation of all the other principles.

- **Public data will be published in reusable, machine-readable form** – publication alone is only part of transparency – the data needs to be reusable, and to make it reusable it needs to be machine-readable. At the moment a lot of Government information is locked into PDFs or other unprocessable formats.

- **Public data will be released under the same open licence which enables free reuse, including commercial reuse** – all data should be under the same easy to understand licence. Data released under the Freedom of Information Act or the new Right to Data should be automatically released under that licence.

- Public data will be available and easy to find through a single easy to use online access point (data.gov.uk) – the public sector has a myriad of different websites, and search does not work well across them. It's important to have a well-known single point where people can find the data.

- **Public data will be published using open standards, and following relevant recommendations of the World Wide Web Consortium.** Open, standardised formats are essential. However to increase reusability and the ability to compare data it also means **openness and standardisation of the content as well as the format.**

- Public data underlying the Government's own websites will be published in reusable form for others to use – anything published on Government websites should be available as data for others to reuse. **Public bodies should not require people to come to their websites to obtain information.**

- Public data will be timely and fine grained – Data will be released as quickly as possible after its collection and in as fine a detail as is possible. Speed may mean that the first release may have inaccuracies; more accurate versions will be released when available.

- **Release data quickly, and then re-publish it in linked data form** – Linked data standards allow the most powerful and easiest re-use of data. However most existing internal public sector data is not in linked data form. Rather than delay any release of the data, our recommendation is to release it 'as is' as soon as possible, and then work to convert it to a better format.

- **Public data will be freely available to use in any lawful way – raw public data should be available without registration**, although for API-based services a developer key may be needed. Applications should be able to use the data in any lawful way without having to inform or obtain the permission of the public body concerned.
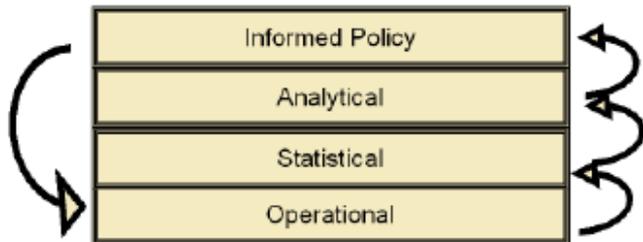
- **Public bodies should actively encourage the re-use of their public data** – in addition to publishing the data itself, public bodies should provide information and support to enable it to be reused easily and effectively. The Government should also encourage and assist those using public data to share knowledge and applications, and should work with business to help grow new, innovative uses of data and to generate economic benefit.

- Public bodies should maintain and publish inventories of their data holdings – accurate and up-to-date records of data collected and held, including their format, accuracy and availability.

What data might we publish?

LeGSB
Local e-Government Standards Body

- **About the Council**
  - Members of the Council
  - Members Expenses
  - Committees, Dates
  - Roles, Contacts, Hours, etc
- **Our Finances**
  - Payments to Suppliers over £500
  - Contracts
  - Council Tax Annual Charge for each Band in each Parish
  - Senior Officer Salaries
  - The Budget and the Chart of Accounts (Last year actual, This year budget, This year actual)
- **Our Services**
  - A-Z of Services
  - When is the rubbish collected for my property?
  - Web Site Visitor Statistics
- **Hot Topics**
  - Impact Assessment Data
- **Our Plans**
  - Policies and Strategies
  - Projects
- **Our Performance**
  - How quickly and accurately does the council handle applications for Housing Benefit?
- **About our Area**
  - River Water Quality
  - River Biological Monitoring Data
  - Air Quality Data

- **What are we dealing with?**
  - Abandoned Vehicles
  - Potholes
- **The people and communities in our area**
  - Population Estimates by Ward and LSOA
  - Quality of Life Survey
  - Mosaic Public Sector Data
  - Indices of Multiple Deprivation
- **The businesses in our area**
  - Food Safety Inspections
- **Where things are**
  - Electric Vehicle Charging Points (Travel)
  - Venues
  - Location of Car Parks
  - Location of Libraries and Leisure Centres
  - Location of Recycling Points
  - Schools
  - Allotments
  - Parks and Open Spaces
  - Polling Stations
  - Public Toilets
  - Archaeological Sites
  - ATMs (Free and Charging)
  - Flood Risk Zones
  - Pharmacies
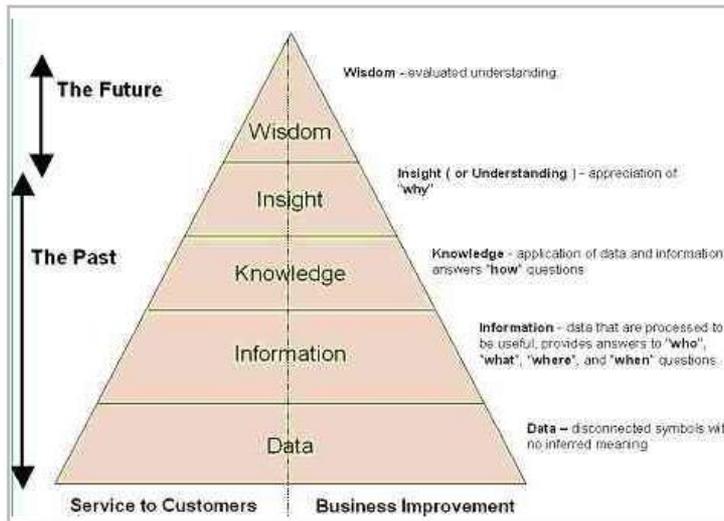 - Listed Buildings
  - Gritting Routes

LeGSB
Local e-Government Standards Body

Informed Policy
Analytical
Statistical
Operational

- **Operational** – real people, with real circumstances, needing real services, i.e. Case Work
- **Statistical** – using common terms for 'segmentation', 'regions' and so on, i.e. Aggregated Data
- **Analytical** – patterns, predictions, inferences, opinions, and so on. i.e. the conclusions drawn from the analysis of evidence and statistics.
- **Informed Policy** – Supporting evidence-based policy in terms that can shape operational delivery

The Information Architecture will provide a semantic and data quality framework that promotes information sharing in each of these contexts.

The four contexts above map well to the well-known 'WIKID' view of deriving value from data.

The Future

The Past

Wisdom - evaluated understanding.

Wisdom

Insight

Insight (or Understanding) – appreciation of 'why'

Knowledge

Knowledge - application of data and information; answers 'how' questions

Information

Information - data that are processed to be useful, provides answers to 'who', 'what', 'where', and 'when' questions.

Data

Data – disconnected symbols with no inferred meaning

Service to Customers | Business Improvement

How might Public Data be used?

- **Scrutiny**
    - Finding exceptions and correlations in the data

- **Non-Public Sector Web Sites and Mobile Apps**
    - The Asboromoter
    - http://openlylocal.com/
    - http://armchairauditor.co.uk/
    - http://www.wheredoesmymoneygo.org/

- **Informed debate on local issues**

- **Improved access to Services**
    - Joined up information on Services
    - Self-Service
    - 'Audience themed' Web Sites

- **Reduce Duplication**
    - Who is the Chief Executive of ....

... for scrutiny

575	CLG	HARDWARE	FUJITSU SERVICES	354,580.03
576	CLG	HARDWARE	PORTAPROMPT LTD	1,414.50
577	CLG	HARDWARE	SPECIALIST COMPUTER CENTRES PLC	187,846.96
578	CLG	HARDWARE	STERIA LIMITED	1,375,506.72
579	CLG	HARDWARE	SUNGARD AVAILABILITY SERVS (UK) LTD	7,434.75
580	CLG	HARDWARE	VERIZON UK LTD	106,975.12
581	CLG	HOTELS	ALDERMAN APARTMENTS LTD	2,899.89
582	CLG	HOTELS	CLEVELAND FIRE AUTHORITY	10,611.64
583	CLG	HOTELS	COMPANY BARCLAYCARD	4,868.72
584	CLG	HOTELS	DERBYSHIRE FIRE & RESCUE SERVICE	773.44
585	CLG	HOTELS	DONNINGTON MANOR HOTEL	3,241.68
586	CLG	HOTELS	DORSET FIRE AUTHORITY	8,105.22
587	CLG	HOTELS	EAST MIDLANDS REGIONAL FIRE CONTROL	1,568.76
588	CLG	HOTELS	FIRE SERVICE COLLEGE	518.00
589	CLG	HOTELS	HOLIDAY INN	7,537.88
590	CLG	HOTELS	HOTEL METROPOLE	570.00
591	CLG	HOTELS	INTERFAITH YOUTH CORE	2,268.62
592	CLG	HOTELS	LINCOLNSHIRE CC FIRE & RESCUE SERV	1,485.34
593	CLG	HOTELS	NOTTS & CITY NOTTINGHAM FIRE AUTH	710.10
594	CLG	HOTELS	OCEAN MEDIA GROUP LTD	511.13
595	CLG	HOTELS	RAMADA HOTELS	3,345.26
596	CLG	HOTELS	REDFERN	314,292.16
597	CLG	HOTELS	RIGHTNOW TECHNOLOGIES (UK) LTD	628.49
598	CLG	HOTELS	RUBENS AT THE PALACE HOTEL	932.31
599	CLG	HOTELS	SHEFFIELD CITY COUNCIL	1,322.70
600	CLG	HOTELS	SUFFOLK COUNTY COUNCIL	867.74
601	CLG	HOTELS	THISTLE HOTELS	11,221.66
602	CLG	HR CONSULTANCY	ALPHA CONSOLIDATED TRAINING LTD	1,656.00
603	CLG	HR CONSULTANCY	BROOKLEIGH CONSULTING LTD	3,457.63
604	CLG	HR CONSULTANCY	CONNAUGHT COMPLIANCE SERVICES LTD	19,843.71
605	CLG	HR CONSULTANCY	DENTON WILDE SAPTE LLP	1,486.31
606	CLG	HR CONSULTANCY	DEPT FOR TRANSPORT	8,280.00
607	CLG	HR CONSULTANCY	E L B A INTERNATIONAL	1,112.63

... Where does my money go ...
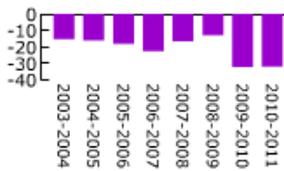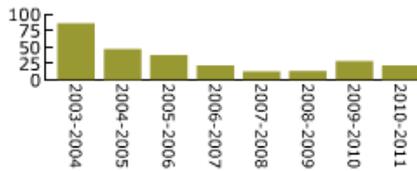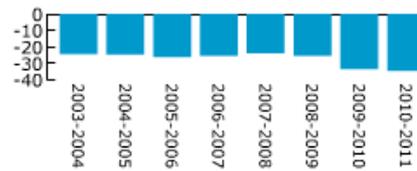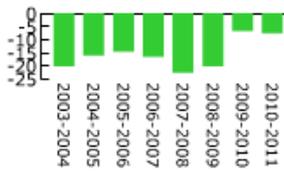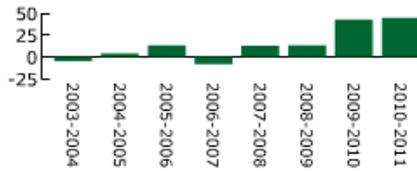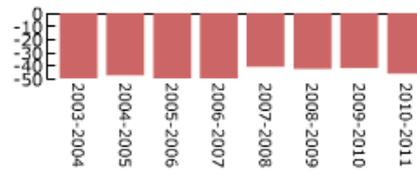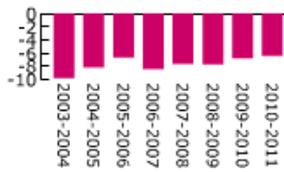
Rights - An Open Licence?
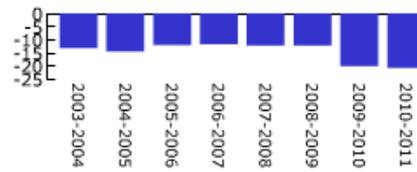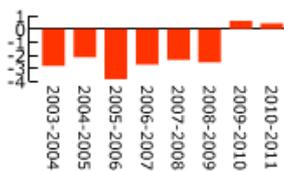
- http://www.nationalarchives.gov.uk/doc/open-government-licence

Open Government Licence
for public sector information

delivered by
The National Archives

Back to The National Archives

You are encouraged to use and re-use the Information that is available under this licence, the Open Government Licence, freely and flexibly, with only a few conditions.

Using information under this licence

Use of copyright and database right material expressly made available under this licence (the 'Information') indicates your acceptance of the terms and conditions below.

The Licensor grants you a worldwide, royalty-free, perpetual, non-exclusive licence to use the Information subject to the conditions below.

You are free to:

✓ copy, publish, distribute and transmit the Information;

✓ adapt the Information;

✓ exploit the Information commercially for example, by combining it with other Information, or by including it in your own product or application.

You must, where you do any of the above:

⚠ acknowledge the source of the Information by including any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence;

Syntax?

- **For Humans**
  - Web pages (e.g. HTML)
  - Documents (e.g. PDF, RTF)

- **For Download and Manipulation**
  - Tabular (e.g. csv)
  - Geographical (e.g. KML)

- **For Machines**
  - For Applications (e.g. XML)
  - Linked Data (e.g. RDF, OWL, URIs ...)
  - Over the Web (e.g. Web Services, APIs)

A Journey
from Open to Linked Data

Increasing skills, capacity and capability

LINKED DATA

LINKABLE DATA

GML

OPEN STANDARDS DATA

CSV

RTF

MACHINE READABLE DATA

PDF

OPEN DATA

Adapted by LG Group from diagram produced by Ian Painter, Snowflake

LeGSB
Local e-Government Standards Body
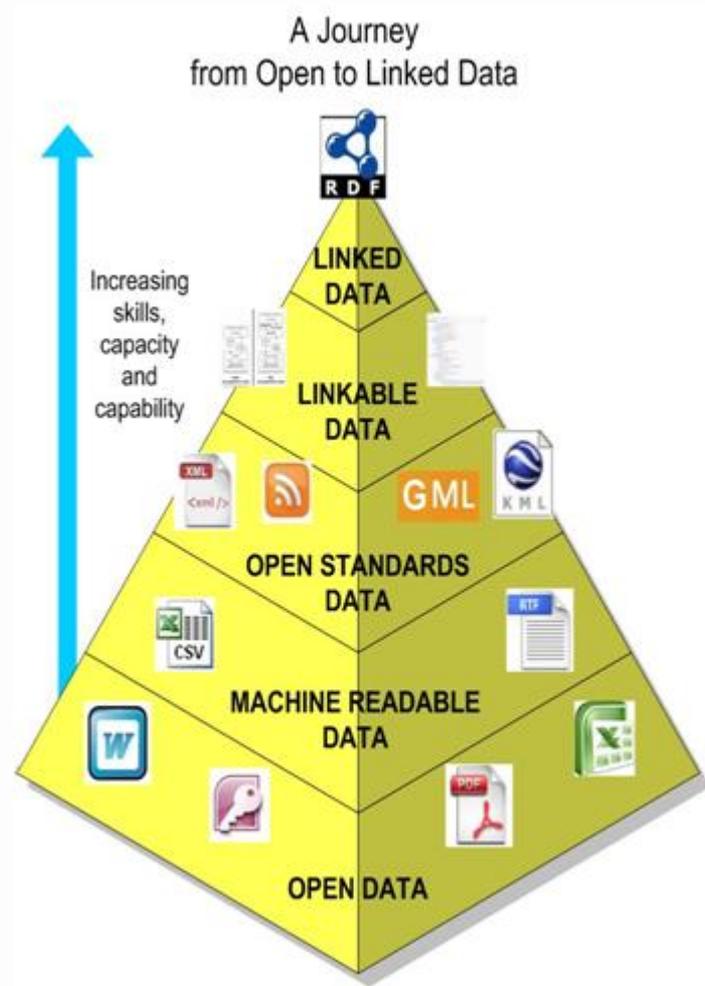
Defining formats for the local sector

- Guidance on how to publish data sets from the Local Government Association

Appendix A: Spending data template

Column	Field Name	What Is Required	Reason For Inclusion	Additional Information	Inclusion status
1.	Body name	Name of the organisation making the payment	Aids readability for casual reading		Minimum
2.	Body	A unique code to identify an organisation.	To allow the file to be self-describing	The code can be the ONS SNAC code or a Unique resource set (URI). Preferred is a Unique resource Identifier (URI) used in Linked Data.	Desirable
Description of Services and purchase details					
3.	Service Area Categorisation	The service area responsible for spending the money.	To identify the service that has spent the money. This is useful to compare payments between different authorities	Represents the highest level to which costs are charged. We propose to use the BVACOP objective category for service areas. For example:Adult Social Care. BVACOP headings currently have no 'codes' so this would have to be the full text of the heading. The BVACOP categories are the most desirable type to use and is a mandatory field in expenditure returns to central government	Minimum
4.	Service Division Categorisation	Represents the Divisions of Service and enables the sub-aggregation of costs.	To identify the specific division within the service that has spent the money	We propose to use the BVACOP objective code for Division of Service e.g. Older People inc Older Mentally ill.	Desirable
5.	Responsible Unit	Department responsible for spending the money.	To improve accountability identify the authority department that spent the money to	Some authorities may want to name the department in addition of the service categorisation	Optional
6.	Expenses Type	General headings that describe the nature of expenditure (e.g. Travel Costs)	To identify the general nature of the spend	The description of expenditure used against account codes held on departments' own finance systems. The two candidate classification used are BVACOP Subjective Group codes for Goods/	Minimum

http://lgtransparency.readandcomment.com/

Making data meaningful?

- Spending on maintaining the roads has more meaning when you know the length of roads in the council area, and perhaps the number of road surface related traffic accidents in the area.

- Spending on Schools has more meaning when you know the number of school age children and the exam pass rates.

- You may want to find those councils that spend a lot more or less than their neighbours on a particular type of product

- You might find a pattern between the spending on a certain type of product and the political control of a council.

- For each data scenario ...
  - Need light-weight ontology and URI sets
  - Capacity and expertise in the local sector
  - Co-ordinating and guidance
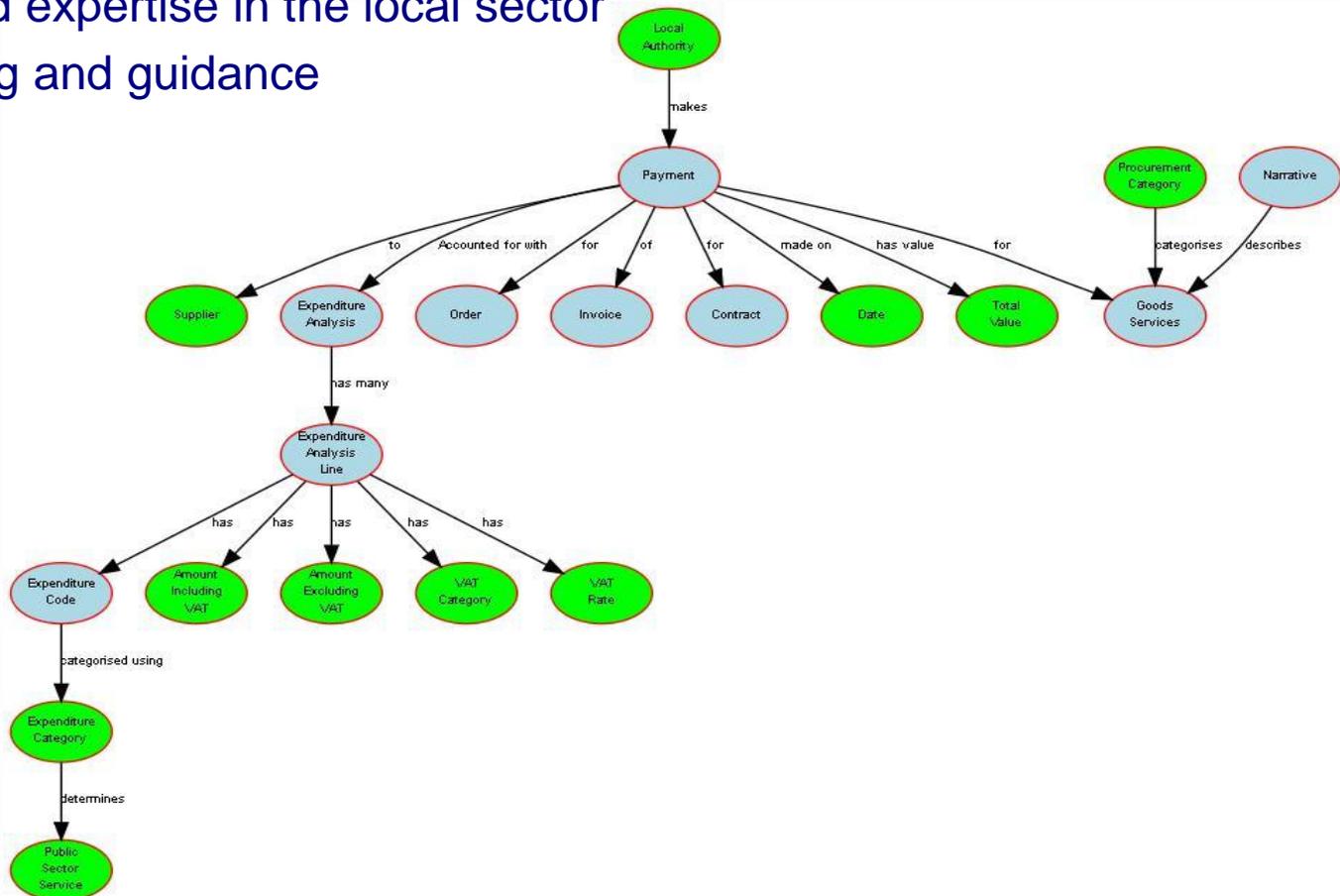
Aggregating and Querying

http://linked4.org/lsd/explore.html

The Architecture of Linked Data

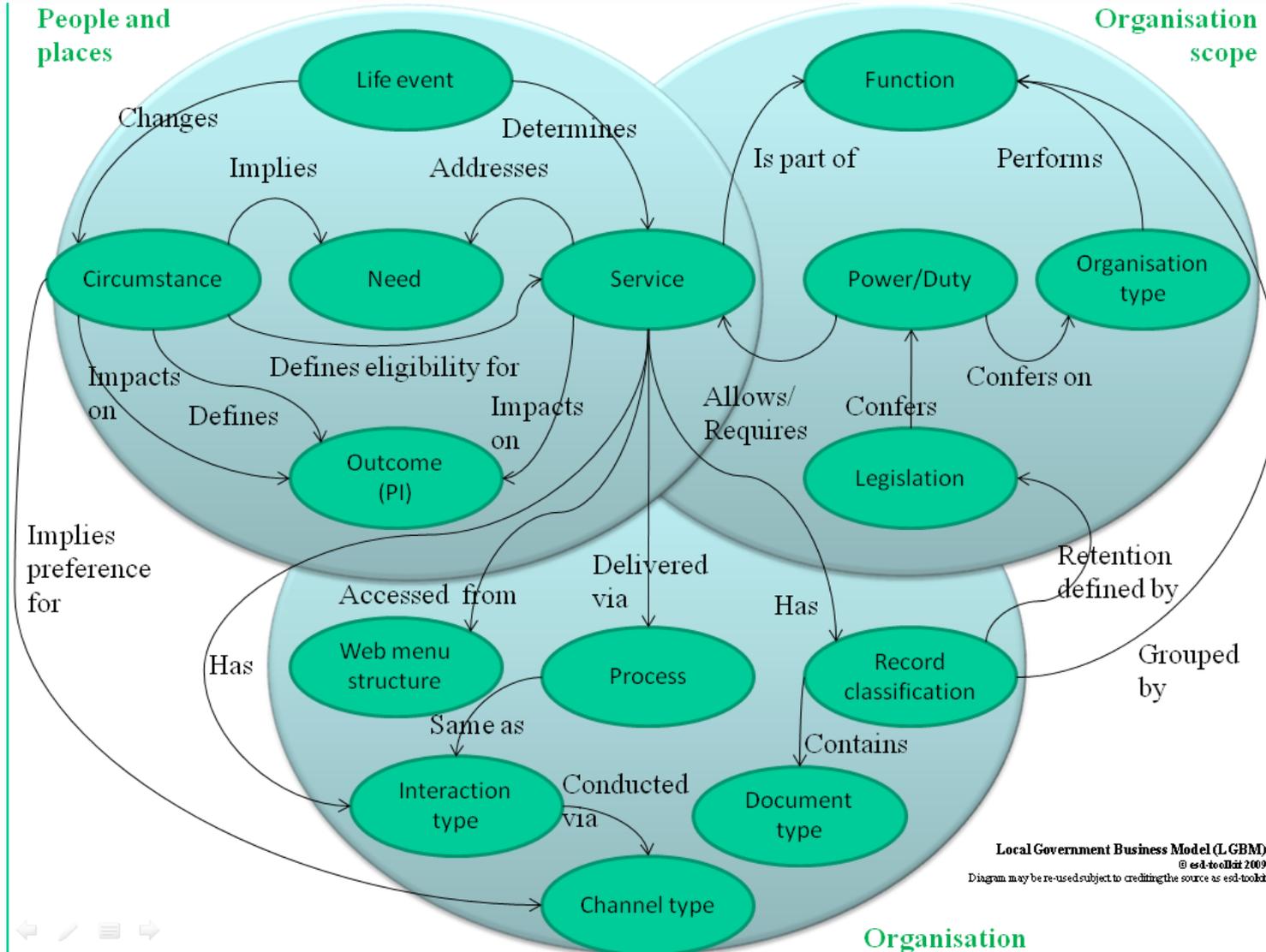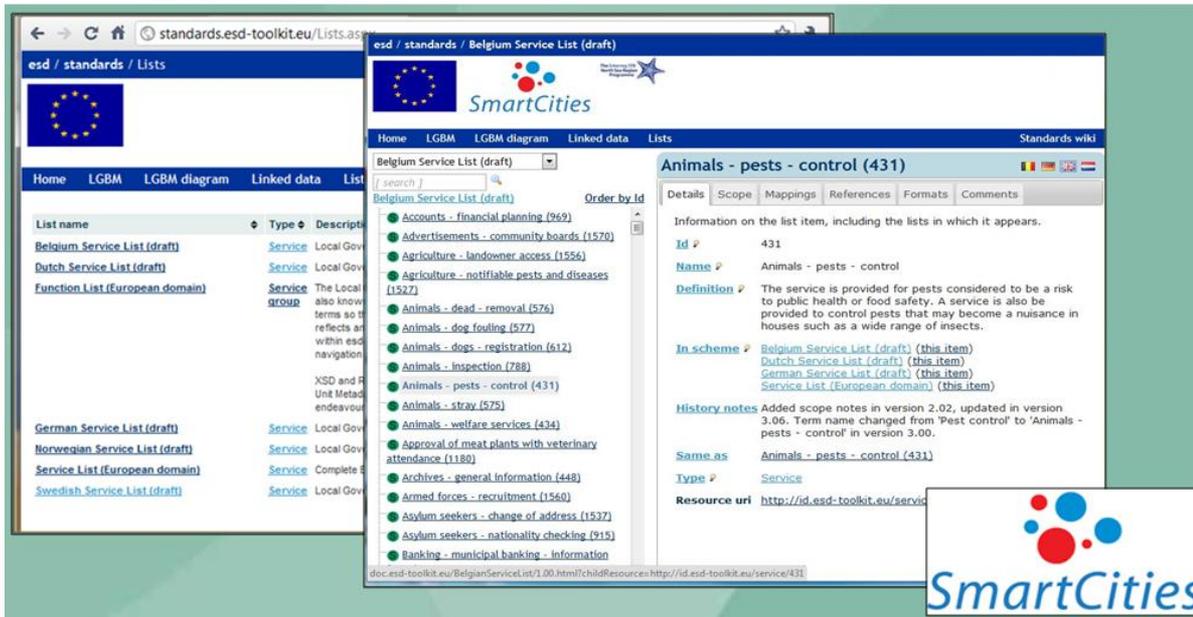| Linked Data 'Role'	
Many Data Publishers.	Types of Data • Transactional Data e.g. Spending, Pothole Incidents • Reference Data, e.g. Councils • Controlled Lists , e.g. Types of Councils, Types of Local Services A Data Publisher can publish their data in a linked data form as static text files in a flavour of RDF, that can be 'dereferenced' over the web. May register data with one or more registries.
Many Registries	Enables a 'Data Publisher' to register that they have published data of a particular type, using an ontology, and with a set of metadata. Provides query and subscriber services so that data can be discovered, and gives alerts to new versions.
Many Aggregation Points	Operates a 'Triplestore' • Pulls in selected data sets • Provide Query and 'api' services • Has 'Service Levels' Draws on Linked Data referenced by one or more Registries. Will be motivated by one or more of • Commercial • Lobbying – singe issue • Public Service • Hobbvist
Many Consumers	Providing web based facilities such as • Web Sites • Mobile Apps Queries one or more Aggregation Points
Many End-Users	Uses services provided by a Linked Data Consumer

Local Government Business Model

Local Government Business Model (LGBM)
© esd-toolkit 2009
Diagram may be re-used subject to crediting the source as esd-toolkit

Definitions and Taxonomies

- URIs and Ontology
- INSPIRE

- **Semantics**
  - the meaning of information
  - People, Organisation, Location, Payment, etc .....
- **Syntax**
  - the format of information
- **Quality**
  - the confidence to re-use information
  - Provenance, Verification, Accuracy, etc
- **Rights**
  - the right to use information for a defined purpose
  - Licenses, Legal Gateways
- **Authentication**
  - who is using information
  - Trust, Identity
- **Transport**
  - how to move information
  - Integrity, Availability
- **Governance and Assurance**
  - the behaviour and culture to protect and exploit information
  - Data Handling, Data Protection, Assurance, Information Risk

Legsb
The Local e-Government Standards Body

Info@legsb.gov.uk

www.legsb.gov.uk