	Big Data from Small Data Revision 2, final, 10th October 2012

Paul Davidson, LeGSB Director of Standards, and CIO Sedgemoor District Council, and Steve Peters, Department for Communities and Local Government

1   Summary

This note outlines proposals for an ambitious programme to unlock data about how public services are organised, delivered and their providers held to account at the local level. We estimate that the programme would run for 3 years, at an estimate cost of £2m.

The core of our proposals are:

- To develop standards and patterns which enable local public service providers – in particular local authorities – to routinely and systematically publish information on their priorities, activities and outcomes in a fully open, re-usable and accessible form,

- To demonstrate how these standards and patterns will enable this disparate local knowledge to be blended and combined in various ways, to provide a consistent national picture.

We believe that this will stimulate significant new national on-line information resources, and therefore enable growth in the knowledge economy. More widely, our proposals will also provide the evidence base to help build a strong, responsive and competitive economy.

For example, we believe we can enable the creation of national datasets which support the Government's ambitions in its National Planning Policy Framework:

> **"ensuring that sufficient land of the right type is available in the right places and at the right time to support growth and innovation; and by identifying and coordinating development requirements, including the provision of infrastructure"**

Our proposals are founded on a new model for local public service providers to describe, inter-link and publish open, re-usable data on their priorities, roles, outcomes etc. This model has been tested in various local and national scenarios.

We are now ready to work with the Data Strategy Board and other key partners to begin implementing the model on the national scale. Crucially, success will depend on various key enablers, outlined below. This includes developing a spine of maintained and trusted core reference data: so that publishers can confidently publish using consistent, standardised terms; and so that users can confidently re-use, aggregate and create innovative new national resources, tools and insights.

Geographic Information is a key building block.   We believe our model can and will enable development of a rich, new set of core reference data: ranging from address/property level data, through to information describing and defining administrative and other geographies.

2   Small Data that could be Big

Local public service providers, in particular local authorities, hold a wealth of unique, definitive data about their priorities, plans, and progress in their particular localities. Examples include:
- Planning Applications
- Location and facilities of Recycling Centres
- Food Premises Inspections
- Reports of Potholes
- Incidents of Abandoned Vehicles

Many of these datasets are also valuable in the national context: "The whole is more than the sum of its parts".   However, much of this data is locked away inside 100s of local public sector organisations.   It is therefore extremely difficult and costly for users to bring these disparate sources together, and create a consistent, national picture.

We believe that these national resources would provide significant benefits to the knowledge economy, and also the government's wider drive for economic growth and deficit reduction.   It should be possible for users to create, analyse and combine these national information resources: and use them alongside established national datasets on Health, Employment, and Crime.   In short, this is about enabling big data, from small data.

3   What could be done with small data when it is Big?

Taking the example of 'Planning Applications', illustrates how this data is useful over the four information contexts in LEGSB's model.
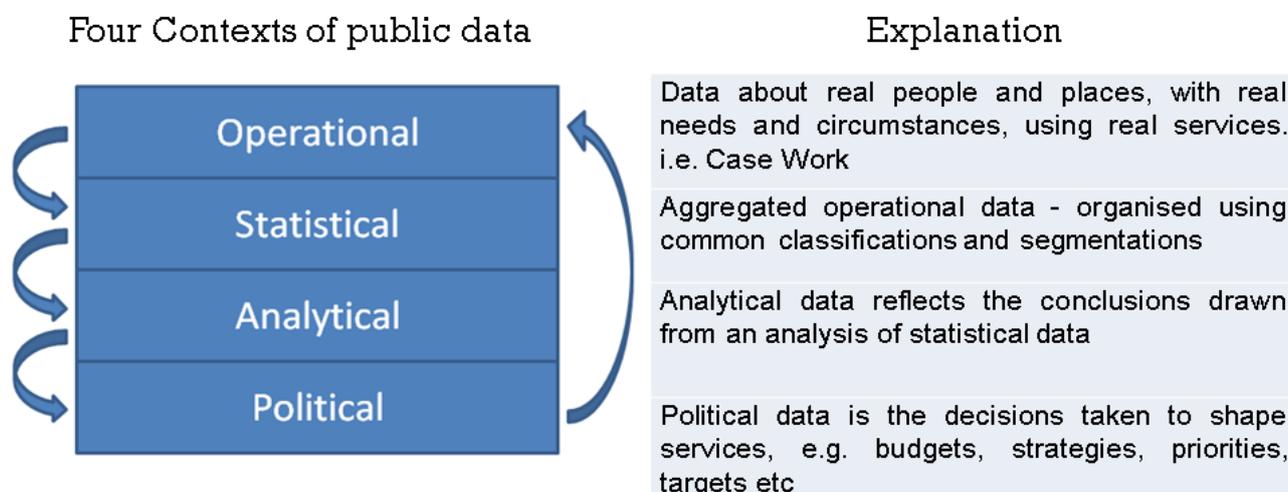

Four Contexts of public data | Explanation

	Explanation
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	Analytical data reflects the conclusions drawn from an analysis of statistical data
Political	Political data is the decisions taken to shape services, e.g. budgets, strategies, priorities, targets etc

Figure 1. Contexts of public data

So

Context	Insight / Service	Data
Operational	What planning applications are we dealing with right now?	Information about each current application. Ability for the public to view and comment on them.
Statistical	How many applications have we approved and rejected in the past month?	The number of applications decided over time, and the cost of the service.

	What is the cost of providing the service?	
Analytical	Are we meeting our housing targets? Are our planning decisions enabling growth in the local and wider economy?	Types of applications decided. Time taken to complete proposed development Numbers of jobs created locally.
Political	What are you doing to support sustainable development?.	Targets, Interventions, Objectives, Contracts

Similarly for 'Food Premises Inspections'.

Context	Insight / Service	Data
Operational	What is the rating of a particular food premises,	Inspection report and rating of an identified premises
Statistical	How are ratings and failures spread across premises types?	Types, Locations, Outcomes of Food Premises inspections.
Analytical	How safe are our food outlets?	Correlation to health incidents.
Political	What are you doing to make our food outlets safer?	Targeted training programmes. Risk Based inspections.

4 Building useful information services from small data when it is Big?

Taking the example of Planning Applications, it would be possible for various commercial intermediaries to aggregate and provide value added information services. For instance, we can imagine a more extensive, competitive and innovative information market place to support property development and retail planning sectors.

Taking the example of 'Food Premises Inspections' data, it would become practical for a commercial web site or app to incorporate this data into a 'Where to Eat?' service.

In both cases, these types of service are likely to be most successful when they have national coverage, rather than a series of disconnected sites or apps targeted at particular towns and other localities.

5 How to make Big Data from Small Data

This approach requires the use, and re-use, of some key enablers.

Enabler	
Core Reference Data	**To enable data to be joined up, we need to be able to link directly to definitive lists of 'things'. In particular** • **Location – coordinates, addresses, geographies, etc** • **Organisations – companies, charities, public sector bodies, etc**

	• **Functions – services, facilities** • **Metrics – statistics, demographics, etc**
Lightweight Patterns for publishing data	**Data formats and standards are crucial to be able to combine otherwise disparate data, however, many of these sets follow a similar pattern, e.g.** • **Location, Category, Outcome of a report or service request** • **Location of an Incident or Facility** • **Count or Measurement of a 'thing' at a Location.** **So, rather than create detailed data formats for each potential data set, we should look for more generic patterns that can be used to publish a wide range of otherwise unrelated data. More detailed standards can be added later where there is value.** **Re-useable standards can be applied to communicate general ideas such as licence, provenance, quality, etc.**
Federated Publishing points	**In preference to a single national public sector warehouse for all of this data, local organisations should be able to simply publish their data, to the agree standard, and update a registry to declare that it is now available.** **Opportunities for repositories where this data can then be queried as 'Big Data' can then arise naturally, to meet a demand.**

6   Enriching Small Data by linking to Big Data

As a consequence of creating Big Data from Small Data, the original publisher should find that they can now do more with their own data. For example, by linking coordinates to administrative geographies, the planning applications example could be queried for new applications, or trends in a 'District Ward' – which can then be pushed to the local Ward Councillor.

Appendix A – LeGSB Concept Model for Open Public Services

The LeGSB concept model shows how lightweight definitions of key concepts can be linked together over many publishers to provide a community or locality based view of local services and local objectives.
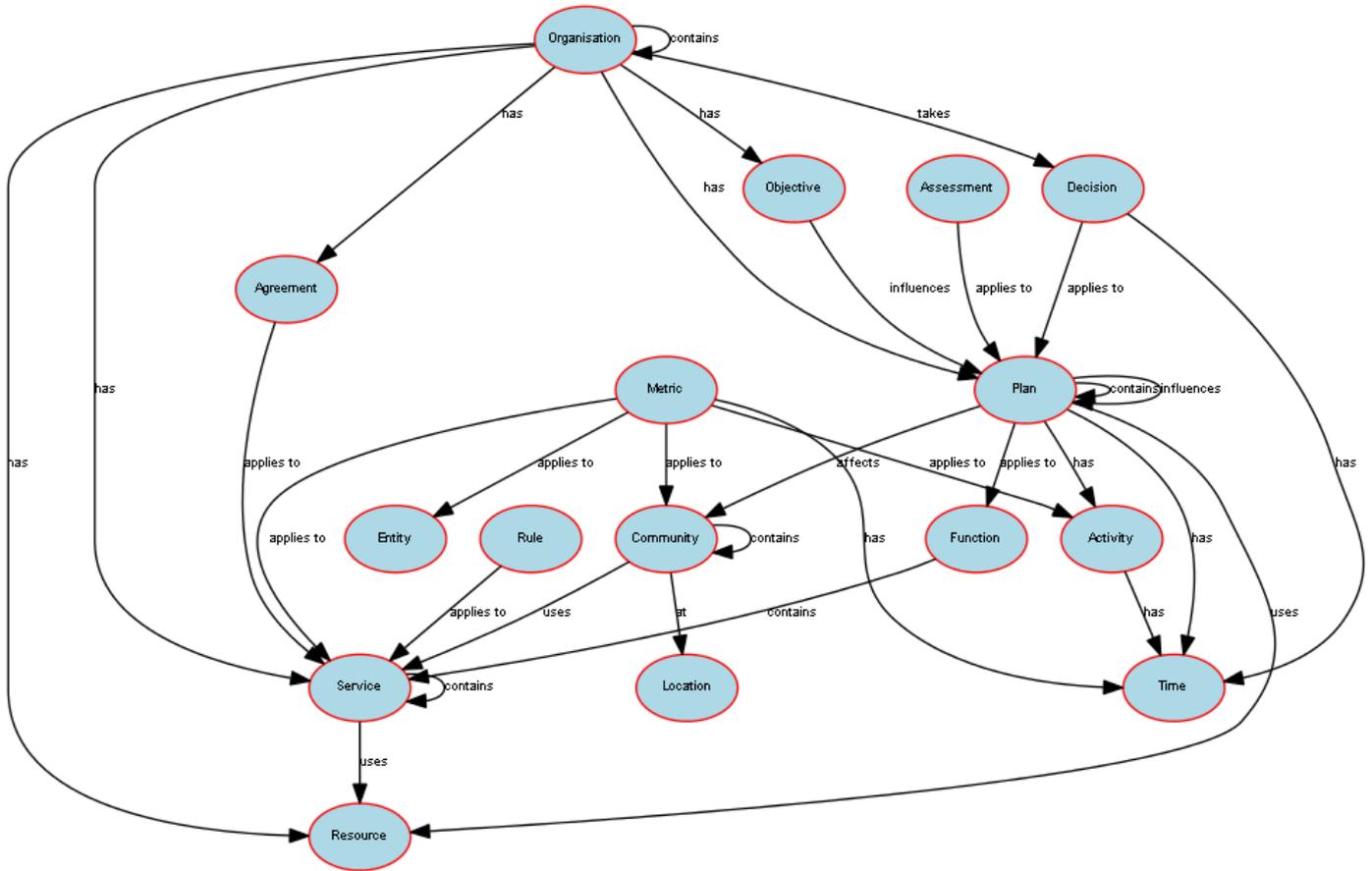

Figure 2. LeGSB concept model linking data.