

<b>Title</b>	Blueprint for a Local Data Ecosystem
<b>Revision</b>	3
<b>Status</b>	Draft
<b>Date</b>	19 <sup>th</sup> October 2014
<b>Author</b>	Paul Davidson, LeGSB Director of Standards, and CIO Sedgemoor District Council
<b>Purpose</b>	To scope the products for the iNetwork 'Local Data Ecosystem' project.
<b>Type</b>	Blueprint

Revision	Date	Author	Notes
1	18/02/2014	Paul Davidson, for LeGSB	
2	12/03/2014	Paul Davidson	Remove table for Troubled Families audience.
3	19/10/2014	Paul Davidson	Revise scope diagram. Change 'Authentication' to 'Trust'

## 1 Background

On the 10<sup>th</sup> February 2014, the Open Data User Group awarded funding to Tameside Council, to develop a 'Local Data Ecosystem', illustrated using a 'Troubled Families' theme.

This document provide a blueprint for the 'Local Data Ecosystem'.

## 2 Purpose of the Local Data Ecosystem

The purpose of the 'local data ecosystem' is to provide a means where data can be contributed, linked up, and consumed across many agencies and the public, and replaces the more traditional approaches where

- Open Data is published as spreadsheets across many web sites
- Protected data is transmitted between individual agencies.

## 3 Components of the Local Data Ecosystem

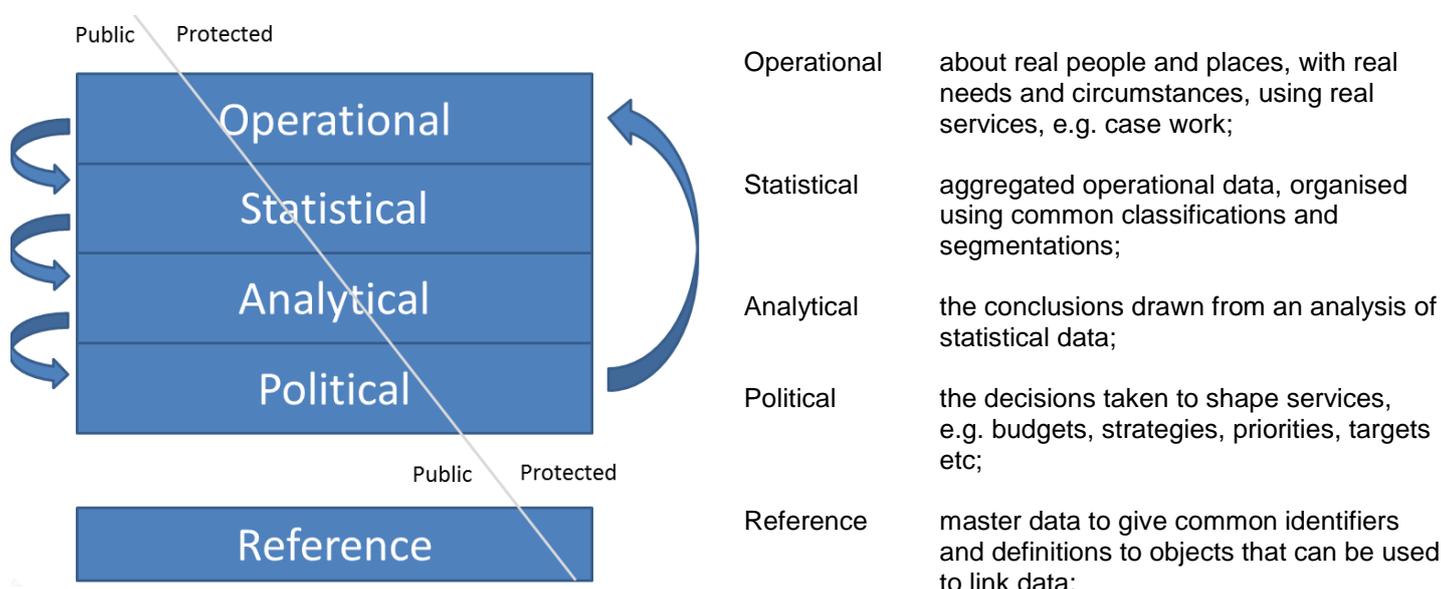
The 'local data ecosystem' will contain

- role based access to a multi-agency secure case data sharing system
- community access to a 5\* open data publishing platform
- personal data access via personal data store
- dashboards for various stakeholders

Data will flow across the components via common standards and frameworks.

## 4 Scope of data in the Local Data Ecosystem

The scope of data in the local data ecosystem is as described in the LeGSB white paper, Strategic Approaches to Data Sharing<sup>1</sup>. Data in the ecosystem will be considered over five contexts:



The use of common standards will provide a ‘golden thread’ where data at each level is supported and evidenced by the layer below. Ultimately, the impact of political data, should be observable back in future operational data.

In each context, data will either be ‘Public’ or ‘Protected’.

Public Data	“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.” <a href="http://data.gov.uk/opendataconsultation/annex-2">http://data.gov.uk/opendataconsultation/annex-2</a>
Protected Data	Data ... <ul style="list-style-type: none"> <li>containing personal information which is covered by the Data Protection Act, or</li> <li>for which there is a relevant exemption from legislation such as the Freedom of Information Act.</li> </ul> or <ul style="list-style-type: none"> <li>containing IPR, which itself maybe from a 3<sup>rd</sup> party, which requires special licence conditions or fees as described by the regulations for the re-use of public sector information.</li> </ul>

## 5 Illustrating the local data ecosystem with ‘Troubled Families’.

The objectives of a Troubled Families program are likely to include

- Improving the quality of life of members of selected families;

<sup>1</sup> <http://legsb.i-network.org.uk/wp-content/uploads/2013/08/LeGSB-aStrategicApproachToDataSharing-Rev1.pdf>

- Reducing the impact of certain families on the surrounding neighbourhood;
- Reducing the demand on public services and hence reducing costs.

The data associated with a Troubled Families programme can be considered over the four contexts, for example

Operational	<ul style="list-style-type: none"> <li>• the characteristics of each selected family;</li> <li>• services used by each member of each selected family;</li> <li>• interventions applied to each case;</li> </ul>
Statistical	<ul style="list-style-type: none"> <li>• caseload by 'risk factor';</li> <li>• cost of providing reactive services;</li> <li>• cost of providing the 'troubled families' programme.</li> <li>• counts of selected types of incidents by area.</li> </ul>
Analytical	<ul style="list-style-type: none"> <li>• assumptions about the future costs of services;</li> <li>• effectiveness of types of interaction;</li> <li>• success rate for 'turning around' families;</li> <li>• value for money of the programme;</li> </ul>
Political	<ul style="list-style-type: none"> <li>• objectives and plans;</li> <li>• targets for future demand on reactive services;</li> <li>• targets for the number of 'troubled families'</li> <li>• decisions to allocate resources to the programme;</li> </ul>

Working with councils in Greater Manchester and the wider North West of England, LeGSB has already undertaken initial research and analysis work to explore the data ecology relating to Troubled Families. This has identified a number of gaps and inconsistencies with regards data standards that make aggregation of data difficult and time consuming. This area is a key theme that cuts across welfare, adult and childrens' social care, police, ambulance, probation and other services and which therefore extend the benefits of this work to a range of other communities and services.

## 6 The Standards Framework for the local data ecosystem.

The framework for standards for the local data ecosystem is as described in the LeGSB white paper, Strategic Approaches to Data Sharing<sup>2</sup>. Standards will be considered over seven themes:

- **Semantics** the meaning of information
- **Syntax** the format of information
- **Data Quality** the confidence to re-use information
- **Rights** the right to use information
- **Trust** who is accessing information
- **Transport** how to move information
- **Information Governance** the behaviour and culture to protect and exploit information

## 7 Semantics for the local data ecosystem.

<sup>2</sup> <http://legsb.i-network.org.uk/wp-content/uploads/2013/08/LeGSB-aStrategicApproachToDataSharing-Rev1.pdf>

## 7.1 Concepts

The 'local data ecosystem' will use the 'Smart City Concept Model' being developed by LeGSB for the British Standards Institute (BSI) and the Department of Business Innovation and Skills (BIS). The concept model will be published as a BSI PAS182.

The Smart City Concept Model provides sector neutral concepts that can be used to organise data, including

- ORGANISATION
- PERSON
- COMMUNITY
- OBJECTIVE
- PLAN
- RESOURCE

... and describes how they can be related, e.g.

ORGANISATION has OBJECTIVE

PLAN delivers OBJECTIVE

PLAN impacts COMMUNITY

## 7.2 Data Dictionary

Concepts are not sufficient to describe the contents of data. An extensible Data Dictionary is required for the local data ecosystem, where specialisations of each relevant concept are defined together with

- classes;
- attributes;
- identifiers;
- code lists;
- classifications.

The Data Dictionary may refer to sources of data that are not definitively contained in the local data ecosystem. These are typically national reference data, and code lists.

The 'local data ecosystem' will have facilities to define and extend a Data Dictionary, to be applied to certain themes ( such as Troubled Families )

## 8 Syntax for the local data ecosystem.

### 8.1 Syntax for source information

Data will typically originate in a computer system, or in a spreadsheet.

The 'local data ecosystem' will have facilities to

- map source data to the Data Dictionary
- transform and import data using a series of bespoke routines for specific datasets;
- transform and import data using a configurable general purpose utility

Source formats may include

- spreadsheet style, e.g. csv;
- xml;

## 8.2 Syntax for data in the local data ecosystem

Data will be stored as a series of triples in named graphs, in RDF format, in the local data ecosystem.

## 8.3 Syntax for querying and extracting data from the local data ecosystem

The ability to extract data will be restricted based on the rights/authentication/transport/governance parts of the framework.

The local data ecosystem will be query-able using

- SPARQL, as a general query language
- Search parameters using terms from the Smart City Concept Model, the extensible Data Dictionary

Data will be extracted in a variety of syntaxes including

- flavours of RDF;
- csv
- xml
- json
- html

## 9 Data Quality for the local data ecosystem.

### 9.1 Provenance

Provenance information describes the processes and actors that data has passed through, such as collection, verification, audit, aggregation and so on.

The 'local data ecosystem' will include a standard structure for describing the provenance of

- Each dataset
- Each statement in a dataset

### 9.2 Fitness for Purpose

Provenance information describes expectations such as precision, accuracy, timeliness, completeness and so on.

The 'local data ecosystem' will include a standard structure for describing the fitness-for-purpose of

- Each dataset
- Each statement in a dataset

## 10 Rights for the local data ecosystem.

### 10.1 Establishing a right to data

For data to be shared, the data controller<sup>3</sup> needs to be assured that

*A **person** is **empowered** by their **organisation** to act in a **role** that has a **right** to a set of **data items** for a **purpose**, and agrees to the **terms** by which the data is to be used and handled.*

The 'local data ecosystem' will provide facilities to enable a data controller to

- define the 'purposes' for which data was originally collected;
- define 'purposes' for which data may be re-used;
- define the 'terms' for access to data;

... and associate them with a dataset.

### 10.2 Licencing data

Re-use of open data is typically via a licence, which is a simplification of the right to protected data. In that case, the 'Terms' are the Licence Terms.

## 11 Trust in the local data ecosystem.

### 11.1 Role based access

The 'local data ecosystem' will enable participating organisations to define 'roles' for their staff, and associate them with 'purposes'.

The 'local data ecosystem' will have a consistent vocabulary for determining the 'risk level' for:

- Loss of confidentiality

for a dataset, or collection of data.

the 'local data ecosystem' will enable a trusted organisation to assert that

*A **person** is **empowered** by their **organisation** to act in a **role** that has a **right** to a set of **data items** for a **purpose**, and agrees to the **terms** by which the data is to be used and handled.*

---

<sup>3</sup> [http://www.ico.gov.uk/for\\_organisations/data\\_protection/the\\_guide/key\\_definitions.aspx](http://www.ico.gov.uk/for_organisations/data_protection/the_guide/key_definitions.aspx)

... and that assertion must satisfy the risk level for the data.

The 'local data ecosystem' can maintain a log of assertions against which data has been accessed.

## 11.2 Anonymous access

Where the 'Risk Level' is appropriate, access to data in the 'local data ecosystem' can be anonymous.

## 11.3 Personal access

Where data is about an individual, or an organisation, there may be a right to data for that individual. Similarly, an individual may wish to provide information about themselves.

The 'local data ecosystem' will integrate with a Personal Data Store in which individuals can exchange data securely.

# 12 Transporting data over the local data ecosystem.

## 12.1 Network access

The 'local data ecosystem' will have a consistent vocabulary for determining the 'risk level' for:

- Loss of integrity
- Loss of availability

for a dataset, or collection of data.

Access to the 'local data ecosystem' will then be dependent upon the use of a network that meets that requirement.

# 13 Governance for the local data ecosystem.

## 13.1 Information Governance Measures

The undertakings that an organisation must agree to, to have access to certain risk levels of information can be described as 'measures'.

The 'local data ecosystem' will have facilities to define 'measures' and associate them to risk levels.

For open data, there will be no measures.

Other measures will be derived from existing governance regimes such as

- The Public Services Network Code of Connectivity
- Health and Adult Social Care Information Governance Toolkit.

## 13.2 Conforming

An organisation will need to achieve a set of measures to be able to access the ecosystem to a risk level.

This should re-use existing accreditations, and provisions that provide access to secure networks.