

Title	Registries and Inventories for Local Data A Technical Companion to the DCLG 'Code of Recommended Practice for Local Authorities on Data Transparency'.
Revision	6
Status	Draft
Date	15 th August 2013
Author	Paul Davidson, LeGSB Director of Standards, and CIO Sedgemoor District Council
Contributors	Mike Thacker, Porism Limited for Peterborough Council and esd-toolkit programme Catriona Taylor, Peterborough Council
Publisher	LGA convened local open data group with representatives from: • Department of Communities and Local Government • esd-toolkit programme • Local eGovernment Standards Body • Local Government Association • London Borough of Redbridge • Peterborough City Council • Sedgemoor District Council
Purpose	To propose options and opportunities for registries of published data, and inventories of all data, owned by Local Authorities in England implementing the DCLG code of practice. To propose requirements to data.gov.uk

Revision	Date	Author	Notes
1	25/07/2013	Paul Davidson, for LeGSB	
2	30/07/2013	Mike Thacker for Peterborough Council and LGA Transparency group	Add sections on "Requirements for data.gov.uk" and "Requirements for councils"
3	02/08/2013	Paul Davidson	Refinement of data.gov.uk requirements
4	05/08/2013	Mike Thacker	Improvements from meeting between Paul Davidson and Mike Thacker 5/8/13
5	13/08/2013	Mike Thacker	Modifications to address wider group feedback and note outstanding issues before passing to data.gov.uk
6	15/08/2013	Mike Thacker	Further modifications to reflect conclusions of tele-conference with the Information Commissioner's Office

The expectation on Local Authorities

The Department for Communities and Local Government, issued 'The code of recommended practice for local authorities on data transparency' in 2011, and will shortly issue a revised version for 2013, following consultation earlier this year. It includes a requirement for Inventories.

Local authorities should build and maintain an inventory of the public data that they hold so that people are able to know what is available to them. If public data would be released under Freedom of Information it should be included in the inventory. As this inventory is highlighted to the widest possible audience demand should grow and local authorities should expect to publish more.

These inventories should be registered on data.gov.uk to support a single point of access for all public data from national and local government.

The code of recommended practice for local authorities on data transparency

<https://www.gov.uk/government/publications/local-authority-data-transparency-code>

The broader vision is to enable data from multiple local organisations, including Local Authorities, to be more quickly and easily discovered, combined and re-used. Local data might include data on transport, crime, the environment etc to aid join-up between organisations.

The Government's response to the Shakespeare Review

(https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207600/Government_Response_to_Shakespeare_Review_of_Public_Sector_Information.pdf) commits government departments to publishing Inventories of **all** data that they hold. Local Authorities may wish to provide this information in their Inventories so that the public can request public datasets that can be derived from otherwise non-public data.

A code of practice on publishing datasets was issued under Section 45 of the Freedom of Information Act in July 2013, see

<http://www.justice.gov.uk/downloads/information-access-rights/foi/code-of-practice-datasets.pdf>.

The Information Commissioner's Office (ICO) has published a Model Publication Scheme; updated for new requirements from 1 Sept 2013, see

http://www.ico.org.uk/for_organisations/guidance_index/~media/documents/library/Freedom_of_Information/Detailed_specialist_guides/model-publication-scheme.pdf. ICO has also issued guidance on the kinds of

information that principal local authorities are expected to provide in order to meet their commitments under the model publication scheme at

http://www.ico.org.uk/for_organisations/freedom_of_information/~media/documents/library/Freedom_of_Information/Detailed_specialist_guides/definition_document_local_authorities.pdf. There is also a guidance document on using the definition documents at:

http://www.ico.org.uk/for_organisations/guidance_index/~media/documents/library/Freedom_of_Information/Practical_application/usingthedefinitiondocuments.ashx.

Local Authorities meeting the expectation to publish will need to create, maintain, and publish an index of their data containing

- An 'inventory' listing at least their public data, and potentially, all of the data that they own, sometimes referred to as an Information Asset register
- A 'registry' of datasets that have been published, as a subset of items in the Inventory, locating one or more formats, known as 'renditions', in which they are published.

Extra challenges for local government open data

Local government data presents extra challenges and opportunities that are not found in national/departmental data.

- local authorities provide hundred of types of service, and consequently, can publish a hugely diverse range of sets of data;
- there are 100s of local authorities, each publishing a subset of the same type of data, potentially in different formats;
- given the tiered structure of local government, each community can be served by a number of local

authorities of various types;

- the value of local data greatly increases if it can be combined and compared, across similar data from many sources;
- the context for each community and local environment is crucial in making meaningful conclusions.

A solution to Inventories and Registries implemented by data.gov.uk/local needs to accommodate these challenges.

Why would a Local Authority want to use data.gov.uk?

Local authorities may wish to use data.gov.uk to ...

- consume metadata about their own open data, for example using it to lay out open data pages at their own web site, or listing relevant open data in pages that describe or transact each local service. This would ensure that the metadata provided to data.gov.uk by a local authority, was always accurate and up to date;
- find data formats (schemas) that have been used by other local authorities for the same type of data, so that they can consider using the same formats leading to greater opportunities to join up data, and reduce the resource required to publish data.

Why would a local resident want to use data.gov.uk?

[Data.gov.uk](https://data.gov.uk) is the data portal promoted by government to signpost open public data. users benefit from being able to find data through one portal.

A local resident or user of local public services may wish to use data.gov.uk to ...

- discover data from each of the local authorities that serve their communities, in preference to searching through the open data pages of each separate council and local public sector bodies.

Why would a developer/innovator want to use data.gov.uk?

Someone who wishes to build an app or analyse data may wish to use data.gov.uk to ...

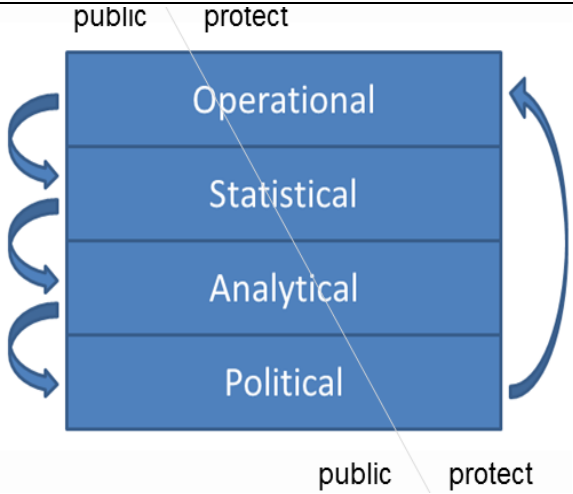
- discover and bring together consistent results across many local authorities, about selected types of data, services, and geographies.

Why would organisations such as a Business/Charity/Lobby Group want to use data.gov.uk?

- to perform alternative analyses for the purposes of development and location of new services and business opportunities
- to challenge statistical and analytical outputs

The contexts of data that can be published by Local Authorities

Local Public Data can be considered over four contexts.

	• Operational about real objects, people and places, with real needs and circumstances, using real services. i.e. Case Work. This is raw data about real objects, people and places as opposed to aggregated, interpreted data • Statistical aggregated operational information organised using common classifications and segmentations. This might include performance data. • Analytical the conclusions and assumptions drawn from an analysis of statistics, e.g. patterns, predictions, inferences, opinions. This represents interpreted data and might express a judgement on performance data. • Political/ Democratic the decisions taken to shape services, e.g. budgets, strategies, priorities, targets etc
--	---

Examples of local data across the contexts are shown at Appendix A.

Openness

Each dataset, will have been assessed by the owner as to whether it can be opened for public access or is subject to protection which may limit publication.

data.gov.uk defines public data at <http://data.gov.uk/opendataconsultation/annex-2> as follows:

“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.”

but the terms ‘public’ and ‘open’ have different means in different contexts.

It is simply necessary to identify if a dataset is protected or not and what the nature of any protection might be. Categories of protected data are:

Data ...

- containing personal information which is covered by the Data Protection Act

or

- for which there is a relevant exemption from legislation such as the Freedom of Information Act other than the Data Protection Act. This includes any sensitive data exempt under national security etc.

or

- containing Intellectual Property Rights (IPR), which itself maybe from a 3rd party, which requires special licence conditions or fees as described by the regulations of the re-use of public sector information. Data with IPR can still be released as open but there are limits on its re-use

Data that may be protected in one context, may become public when processed into another context. For example, data about people and their circumstances may be protected, but statistics about caseloads and segmentations of customers may be public, and valuable insight might be released. It would be useful to be able to make links across an Inventory to show where a non-protected dataset has already been derived from a protected dataset.

Classes of open data

To demonstrate meeting the expectations of the Model Publication Scheme, datasets published need to be classified under one of these headings:

Who we are and what we do	Organisational information, locations and contacts, constitutional and legal governance
What we spend and how we spend it	Financial information relating to projected and actual income and expenditure, tendering, procurement and contracts
What our priorities are and how we are doing	Strategy and performance information, plans, assessments, inspections and reviews
How we make decisions	Policy proposals and decisions. Decision making processes, internal criteria and procedures, consultations
Our policies and procedures	Current written protocols for delivering our functions and responsibilities
Lists and registers	Information held in registers required by law and other lists and registers relating to the functions of the authority
The services we offer	Advice and guidance, booklets and leaflets, transactions and media releases. A description of the services offered

Information might be further categorised according to document types given by the ICO under each of the above classifications.

Schemas and formats

A schema defines the structure of data and any guidance to which it conforms. Local authorities may publish datasets following centrally published schemas or use the schemas of other local authorities to allow datasets to be combined across administrations. There is therefore a need to be able to associate a schema with each dataset and to make schemas publicly available for reference and re-use.

Formats define the open or proprietary means of making datasets human and machine readable, for example a spreadsheet might be published as comma separated variables (CSV) or in the proprietary Microsoft Excel format.

Categorising Local Authority Datasets

In addition to the existing metadata provided by data.gov.uk, data published by Local Authorities can usefully be further classified by

Classification	Description	Enabling
The Local Authority, and consequently the geographic area that it serves.	In addition to categorising the data against the name of the data publisher, it is also useful to link the Local Authority to the geographic area it serves. There is a well established URI set for Administrative Geographies from the Ordnance Survey, which could be linked to a URI for the Local Authority. The Local Authority can also be categorised by Type e.g. • District Council • County Council • Unitary Authority ... and so on.	Searching for data using a map. Finding datasets from many councils who cover the same area.
The Function of Local Services	The esd-toolkit (http://standards.esd.org.uk) provides lists that have been created by local authorities, of types of local functions (such as Waste Management), and the types of services within each function (such as Household Waste Collection, Syringe disposal, and so on). Each service is linked to the Power/Duty that gives rise to the type of Service, the type of council that provides each type of service, and to a Navigation term that is commonly used on council web sites.	Classifying local data by Function from a well known list, gives an ability to list data of the same type from across many councils.
Data Format (schema)	Tools are emerging across local authorities that enable staff to create a 'schema' describing the format and meaning of the columns that they use for each dataset. These schemas could be registered at data.gov.uk and associated with the appropriate function (above). Local Authorities can then indicate the 'schema' that they have used when publishing their data. A local domain of data.gov.uk could have forums where local authorities, and those using the data, could promote and discuss these schemas.	Data from many sources that use the same schema can confidently be joined up. Councils looking to publish data of a certain type can look to see what schema(s) have already been used by other councils, avoiding unnecessary divergence of formats and identifiers for the same type of data. Users of data can take part in a structured debate across local government about the content and format of each type of dataset.

Using data.gov.uk to publish Inventories and Registries

Local Authorities may wish to publish their Inventories and Registries, as 'metadata' at data.gov.uk, in two modes

by directly keying data into	• the definitive version will be held once at data.gov.uk;
------------------------------	---

forms at data.gov.uk;	the local authority will want to consume the metadata so that it can list datasets on its own local data pages;
by repeatedly loading data to data.gov.uk from metadata held locally.	- the definitive version will be held locally by the local authority; - the local authority data may be 'richer' (i.e. have more elements) than can be accommodated by data.gov.uk; - a local authority may have more than one local inventory and registry, although a dataset will appear only once, and asset references for will be unique.

For either mode, an open standard for the 'schema' of the metadata will be required so that

- metadata about datasets can be read from data.gov.uk and represented onto local web pages and apps;
- metadata from many local authorities can be regularly harvested (pulled) by data.gov.uk to update their copies for publication;
- depending on the frequency of harvesting, a local authority may wish to 'push' updated metadata to data.gov.uk, via an api;
- a query at data.gov.uk can return with machine-readable results from many local authorities;
- organisations other than data.gov.uk to build lists of Local Authority data;

Requirements for data.gov.uk

To facilitate councils' compliance with the recommended practice's requirement for Inventories, the following functionality is needed from data.gov.uk:

1	An ability for a public agency (a local authority in particular) to be associated with: - the administrative area or geography that it serves; - the type of local public agency. This is to enable a search (see below) that includes a location (e.g. a post code), and replies with results that are from local agencies that serve that location. i.e. the query does not have to include the name of the publisher. Note that the Department for Communities and Local Government (DCLG) already provide this information as URIs at http://opendatacommunities.org/ , for English local authorities.				
2	An ability for a public agency to nominate one or more public URLs where metadata in a standard machine-readable format can be found, that can be used to import Inventory and Registry information.				
3	A means of pulling from these URLs data identifying for each dataset: <table><tr><td>• Identifier</td><td>which will be unique within the public agency</td></tr><tr><td>• Subject(s)</td><td>from one or more pre-determined standard vocabularies (lists). Including The Local Government Function List </td></tr></table>	• Identifier	which will be unique within the public agency	• Subject(s)	from one or more pre-determined standard vocabularies (lists). Including The Local Government Function List
• Identifier	which will be unique within the public agency				
• Subject(s)	from one or more pre-determined standard vocabularies (lists). Including The Local Government Function List				

	http://id.esd.org.uk/list/functions The Local Government Service List http://id.esd.org.uk/list/services Note that a mapping exists listing the 'service' list entries that are relevant to each 'function' list entry. Some datasets are not relevant to a function or service (typically those that describe local communities), so these are not mandatory.
• Description	As free text
• Protected	Protected or not and type of protection from: - protected personal - protected other Fol exemption - protected intellectual property
• Context	One of: - operational - statistical - analytical - political
• Licence	URL to a web page describing the licence, eg the Open Government License published by the National Archives
• Charge	Is a charge levied and the URL of the page describing applicable charges
• Certificate	URL to an Open Data Certificate that gives guarantees and undertakings about the continued availability and quality of the data
• Derivations	A list of identifiers to other entries to indicate if a dataset has been derived from other data in the Inventory.
• Other properties agreed following consultation with LAs	These are likely to be drawn from DublinCore and the existing CKAN metadata card. and INSPIRE

The information above is suitable for Inventory Items. Where data has been published, further 'rendition' information is added. There are typically many renditions of each dataset to provide for

- datasets covering new periods
- a choice formats.

For each rendition

• Period	The date range, or instant, (if applicable) over which the data is applicable. E.g. For payments data where a new file is produced each month.
----------	--

	Location	URL web where the data can be accessed							
	Format	From a controlled list of formats such as csv, xls, odt etc							
	Schema	Linking to a separate registration of schemas (see later)							
	Availability	One of - Download - Service - On application where <table><tr><td>Download</td><td>A copy of the entire dataset can be downloaded and stored locally.</td></tr><tr><td>Service</td><td>Items from a dataset can be queried via an api, and a result set returned.</td></tr><tr><td>On application</td><td>The dataset is not available online, please contact the owner.</td></tr></table>		Download	A copy of the entire dataset can be downloaded and stored locally.	Service	Items from a dataset can be queried via an api, and a result set returned.	On application	The dataset is not available online, please contact the owner.
	Download	A copy of the entire dataset can be downloaded and stored locally.							
Service	Items from a dataset can be queried via an api, and a result set returned.								
On application	The dataset is not available online, please contact the owner.								
Other properties agreed following consultation with LAs									
Where there are no renditions, the data is not yet available as open public data. A metadata schema will be published for this purpose. Many properties will be optional.									
4	An ability to register a 'schema', and associate it with - one or more entries from the Local Government Function List; - one or more entries from the Local Government Service List; - an existing registered schema that it has been derived from.								
5	Online user-driven search and filtering from data.gov.uk/local of the index by parameters including: - Organisation - Geographic location - Subject (eg 'function' or 'service' from controlled lists) - Date - Text within description - Openness - Context - Schema								

	Other parameters to be agreed
6	A public API whereby anyone can extract selected records from the index using the same search and filter parameters as are available in the user interface
7	Ability to download an index of datasets with all entries for all organisations.
8	A list of inventories

Note that non-data requirements, such as publication of guidance and facilitation of discussion, are beyond the scope of this document.

Requirements for councils

To comply with the recommended practice's requirement for Inventories, councils will need to:

1	Inform data.gov.uk of the location of their machine readable inventories/registries by an agreed means
2	Publish inventories/registries at those locations according to the agreed metadata schema
3	Publish datasets at locations and in formats specified by their registries

Appendix A - Examples of data at each context

Car Parks

Operational	Where are Car Parks? What spaces are free now? Individual Tickets and Fines	GIS and real-time monitoring. Ticket machine data
Statistical	Average number of free spaces at locations and times of the day/week. Income and Expenditure on Car Parks, year to date, previous years.	Locations, Times. Accounting
Analytical	Capacity of car parks to meet retail and employment needs.	Correlation to Town Centre shopping and employment data
Political	Fees and charges Target income/expenditure Car Parking Strategy	Budgets, Outturn Decisions on future capacity, charges, 'park and ride' initiatives.

Food Premises

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 is being done to make our food outlets safer?	Targeted training programmes. Risk Based inspections.