

## Local Government Open Data Breakthrough Projects 2014/15 Evaluation Report

<b>Project title:</b>	<b>Leeds Data Mill</b>
<b>Lead organisation:</b>	<b>Leeds City Council</b>
<b>Date report is submitted:</b>	<b>30<sup>th</sup> April 2015</b>
<b>Type of project:</b>	<b>Open Data Release, Platform/Tool Development, Engagement and Data Usability.</b>
<b>Total grant:</b>	<b>£165,000</b>

---

---

### 1. Summary of the project

The project is designed to continue the release of open data from Leeds City Council (LCC) and other external organisations and to derive the most value out of data within Leeds. The work included training, workshops, and internal “hack” events. Platform enhancements have been developed to support the release of data, and community involvement, which are open source by default so other cities can re-use them. Building on the initial community, the project has enabled full engagement.

As well as bringing data owners on board (public, private, and third sector) the project has enabled assets that make open data engaging and understandable to a wide range of users.

Dashboard functionality has enabled the transformation of complex data into understandable story images which visualise the data. Each image is created by users and linked to different themes (health, transport etc.). The project has enhanced the dashboard and created an open data “Story Hub” which houses developed stories so users can create their own themed dashboards.

### 2. The Challenge/opportunity:

Most local authorities only publish data that they are mandated to release. This means that most city councils have significant amounts of data which is not yet released as open data. Releasing this data will improve transparency, increase insight and stimulate economic growth.

This project is designed to:

1. Accelerate the release of data from across the city led by LCC.
2. Build an open data community of SME’s, developers, analysts and enthusiasts (both within council as well as external)
3. Develop a culture of openness within data managers and users from across the city.

4. Enhance the Leeds Open Data Platform (Leeds Data Mill) by adding new features such as ODI Certificates and Local Government Schema Integration.
5. Make open data understandable to all, creating insight on key city topics
6. Empower developers of all skill levels to turn Data into stories and form a simple dashboard

### **3. Approach to publishing data**

Data on the Data Mill is published in an open and re-usable format. Leeds City Council has informally adopted CSV (Comma Separated Values) as this standard, though others are in common use.

As a general principle published Data does not contain any personal or business sensitive data and is published to as granular level as possible.

Leeds City Council, where possible, is publishing its datasets using an Open Government Licence. This allows developers to re-use the data in any way they wish, subject to a few conditions, and protects the council at the same time. Other similar licences are available (e.g. Creative Commons).

The Data Mill has developed a guide to anonymisation and open data principles which data owners are asked to apply. These guidance documents have been published on the Data Mill to allow others to use. Leeds City Council is currently working on updating these documents and creating an Anonymisation Code of Practice which will also be published once completed.

### **4. Open Data Awareness**

The Leeds Data Mill is actively publicised both inside the Council and in the wider City. Leeds Data Mill has strong links with the Open Data Institute (ODI) which has a node in Leeds. ODI Leeds promotes and raises awareness of Open Data and in particular the Leeds Data Mill.

ODI offers a series of Open Data training courses and is the space that it is used for Leeds Data Mill events. Within LCC regular Open Data awareness events are held attended both Officers and Elected Members.

### **5. Benefits and Impact**

In order to drive out value from open data Leeds Data Mill takes an active approach by engaging with the city community to work on city problems and locate supporting datasets. The aim is to bring about culture change within organisations which will lead to increasing amounts of data being published and the delivery of new and innovative services via websites, apps and data analysis.

From an LCC point of view this means engaging with service directorates (e.g. Environment & Housing, City Development etc.) promoting the 'Open by Default' principles and working with them on potential areas of development.

This approach results in real problems being solved and increases the flow of data which can be used to identify solutions. Bringing together data owners, directorate managers, local SME's, developers and designers into a common space (Leeds ODI) enables new insights into solutions and new ways of working.

Publishing open data can have a positive impact on reducing FOI requests. In line with changes to the FOI Act in September 2013, any datasets published resulting from an FOI request are always published on the Data Mill and maintained going forward.

#### **4. Open Data Platform and User Engagement Enhancements**

The Leeds Data Mill is built using an Open Source approach. The core platform is CKAN ( the same technology that is supporting the London Data Store and Data. Gov.uk). This project has enabled a new user interface to be developed which is easy to use and enables users to request new data sets and supports, blogs, projects and community engagement.

The project has also enabled the development of a dashboard solution allowing users to create data visualisations from open data which can be published and shared with others.

Full details are available in the appendix.

#### **5. Data published and uses:**

**The Data Mill has over 160 data sets from a range of organisations:**

- **Department for Communities and Local Government (DCLG)**
- **Leeds and York NHS Partnership Trust**
- **Leeds City Council**
- **Northern Powergrid**
- **SuperFast West Yorkshire (DCMS)**
- **UK Sport**
- **West Yorkshire Combined Authority**
- **West Yorkshire Fire and Rescue**
- **Yorkshire Water**

**Usage Stats:**

- **Page Views Since Launch (07/03/2014) 189,894**
- **Page Views in March 2015 17,395**

**The organisations providing the most data sets are:**

- **Department of Communities and Local Government (DCLG)**
- **Leeds City Council (LCC)**
- **Northern Powergrid**
- **West Yorkshire Combined Authority (WYCA)**
- **Yorkshire Water**
- **West Yorkshire Fire and Rescue Service**

**There are 246 registered users of the Data Mill**

**The most popular data set tags are**

- **Residents and Communities**
- **Business and Economy**
- **Tourism and Shopping**
- **Streets, Roads and Transport**
- **Housing**
- **Running of the Council**
- **Environment**

## Appendix 1: Project Information

### A1. Open Data Platform and User Engagement Enhancements

The Leeds Data Mill is based on the following technical environment which has 3 elements:

- **A core CKAN (V2.3) based installation**
- **A Word Press based user interface**
- **An independent dashboard service**

#### A1.1 Core Platform Development, Support and Maintenance

Hosting, technical support and maintenance for the Leeds Data Mill is provided by ODI (London) based company **Data Press** who specialise in CKAN and supported the development of Data.Gov.Uk and the London Data Store.

**Data Press** maintain the CKAN core version (Open Source) and have been commissioned to develop new features and enhancements.

The project has enabled a number of enhancements to the Core CKAN service:

##### a) Data Requester Service

The Leeds Data Mill now includes a 'Request Data' page. This allows requests for new data sets to be submitted by registered users and is similar to that available on Data.Gov.UK.

Data set requests are monitored and actioned by the information governance team within LCC. The Requester service includes updates on progress.

##### b) CSV Validation Integration

The Data Mill now includes access to a CSV Lint tool. This tool allows CSV format files to be checked via a background checking process and can incorporate schema validation. This is included in the Data Mill upload workflow.

More details on the ODI CSV Lint tool can be found in the appendix.

##### c) ODI certificate integration

ODI Certificate support has been incorporated in the data set metadata on the Data Mill and includes the ODI logo where a certificate has been created. Further development work (although not part of this project) will continue on this to fully integrate the certificate process into core CKAN.

More details on ODI Certificates can be found in the appendix.

#### **d) Inventory integration (Schema)**

The XML Schema for Leeds Data Mill has already been incorporated in the esd website and conforms to the inventory schema

The inventory details the datasets published on the Data Mill and supported schemas.

#### **A1.2 User Engagement Enhancements**

The project supported the development of a new Data Mill user interface and Dashboard Service The work was undertaken by local media company HeBe Works. Link: <http://www.hebeworks.com/>

#### **e) User Interface**

Although direct access to CKAN is possible the interface can be more suitable for a technical user. Configuration requires a level of technical knowledge which is not always readily available. The Project has enabled the development of a feature rich accessible interface with Project and community elements. This is the main Data Mill access page and is fully integrated with the CKAN core service. Access to data and user registration are supported by CKAN. Word Press and CKAN elements have been configured with the same branding

This interface has been developed in Word Press and is local administrator configurable. This means that changes can be quickly implemented, blogs and project updates easily maintained and user comments, requests and data uploads simplified.

#### **f) Dashboard and Story Hub**

Although direct access to Data Sets on the data mill is fully supported, not all users have the ability to view and integrate data sets (which may have a number of different formats and interfaces) into their own local environments.

To enable all levels of user to have a least a basic view of all data sets a Dash Board service has been developed.

The Dashboard concept enables a visualisation of any data set. These visualisations are attached to an underlying data set and present the data as a picture. The service supports a number of different views e.g. a chart, a graph, a map etc.

#### **The Dashboard**

The Dashboard has been developed using Open Source Java Script libraries.

The Dashboard is an independent web application which accesses Data Mill data sets.

The Dashboard has a default view when first launched – this is an LCC curated series of visualisations of key corporate data sets e.g. Air quality, recycling etc. Registered users can start to build their own ‘curated’ dashboards based on the data sets of most interest to them.

Dashboards (Also known as a Canvas) can remain private or can be ‘published’ i.e. made available for others to use directly or extract specific visualisations (also known as stories) for re-use perhaps to augment with other data.

### **The Story Hub**

Individual visualisations (Stories) can be easily created by anyone using a ‘wizard’ approach which is a core part of the Dashboard service. In addition more advanced stories can be created by developers using open source development tools. These stories can also be ‘published’ and will therefore be available for use by any user.

### **Dashboard Value**

The Dashboard approach can deliver value to citizens and communities enabling users to build dashboards reflecting local and personal interests.

Like many public sector organisations LCC has a high volume of data but much of this is in departmental silo systems. By releasing Open Data on to the Data Mill and reflecting that back through themed Dash Boards managers within LCC will gain valuable new insight into cross Council issues.

**Link:** <http://dashboard.leedsdatamill.org>

## **A2. Meta Data and Data Set Details**

Metadata tags are in place for each data set which aid categorisation and searching. In addition each dataset has key information about the data and details of the **Author, Maintainer, Update Frequency and Next Review Date.**

## **A3. Data Set Formats**

Data Sets on the Data Mill are presented in a range of formats depending on the source and intended application. As a general rule all data is available as CSV but may also be available in additional format types.

### **Format Types Supported:**

- CSV
- XLSX

- XLS
- XML
- XSD
- API
- PDF
- ODT
- JSON
- ZIP

## A4. Case Studies

Over the course of the project several applications and services have been developed under the banner of the Leeds Data Mill. These both use and create open data.

The approach to these solutions is based on a 'Data Dive' methodology. This is a series of engagements gaining an understanding of key issues and supporting data. The Data Dive itself includes developers, designers and data owners working together to discover new ideas and service delivery models. The services below were developed using the Data Mill Data Dive approach.

Details of the Data Dive approach can be found on the Data Mill

**Link:** [Data Mill Information and Guidance Documents](#)

### A4.1 Leeds Art Crawl

Leeds Art Crawl is the result of a request from City Development for help to locate public art across the City. A local digital SME (WetGenes) stepped up and created the Artcrawl service.

Leeds Art Crawl is a tool to generate real-time location based data about art. People take photos of art they find within Leeds, and upload them to twitter (with location services turned on) tagging each entry with #leedsartcrawl. Over time a comprehensive database of places where art lives is created helping us not only know where art is within the city, but also what it looks like, and their current state of repair.

Meta-information attached to each tweet forms datasets with the following information:

- Twitter username
- Profile ID
- Date
- Latitude and longitude
- Avatar picture
- Art photo
- Link to the original tweet

**Link:** <http://leeds.artcrawl.club/>



## A4.2 Find My Bin Day

Many people in Leeds are unsure which bin goes out on which day despite considerable efforts by the Council to communicate this information.

Local media company HEBE Media produce a local newspaper (The City Talking) which focusses on cultural life in Leeds and has a large audience. They felt that by taking our bin route and bin day datasets they could create an online application linked to The City Talking which would be used by their subscribers.

Feedback has been very positive and the application is to be extended to include automatic reminders.

This is an example of self-initiated project from the local digital community using established data sets.

**Link:** <http://www.findmybinday.com>

## A4.3 Viaggiart (Tourism),

Italian company ViaggiArt (<http://www.viaggiart.com>) develop a tourist application covering city events and destinations delivering information based on location.

Having discovered the Leeds Data Mill they have developed a Leeds version using cultural and other data from the Mill.

This is again an example of where an established company has extended their service using open data from the Leeds Data Mill.

**Link:** <http://www.viaggiart.com/en/search?q=leeds>

## Other Applications

Other applications are in the early stages of development and prototyping

## A4.4 Roadee

Roadee is an application developed by Professor Berringer and creative technology students of Leeds Beckett University. Measuring the surfaces of roads in Leeds has a very high cost and does not cover the whole area. This process is designed to discover potholes and roads that are about to deteriorate.

Roadee is designed to utilise sensors present within modern smart phones to measure the roughness of the road as a vehicle travel around the city.

The application will be loaded onto LCC Android smart phones and placed within a range of council vehicles. This will essentially crowd source a view of

the state of roads across the city and return all the data as a heat map via Leeds Data Mill. Highways can then use the data to prioritise repairs and preventative maintenance.

#### **A4.5 Route Rater**

The cycling community in Leeds is concerned with road safety and reducing accidents.

Leeds ODI held a cycling hack (Hack My Route) using data from Leeds Data Mill. This included cycle bay locations, cycle lanes, passenger transport information and accident data from across the city. The result was an idea based on crowd sourcing cycle routes (via an app) and rating them for safety and accessibility.

Route Rater is an app which uses a rating system similar to that used on ski slopes e.g. a Black route is more dangerous and should be used by more experienced users.

Cyclists can create new routes and share with the wider community and less experienced riders and families can find safer easier routes which could be more suitable for them. It also enables riders to pass on useful information such as where the busy junctions are or where there is a good café to stop off at.

**Link:** [Route Rater Blog](#)

## **A5. Data Visualisations**

The project has explored different ways of visualising data.

### **A5.1 Sports Centres**

This is an interactive service showing activities at sports centres

At Leeds City Council we have lots of information on our website pages about the activities and sports on offer at our various leisure centres. We wanted to see how newly published data could present this same data in a new and interactive way. Working with Hebe Works, we published the data on Leeds Data Mill and then considered a different viewpoint.

This visualisation shows how data can be turned into something which is really easy to use. Simply by hovering over a leisure centre or activity, you can easily find out which sports and activities are available at each leisure centre.

This technique is useful where multiple services\ activities\ skills etc. can be shared across many entities.

### **A5.2 Air Quality**

Air quality monitoring is becoming a significant issue for many cities – particularly as there are large fines for exceeding mandated CO2 levels. Leeds has Air quality data recorded for over 10 years and this is available via the Data Mill.

This visualisation consists of a number of videos showing journeys into Leeds from different traveller's viewpoints. These travellers have personas and use different modes of transport and different routes.

As they travel air quality data is displayed for the various parts of the route.

This approach is used to illustrate how transport choices have an impact on the environment and specifically air quality.

### **A5.3 Tour De France**

The success of the Tour De France last year had a significant economic impact on the region and generated a large amount of data. Thousands of people travelled to Yorkshire to view the race and also the other attractions of County. The data includes hotel occupancy rates, cycling growth and footfall.

This visualisation explores the areas of the country people travelled from to visit Yorkshire and further work will explore the legacy impact on Tourism

These visualisations are available via links from the Data Mill.

## A6. Contacts

Please visit: [www.leedsdatamill.org](http://www.leedsdatamill.org)

### Key Contacts:

- Ian Jones Solution Architect (Smart Cities)

[ian.jones@leeds.gov.uk](mailto:ian.jones@leeds.gov.uk)

- Stephen Blackburn Open Data Lead LCC.

[Stephen.blackburn@leeds.gov.uk](mailto:Stephen.blackburn@leeds.gov.uk)

## Key Links

### CSV Lint (ODI):

Link: <http://csvlint.io/about>

### Data.Gov.UK:

Link: [www.data.gov.uk](http://www.data.gov.uk)

### Data Mill ODI Certificate Example

Link: <http://leedsdatamill.org/dataset/public-toilets>.

### Data Mill Inventory:

Link: <http://inventories.opendata.esd.org.uk/153>

### Data Mill Guidance:

Link: [Data Mill Information and Guidance Documents](#)

### Data Mill Requester

Link: <http://leedsdatamill.org/request-data/>

### Data Press:

Link: <http://datapress.io/>

### ESD Inventory Schema:

Link: <http://schemas.esd.org.uk/inventory/inventory.xsd>

### ODI Certificates:

Link: <https://certificates.theodi.org/about>