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

Project title: Structure of the Managing and Publishing of Open Data

Lead organisation: City of York Council

Date report is submitted: 30TH April 2015

Type of project: Data Release / Standards / Engagement

Total grant: £50000



Summary of the project

The City of York Breakthrough fund project proposed to provide an infrastructure to enable CYC to automate collection, quality checking and formatting of data to enable open data release. The activities covered by the funding application included:

- Development and deployment of a data warehouse architecture to provide a core for fully functional enterprise content management (ECM)
- Development of common standards of data attributing
- Data formatting, standardisation, attributing and migration
- Development of automated collection and publishing of open data application (from within data warehouse)
- Creation of connectivity with external open data portals, including data.gov.uk, by exploring API capabilities within these portals
- Training of key personnel
- Project review and evaluation, and identification of further developments and improvements.

The breakthrough project was linked to the council's ambitions to produce a York Open Data platform and develop content. The open data platform was co-created with various partners in the "Media Mill" and one of the key drivers of this project is the identification and transformation of data into local media content. Other drivers are making open data accessible, encouraging its exploitation, and also engaging other public and private organisation data holders in order to facilitate contributions to a wider range of data sets.

Challenge / Opportunity:

The joint York/ Leeds project provided opportunities for the creation of a powerful and high profile open data platform with content in York. However in order to effectively use a data platform the twin problems of preparing and loading data onto the site and maintaining it over time, in the context of the continuing effects of significantly reduced council staffing and capacity had to be addressed.

The city council implemented automation of the council's key performance indicators systems (KPIs), which transformed the methods of holding performance information from their operational formats into a central SQL server database held by the business intelligence hub, with the aim of streamlining KPI processes and significantly improving data quality. This data automation could be extended to transferring open data files to an open data platform. Rather than needing ongoing funding to support regular creation, updating and loading of open data files onto a platform, the automation approach required some one-off funding to build a data warehouse of open data and

potential open data, and undertake the necessary data formatting and integration programming function. This process was designed to be low maintenance.

In setting up an open access platform, populated by an automatic process with its own datasets the Council needed to transfer data across the public sector network, the firewall which encloses public data, without allowing access in the opposite direction. The Council selected the CKAN open source, open data standard for its open data database and platform.

Data published and uses:

Number of datasets published

During this project, CYC has reformatted, cleaned and structured a total of 285 datasets. These datasets are kept in the new database that was developed on the new server the Council acquired thanks to the Breakthrough funding and are sent there automatically through a number of integrations set up internally between council systems.

CYC have tried to publish a wide-range of datasets during the initial stage. Internally there have already been agreements to try and release more data related to “Transparency” in a higher quality format than previous releases, as well as greater release of school admissions and transport data. Examples of these are:

- Allotments
- Arts and Entertainment
- Bus Stops
- Care Homes
- Dog Bins
- Footfall
- Parks and Open Spaces
- Playing Areas
- Polling Stations
- Primary School by Language
- Primary Schools Enrolment
- Public Toilets
- Salt Bins
- Sport Centres
- Waste Recycling Sites
- Transparency related datasets
- KPIs related datasets

For a full list please see Appendix A, Section II of this document.

Volume of data downloads within period

The platform went live on 24th March 2015, and so it is too early to gauge the usage of the platform to date. York has various partnership arenas which have shown interest in whether there is a public/private/resident appetite for the platform in York and so various boards from its Local strategic Partnership, Without Walls, through to its Digital board, have asked for usage to be reported.

The total number of datasets downloaded can only be measured as the volume of datasets that have been downloaded from the public platform www.yorkopendata.org.

- Total amount of downloads 24-3-15 to 31-3-15 = 132

It is also worth mentioning that the CKAN technology which supports this platform also offers the possibility to explore the data and consult it without needing to download it (“preview”) and this may be a better indicator of usage.

- Total amount of previews 24-3-15 to 31-3-15 = 926

Most often downloaded datasets

- Do it Online – New users
- FOI Elderly Persons Home
- Public Toilets
- Allotments
- Protect Vulnerable People Scorecard
- %Primary Schools 25% or more places unfilled
- Avg days to remove obscene graffiti
- Business Births
- YCC % Calls answered in 20 seconds
- Dog Bins

For further information please see Appendix A, Section I of this document.

Number of users within period

Visits to the platform between the 24th of March and the 31st of March;

- 380 sessions corresponding to new users visiting the platform (first time)
- 187 users who had already visited the platform

Approach to publishing data

Approach used to publish data

When facing the task of publishing data the Council holds, the majority of the open data team did not have a wide experience of council working and had come from technical backgrounds so the first challenge was to try, with guidance from the Business Intelligence Unit management team, to make an inventory of possible and requested data held by the different council departments. At this stage the team also needed to keep track of who was responsible for each set of information, in which format was it kept, how often it was updated and which systems and server the data was being held in. To achieve this many heads of departments were asked to have 1:1 meetings where issues such as data quality and sensitivity were discussed, and a log of which data they held was compiled. Regular meetings were held between individuals who had further interest in open data and members of the Business Intelligence Unit.

After this process the next thing to achieve in order to publish data was to look at some of the existing requests to publish datasets and come up with a varied mixture of data-sets for launch. The datasets were chosen based on a mixture of the compulsory Transparency Code, ease of publication and interest registered in some datasets by potential users of the platform.

After picking a dataset, or collection of information, its quality is checked, it is cleaned and depersonalised if needed and reformatted into a new dataset. A schema is developed for the new dataset and it is kept in the database of the new server.

Integration packages have been developed and deployed on this new server. These automatically send the Open Data datasets to the public server where the public platform sits, and there another script runs and publishes those datasets to the platform.

How data users assess the availability and use of data

- Presentation of concept in partnership arena
- Internal meetings with people from the Business Intelligence Unit
- Internal meetings with other departments key personnel
- Presentation at Breakthrough Fund event
- Bringing Hebe Works – a media company based on Leeds / Leeds City Council to York based events to share experience and understanding.

-
- Information on social networks such as Facebook and Twitter and press releases
 - Council's chief executive speech at LARIA 2015 Conference
 - Stall at LARIA 2015 Conference
 - Public events at West Offices
 - Series of public events planned for the near future, such as a Hackathon at the end of May 2015

Ease of access to the datasets for use and download

One of the reasons to choose CKAN as the technology underlying the open data platform is that it is the mainstream technology when it comes to opening up data and many users of other open data platforms are used to its looks and feel. The Council also tried to keep the rest of the webpage as simple as possible, focusing on the data, which meant that the open data team could concentrate energy on providing high quality datasets.

Metadata

Some metadata is published along with the data but full dataset schemas have not been published yet as the ESD team is developing a new functionality for local authorities to register them under <http://opendata.esd.org.uk/>. They are ready to go as soon as it is working.

Sustainability and further plans for improving the system

One of the key drivers for the platform, in the current climate, is that the platform is set up sustainably, so that if any ideas or products come from information on the platform, residents and businesses have the confidence to follow through these ideas knowing that the platform is still going to exist in 3 – 5 years and onwards. Therefore when the technological development for this project implemented was put in place, significant focus was put on attempting to complete this with low costs for future maintenance and support. Examples of this were purchasing servers with enough capacity for several years

York looked in to the experience of other local authorities' approaches to open data and decided that a big bang approach, and then significant drop-off of datasets, was not an approach that it wanted to follow. This is not to say that this approach has not been effective in encouraging the use and growth in open data usage, it is just that York would like to provide fewer, more high quality datasets in a sustained and regular release pattern. Part of this regular release has been held up by "Purdah" due to local council elections which started a fortnight after the platform launch but will re-start in May and June. It is also hoped that through the internal and external events feedback will be received on which datasets businesses and the public will want to see released more regularly

Benefits and impact

It is too early still to provide evidences of the advantages of opening up City of York Council's data so it is worth looking at the original strategy aims to harvest benefits in different ways:

- Reduction of FOIs, as the data will be out there for people to freely consult it. This saves time to public staff which can be used to improve other services.
- Better informed citizens who gain understanding on why decisions are taken and what happens around them.

-
- Economic growth. Create data driven local businesses or new products based on data developed by existing ones.

Report Written by: Claudia / Thomas / Ian

Further information: Ian Cunningham, Group Manager–Shared Intelligence Bureau
t: 01904 555749 | e: ian.cunningham@york.gov.uk

Website: www.yorkopendata.org

Appendix A: Datasets

Section I: Most Popular Datasets



Position	Dataset	Topic	Link
1	Do it Online – Number of new users created	Transparency	http://data.yorkopendata.org/dataset/kpi-ycc27
2	FOI Provision on Elderly Persons Homes	Freedom of Information	http://data.yorkopendata.org/dataset/foi-ccf-5485
3	Public Toilets	Crime and Community Safety, Environment, Health	http://data.yorkopendata.org/dataset/public-toilets
4	Allotments	Environment, Health	http://data.yorkopendata.org/dataset/allotments
5	CYC Protect Vulnerable People Scorecard Q3 2014-15	Transparency	http://data.yorkopendata.org/dataset/priority-board-q3-2014-15-vulnerable-people
6	% of Primary Schools with 25% or more of their places unfilled	Education, Young People	http://data.yorkopendata.org/dataset/kpi-rm3
7	Average days taken to remove obscene graffiti	Crime and Community Safety, Environment	http://data.yorkopendata.org/dataset/kpi-can008
8	Business births	Business and Economy	http://data.yorkopendata.org/dataset/kpi-cige28
9	YCC % Calls answered in 20 seconds	Transparency	http://data.yorkopendata.org/dataset/kpi-ycc117
10	Dog Bins	Crime and Community Safety, Environment	http://data.yorkopendata.org/dataset/dog-bins

A list of datasets present in the database developed on the new acquired server thanks to the funding is publically available on www.yorkopendata.org