

Case Study:

data.gov.uk

Type: Website

Organisation(s): HM Government

Tags: open data, process, metadata, standards

data.gov.uk | Find open data

[Data.gov.uk](https://data.gov.uk) is a datastore that features datasets published by

government departments, local councils, and many other public bodies. It was established with the vision of building greater levels of trust with citizens, whilst delivering more cost-effective and better targeted and tailored services to their needs.

The platform has grown since its establishment in 2010 to encompass hundreds of organisations and tens of thousands of datasets.

Background

Data.gov.uk was launched publicly in 2010 and was intended as a repository for non-personal data from across all UK government departments. This was underpinned by the establishment of the [Open Government Licence](#) later that year, as well as the founding of the Government Digital Service in 2011.

This was partly driven by the government transparency agenda, but it also had an economic strategy component. There was a hope that there would be a strong drive to make use of public datasets in innovative ways. The National Audit Office estimated that [public data was worth £16 billion to the UK economy in 2012](#). The same report also estimated that openly publishing government data that was currently traded, including from the Met Office and Ordnance Survey, would be worth £1.6bn to £6bn to the UK annual GDP.

By summer 2020, data.gov.uk contained reference or links to almost 55,000 datasets, although not all of these are available. Within each of these datasets, there are often several separate formats, years or areas covered. For example, some datasets are released periodically, and each release falls under one dataset. Some of these datasets are added directly to the website, while others are webscraped.

The [Government Transformation Strategy 2017- 2020](#) encompassed the goal of improving standards of quality for the website. These standards are defined by the [Design Principles](#), the [Digital Service Standard](#), and the [Technology Code of](#)

[Practice](#). The overall strategy is aimed at increasing the amounts of data shared between agencies and departments, as well as building government digital services around the needs of its users.

Data.gov.uk has a CKAN API. In comparison to some APIs, users do not need an API key and there are also no rate limits to usage. The portal also enables users to directly request further datasets to be published if they cannot find what they are looking for. 2019 recorded a very significant increase in API requests, with [382,000 API requests made in May 2019](#) alone, suggesting increased awareness of the opportunities around government data.

The platform is one of the largest public data portals in the world. The UK government considers itself a world leader in the field of open data. For instance, Open Data Barometer awarded the UK first place in European and Central Asia in its 2017 [Open Data Barometer report](#). Therefore, the UK has a reputation to maintain by the preservation and enhancement of the website.

Important considerations

Content and quality

Data.gov.uk is a searchable data catalogue with links to data that is hosted by the individual UK Government departments and agencies, local councils, and local data hubs. This inevitably leads to some variation in data standards, but the search engine in the current Beta version is responsive and easy to use. Datasets are organised by category and easily filterable by format, publisher, and most recent. However, format and publisher are not searchable, as they were in previous versions.

There is a large breadth of data available, making it a natural first port of call for a wide range of queries. For example, complete [HM Land Registry stamp duty data](#) stretching back to 1995 was published in 2019.

The format and presentation of the data vary considerably, with some data files available as raw data in a machine-readable format such as CSV, whilst others are only available as pre-packaged data in a human-friendly format such as a PDF file, often with a pivot table widget.

The portal is designed primarily with analysts and developers in mind and does not have the levels of interactivity and intuitiveness non-specialist users might demand. Nevertheless, some data can be previewed, such as geospatial data, which is presented individually on maps.

Usage

There is a disproportionate user interest in a small number of data releases, including road safety data, whilst large numbers of datasets are rarely used. The

most popular API requests are government domain lists and local authority boundaries.

Blockers and challenges

The platform's greatest strength is also its greatest weakness: because of the breadth of datasets and publishers included, it is difficult to maintain a consistent approach. This has created a mixed enthusiasm for open data, not least because there is more likely to be support for data publication where there is a direct and visible benefit for those responsible for publication. This is more likely to occur where a dataset is consistent and standardised across multiple areas, improving the value and usefulness of the data for re-use. When datasets are inconsistent, or not backed by a standard, there is a cost in bringing datasets together.

The [initial response from data users was less than expected](#) in the early years of operation despite significant investment. The [National Audit Office's report in 2012, *Implementing Transparency*](#), shows that usage levels did not match early expectations.

Moreover, GDPR has had an impact on engagement and involvement. For instance, [the Higher Education Policy Institute observed that the DfE tightened procedures in response to GDPR](#) and has removed the assumption that anonymised data should be published.

User experience

Because the datasets originate from many different sources and links redirect to external sites, user experience is not fully consistent. The accessibility and searchability of individual data points largely depend on who has published them.

There also appears to be limited curation of datasets, with no identified work to assess the currency, validity or completeness of datasets. Many are not up to date or are no longer available, and there is no way currently to filter out inactive data sets.

What can Greater Manchester take from this?

- The publication of large amounts of data does not automatically generate levels of user traffic. Awareness raising and engagement with the data community to explore potential uses is at least as important as publication itself.
- Open data project leaders need to be aware that if they take a blanket approach to publication, many datasets may go unnoticed or neglected. This may be unavoidable since it is difficult to predict which datasets will be most

useful to users, but could be mitigated by better promotion around ‘data drops’.

- It is important to focus on quality and standardisation from the very beginning, since any problems are hard to rectify later once many data sets have been published. Early design mistakes can become ‘baked in’ to the structure of a datastore.
- There needs to be a way to select only active and regularly updated datasets to avoid wasting time sifting through empty entries. A curated approach would allow for review of dataset usage, timeliness and currency.
- There are limits to what is achievable for a portal that allows publication by many separate organisations, which each have different policies and approaches. Stronger central coordination and curation of data publication would help but may not be enough to completely overcome inconsistencies.

Find out more:

[Government Transformation Strategy](#)

[Guardian article on UK open government data – the results of the official audit](#)

[Government guidance: make better use of data](#)

[Public Technology article on government usage of data registers](#)