

Case Study:

European Union Open Data Portal

Type: Website

Organisation(s): European Commission

Tags: open data, process, standards, website

The [European Union Open Data Portal](#) (EUODP) provides a single point of access for European Union data. EU open data aims to serve the twin goals of boosting the EU's economic development and improving transparency in EU institutions.



The data hosted on the portal is provided exclusively by European Union institutions and agencies.

The European Union Open Data Portal should not be confused with the [European Data Portal](#), which is also run by the European Commission but has the wider remit of linking data produced across the continent by the Member States and affiliated countries. Unlike the EDP, the EUODP contains individual datasets in a datastore, rather than harvesting metadata.

Background

The [European Union Open Data Portal](#) (EUODP) was established in 2012 following the European Commission's Decision [2011 / 833 / EU](#) regarding the reuse of internal documents. Since then, the EUODP has grown to become a major datastore holding over 15,000 datasets on a wide range of topics from crime data, to election results and legal acts. All datasets are openly available and can be reused freely, as long as the user fully cites the EUODP as their data source.

The portal provides a catalogue of datasets, but also a library of applications that have been built using EUODP data and detailed information for the European Union's digital and data strategy.

As is also the case for the European Data Portal, a small team at the [Publications Office of the European Union](#) in Luxembourg is responsible for maintaining and updating the site.

Technical design and data access

The portal is built using open source solutions such as the [Drupal](#) content management system and [CKAN](#).

Metadata can be provided either manually through the data provider's section of the website or automatically in RDF format using the [Virtuoso RDF database](#). Metadata is then stored in two locations: in a human-readable database and a machine-readable triplestore database.

Data can be accessed using the [REST API](#), as well as the legacy [CKAN API](#), and users can also use [SPARQL EndPoint](#), which allows queries on the RDF descriptions of datasets.

Users are able to suggest new datasets that should be included using an [online form](#).

EU Data Strategy and data sharing

In 2019 the European Commission made the digital agenda one of the key pillars of its five-year strategy. Data is a key part of this strategy, and the European Union aims to create a [single market for data](#) to boost both competitiveness and data sovereignty. The expected benefits of this strategy include improvement in healthcare, creating better transport systems, increasing energy efficiency, and reducing the cost of public services. This strategy includes a wide range of policies that include, but are not limited to, opening up data. These include:

- Investing €2 billion in a [High Impact Project](#) to develop data processing infrastructures, data-sharing tools and governance mechanisms for data sharing mechanisms, facilitating a single European data market
- Investing in digital skills for SMEs
- Creating common European data spaces in strategic sectors like the green deal and industrial manufacturing
- Facilitating a procurement marketplace for data processing services
- Making more high-value publicly held datasets available.

B2G data sharing strategy

As part of this, the European Commission set up a High-level Expert Group on B2G (business to government) data sharing. The group's [report issued a series of best practice studies](#) and pledges to help guide the use of privately-held data to achieve public policy goals. Pledges include the creation of a [European data stewards' network](#) to act as champions for B2G collaboration, Another is the creation of [European regional data space](#) for collaboration between universities, governments, private companies, and NGOs, with support from the [Data-Pop Alliance](#). This

network will be used to support data literacy and provide strategic support to leverage private-sector data and AI.

An example of good practice given here is [Finnish forestry data](#). The Finnish Forestry Commission is setting up a data platform to share anonymised cloud-based data such as canopy height data, laser-scanned data, and aerial photography. This aims to help improve the cost-efficiency of timber management in the forestry sector, which is a strategically important industry.

Important considerations

Content and quality

A total of 15,399 datasets are available on the portal. Datasets can be accessed in several ways. Users can use the free-text search function, search through the thirteen categories, or choose one of the 'most used' or 'recently updated' datasets.

Each dataset is supported by a wide selection of separate resources. There is generally a list of sub-files containing key segments of larger datasets, and a list of visualisations to help users gain a quick overview. There is also extensive documentation providing detailed information on how data were gathered and the form they are presented in. Datasets can often be downloaded in several formats, including as an RSS feed where applicable. There is also good quality metadata, showing the temporal and geographical coverage, key dates of updates, and contact details for the data holder.

There are several datasets on the portal which have been extensively used. The three most frequently viewed datasets are the COVID-19 dataset (327,828 views); a list of persons and entities subject to EU financial sanctions (178,296 views); and TED, a list of public procurement notices (106,965 views).

A total of over 70 agencies, institutions, and departments have published data on the portal. The largest publisher by far is the European Commission, with 12,966 datasets, of which the largest sub-publisher is [Eurostat](#), with 7,770 datasets. This is unsurprising given that this is the Directorate-General that specialises in producing data to support EU decision-making. In contrast, some publishers within the European Commission, such as the Directorate-General for Trade, have only published one dataset.

Web Applications

A large number of web applications have been created by third-party developers based on EUODP, and a [comprehensive list of these web applications](#) is shared on the website. One example is [InfringEye](#), which provides a database and visualisations of past and ongoing EU infringement procedures. The aim is to make it

easier for journalists and citizens to keep track of member state governments that have taken actions that do not comply with EU law.

Another is [OpenTED](#), which aggregates all public procurement opportunities in European Union member states. The app takes EU-data held on [Tender Applications Daily \(TED\)](#) data and presents them in a more user-friendly way, allowing businesses to identify procurement opportunities more easily.

Blockers and challenges

The EU data portal is easy to use and provides consistent quality across data publishers. It has been noted that there is still [more work to be done in the data standardisation process](#) to ensure metadata and file types are equivalent across participating organisations. However, this is inevitably an ongoing process.

One potential blocker however is the fact that this data portal is limited to European Union agencies and institutions only. In the context of multi-layered governance, with close interaction between national, European, and regional levels, distinctions between datasets relating to these different layers are becoming increasingly arbitrary. The European Data Portal provides these links but has less consistent interoperability than the EUODP. The contrast between these two portals exemplifies the difficult balance that must be struck between the quality and quantity of data in international data programmes.

What can Greater Manchester take from this?

- The EU's digital strategy highlights the fact that open data is only ever one half of the puzzle. It is best to see open data as part of a wider strategy that supports digital access, skills, and the innovative application of data.
- Where third party developers have found innovative uses of a dataset, the portal that hosts the underlying data should promote these applications and list them in one place to ensure data use is always prioritised.
- Datasets about public procurement projects, as in the case of infrastructure and property development, is one area where timely open data releases could have significant commercial benefits.
- In the case of key datasets, it can be helpful to publish detailed and sophisticated contextual information about the background of the dataset and what it relates to. Whilst doing so is costly, a standard metadata table with key dates and relevant organisations is not always sufficient to provide the full picture.
- Where several levels of government are interacting, as is also the case in Greater Manchester, it is important to decide on a clear strategy on how to best link up datasets held by national, regional, local and international authorities.

- The case of Finnish forestry data indicates that the commercial value can be most easily captured from shared data where there is an industry-level coordination mechanism to ensure there are sufficient incentives to participate and reduced risks of freeloading.

Find out more:

[European Union Open Data Portal](#)

[European Data Portal](#)

[Good practices and pledges on B2G data sharing](#)