



An Roinn Caiteachais
Phoiblí agus Athchóirithe
Department of Public
Expenditure and Reform



DATA.GOV.IE

Open Data Publication Guidelines

Version: 1.0

April 2021

Open Data Unit,
Department of Public Expenditure and Reform

Supported by:



Author: Deirdre Lee, Derilinx

Contents

Contents.....	2
Glossary.....	3
Revision of the Irish Open Data Technical Framework.....	4
The EU Directive on Open Data and the Re-Use of Public Sector Information	4
Purpose of the Open Data Publication Guidelines	6
Quick Guide.....	7
Open Data Publishing Guidelines.....	9
1. Publish High-Value Datasets	9
2. Perform Data Audits	10
3. Put in Place an Open Data Publishing Roadmap.....	10
4. Associate Data with an Open Data Licence	11
5. Provide Comprehensive Metadata	12
6. Provide Accurate Timeframe Metadata	13
7. Ensure Data is Up-to-Date	14
8. Enable the Geolocation of Data	14
9. Provide Granular Data.....	15
10. Co-Locate Documentation for the Dataset.....	15
11. Publish Data in Open Formats	16
12. Publish Historical Datasets.....	17
13. Provide Access to Dynamic Data.....	18
14. Publish Data via APIs.....	19
15. Use Data Standards.....	20
16. Use Trusted Identifiers to Link Datasets.....	24
17. Ensure Dataset URLs are Operational.....	25
18. Identify Related Datasets.....	25
19. Support Non-Technical Users in Understanding the Data.....	26
20. Collaborate with User Community	26
21. Provide Examples of Data in Use	27
22. Measure Data Use.....	27

Glossary

API	Application Programming Interface
CC	Creative Commons
GDPR	General Data Protection Regulation
Information Provider	Public Sector Body responsible for publishing the data
Licensor	Public Sector Body responsible for licensing the data
ODAG	Open Data Advisory Group
PBWG	Public Bodies Working Group
PSB	Public Sector Body

Revision of the Irish Open Data Technical Framework

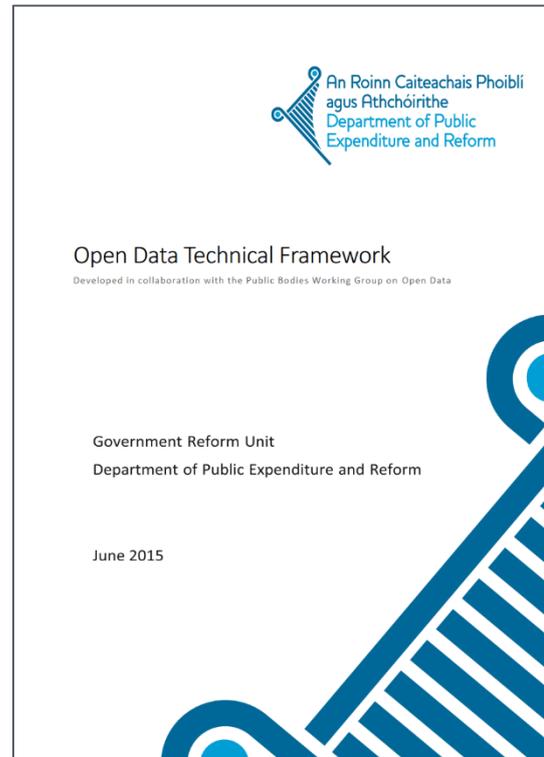
The Open Data Technical Framework was published in June 2015 and is available at <https://data.gov.ie/pages/opendatatechnicalframework>. The document was prepared by the Public Bodies Working Group (PBWG) on Open Data.

The Open Data Technical Framework set out the proposed technical elements that support the ongoing implementation of the Open Data Initiative. It ensured that publication of datasets on the Open Data Portal, <https://data.gov.ie>, was done in a consistent, persistent and truly open manner.

The Open Data Technical Framework comprised five key components:

1. Open Data Licence
2. Recommended Formats for Open Data
3. Recommended Metadata Schema for Open Data
4. Recommended Standards for Open Data
5. Unique Resource identifiers

Since 2015, there has been significant advancements in the area of Open Data and related domains, such as machine learning, artificial intelligence and the internet of things. New types of data are being generated and collected, and there is continuous evolution in technologies for analysis, exploitation and processing of data. In order to address these changes, whilst continuing to support and provide practical guidelines to data publishers, a revision of the Irish Open Data Technical Framework is required. In version 2 of this document, the Open Data Practical Publishing Guidelines, a clear outline of the technical components recommended for publishing Open Data will be presented, building on the solid foundation that the Technical Framework provided. In this way, the Guidelines will continue to ensure discoverability, accessibility and re-use.



The EU Directive on Open Data and the Re-Use of Public Sector Information

Notably, on 20 June 2019 the Directive (EU) 2019/1024 of the European Parliament and of the Council on Open Data and the re-use of public sector information entered into force. This Directive replaced the Public Sector Information Directive, also known as the 'PSI Directive' (Directive 2003/98/EC) which dated from 2003 and was subsequently amended by the Directive 2013/37/EU¹.

The 2019 EU Open Data Directive was published in order to address the remaining and emerging barriers to a wide re-use of public sector and publicly funded information across the Union, to bring

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

the legislative framework up to date with the advances in digital technologies and to further stimulate digital innovation, especially with regard to artificial intelligence.

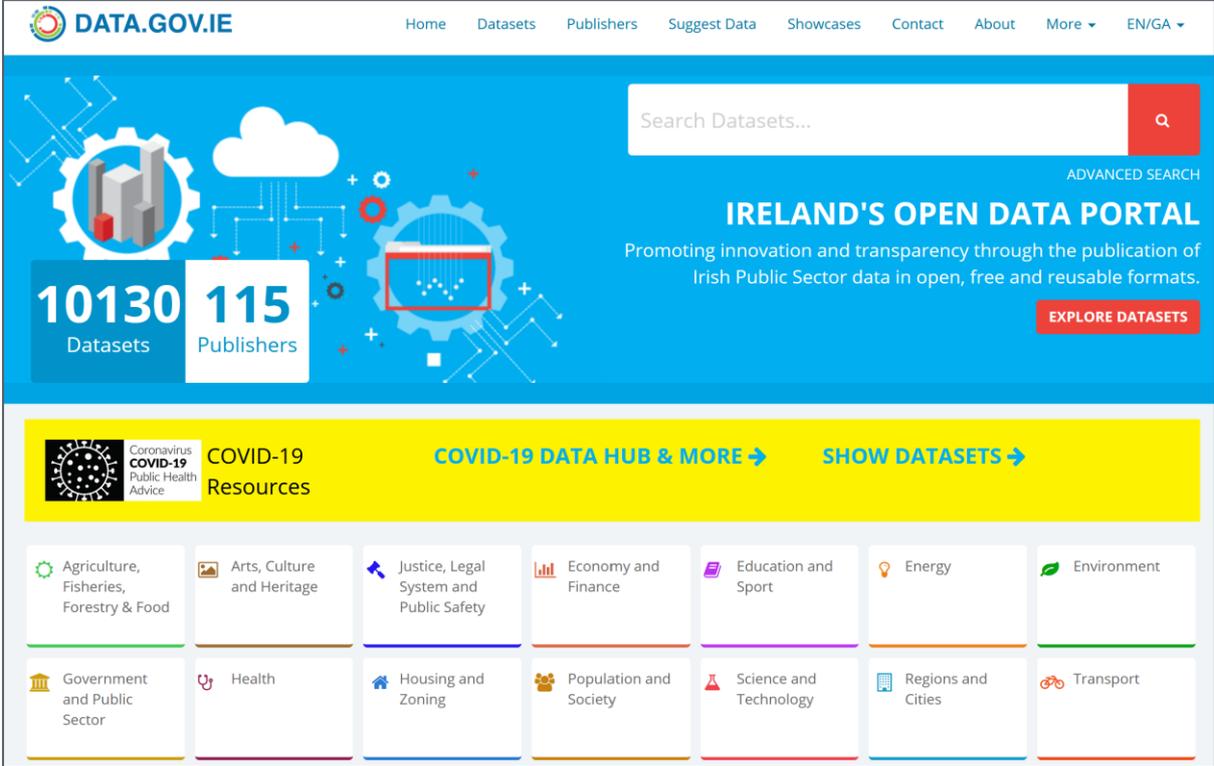
The Directive recommends that the legislation should focus on the following areas, so as to fully exploit the potential of public sector information for the European economy and society:

- the provision of real-time access to dynamic data via adequate technical means,
- the increase of the supply of valuable public data for re-use, including from public undertakings, research performing organisations and research funding organisations,
- the tackling of the emergence of new forms of exclusive arrangements,
- the use of exceptions to the principle of charging the marginal cost, and
- the relationship between this Directive and certain related legal instruments.

The Open Data Publication Guidelines should align with the EU Open Data Directive.

Purpose of the Open Data Publication Guidelines

Since the inception of the Irish Open Data Initiative and the launch of the Open Data Catalogue data.gov.ie, more and more datasets are being published, which has facilitated increased access to the information for users.



While providing data in its raw form supports data reuse for a variety of applications, there are a number of ways data publication could be refined and improved to increase the understanding, potential use, and overall impact of the information. This is why, in these Open Data Publication Guidelines, we focus on recommending best practices that will:

- ensure the discoverability and accessibility of data,
- empower people to understand and use data,
- support increased public-service transparency,
- drive evidence-based decision making and innovation, and
- enable measurable impact.

An Open Data Catalogue is a method of publishing and sharing data for those that need it, while offering a means to discover a wider range of datasets. Therefore, the data published on Open Data Portals should primarily meet the needs of the people who want to discover, understand and access information. Following this user-centric approach, each of the guidelines is tied to delivering a specific benefit(s) to the users of Open Data.

Quick Guide

The following is a summary of how the Open Data Publication Guidelines can be put into practice, highlighting some of the common challenges that potential publishers face.

To get started, publishers will want to identify what datasets to publish as Open Data.

Question: What datasets should we publish first?

- We will publish all datasets immediately.*
- Let's start with a couple of key datasets.*
- No datasets, our data isn't interesting and is of poor quality.*



Answer: (b) Start with some high-value datasets that are easy to publish. Then put in place a roadmap for publication of all suitable datasets.

To help understand what a high-value dataset is, check-out Guideline 1: Publish High-Value Datasets. For guidance on creating a roadmap for your organisation, see Guideline 3: Put in Place an Open Data Publishing Roadmap.

Challenge: But I don't know what datasets we manage, never mind the high-value datasets.



This is a common challenge, as datasets can be created and managed in any of the units across an organisation. One person will seldom have a complete view on all datasets. Also, while technical teams may process the data, it is often business teams who use and understand the data. Therefore, in order to get a better understanding of your organisation's data holdings, data audits can be carried out. This can begin be a simple list of what datasets a unit manages, or a detailed description of all datasets in every unit. To find out more about how to carry out a data audit, check-out Guideline 2: Perform Data Audits.

Question: In what format should I publish my dataset on data.gov.ie?

- As a PDF report, the data is in there somewhere.*
- As a CSV document.*
- I hear APIs are popular...*



Answer: (b/c) To make data as easy as possible to reuse, all data should be published in open, non-proprietary formats, such as CSV, JSON or XML. APIs are a great option if you want to provide programmatic access to data, in particular dynamic datasets or large datasets.

For more information on data formats, go to Guideline 11: Publish Data in Open Formats. If you would like to learn more about APIs, providing access to dynamic data, and publishing historical data, have a look at Guideline 12: Publish Historical Datasets, Guideline 13: Provide Access to Dynamic Data, and Guideline 14: Publish Data via APIs. Associating data with an open licence is also essential to clearly let users know what they are allowed to do with the data. For more information on licencing, go to Guideline 4: Associate Data with an Open Data Licence.

Challenge: People won't understand our data, it's too complicated. They might misinterpret it.



Inevitably, domain experts working with data will be best placed to understand this data. However this doesn't mean that this data can't have value outside of this use-case. Open Data can have numerous applications outside of those that were foreseen when it was initially collected. To help potential users fully understand the data, it is important to provide as much contextual and descriptive information about the data as possible. To find out more, see [Guidelines 5: Provide Comprehensive Metadata](#), [Guideline 6: Provide Accurate Timeframe Metadata](#), [Guideline 8: Enable the Geolocation of Data](#), [Guideline 9: Provide Granular Data](#), and [Guideline 10: Co-Locate Documentation for the Dataset](#).

Question: How will people use my dataset?

- On its own – one dataset = one application.*
- Integrated with a couple of other open datasets from my organisation.*
- Potentially integrated with any number of open and closed datasets from many sources.*



Answer: (c) The potential for Open Data use has no limits. Therefore it's important to ensure that all data published is as interoperable as possible.

Interoperability stems from using data standards, i.e. common ways of describing concepts in your dataset. For more guidance on improving the interoperability of your data, check out [Guideline 15: Use Data Standards](#), [Guideline 16: Use Trusted Identifiers to Link Datasets](#), and [Guideline 18: Identify Related Datasets](#)

Challenge: I already published some Open Data on data.gov.ie in 2017.



One of the most common issues with Open Data is that it is outdated. To build trust with users, datasets need to be kept current and all links should be working. To learn more about maintaining Open Data, read [Guideline 7: Ensure Data is Up-to-Date](#) and [Guideline 17: Ensure Dataset URLs are Operational](#).

Question: What's the point of going to all this effort?

- Publishing Open Data is like a black hole, I don't know if anyone is accessing it.*
- I've heard on the grapevine of a couple of people using it.*
- We have a good relationship with our community of users and have an understanding of their needs.*



Answer: (c) The best way of ensuring that publishing Open Data will achieve positive impact is by being demand-driven and working directly with users of your data.

There are many ways to help get an understanding of user needs, and also to help measure data use. If you know who your users are, you can work directly with them at face-to-face events, or online via social media or user forums. There are also a number of features on data.gov.ie that can help measure user needs and dataset access. To learn more about engaging with users and understanding demand, see [Guideline 19: Support Non-Technical Users in Understanding the Data](#), [Guideline 20: Collaborate with User Community](#), [Guideline 21: Provide Examples of Data in Use](#), and [Guideline 22: Measure Data Use](#).

Open Data Publication Guidelines

1. Publish High-Value Datasets

Much of the focus of Open Data initiatives in recent years, including the European Study on Open Data Maturity, is on the availability of High-Value Datasets (HVDs). While all Open Data publication is encouraged, Public Sector Bodies may have limited resources, and therefore need to prioritise what datasets should be made available first.

As outlined in Article 2 (10) of the EU Open Data Directive HVDs means documents, the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets. In Article 14 of the EU Open Data Directive, the identification of specific HVDs shall be based on the assessment of their potential to:

- a) generate significant socioeconomic or environmental benefits and innovative services;
- b) benefit a high number of users, in particular SMEs;
- c) assist in generating revenues; and
- d) be combined with other datasets.

Statement 66 of the Open Data Directive illustrates the thematic categories that could be covered by the Directive, namely:

- **Geospatial:** postcodes, national and local maps
- **Earth observation and environment:** energy consumption and satellite images
- **Meteorological:** in situ data from instruments and weather forecasts
- **Statistics:** demographic and economic indicators
- **Companies and company ownership:** business registers and registration identifiers
- **Mobility:** road signs and inland waterways

1.1. Recommendations for Publisher

- Identify HVDs within your organisation, using recommendations from the European ‘Impact Assessment Study on the List of High Value Datasets’.
- Review and decide on the low/medium/high intensity intervention measure for each HVD (as outlined in the Impact Assessment Study).
- Publish the HVD identified to the level of intensity intervention agreed.
- For the purpose of ensuring their maximum impact and to facilitate re-use, the HVDs should be made available for re-use with minimal legal restrictions and free of charge.
- the HVDs should also be published via APIs.

1.2. Benefits for Users

HVDs are of most interest to users, as they are the datasets that have the biggest potential for society, the environment and the economy.

2. Perform Data Audits

A data audit is a process to find out what data assets are being created and held within an organisation, and helps establish which datasets are suitable for sharing, both internally and publicly as Open Data. A data audit will result in the compilation of a comprehensive dataset inventory, and the definition of clear recommendations for managing this information and sharing it in a controlled environment.

A PSB can then use this dataset inventory to identify what datasets should be published as Open Data, which datasets should be prioritised, and how the datasets can be published according to these Publication Guidelines. This can be defined in an Open Data Publishing Roadmap (see Guideline 3).

For example, a PSB may hold twenty datasets, but five fall under the high-value dataset category (See Guideline 1). These datasets could then be prioritised. From the data audit, the publisher can also see if the data contains any personal/sensitive information that needs to be removed or anonymised. The publisher may also decide to create an API to provide dynamic access to the data.

Data audits should be carried out on a regular basis by all departments of a PSB to garner a recent snapshot of the organisation's data holdings.

While all datasets identified during a data audit may not be deemed as suitable for publication as Open Data, the data inventory itself could be made publicly available for transparency purposes.

2.1. Recommendations for Publisher

- Perform regular data audits to find out what data assets are being created and held within your organisation.
- Use the resulting data inventory to determine what datasets to publish as Open Data, and how to publish the data according to these Publication Guidelines.
- Publish the data inventory itself as Open Data.

2.2. Benefits for Users

Performing data audits enables PSBs to understand their complete data holdings, and in turn ensures that all data is considered for publication as Open Data to potential users.

3. Put in Place an Open Data Publishing Roadmap

An Open Data Publishing Roadmap maps out what datasets a PSB plans to publish against a specific timeline. In this way, a publisher can focus on publishing datasets it has prioritised in its data audit, while putting in place a delivery plan for when other datasets will be published. This gives PSBs time to prepare datasets that were not initially deemed ready to publish in the public domain, but were identified as being suitable for publication as Open Data.

3.1. Recommendations for Publisher

- As part of data audits, put in place an Open Data Publishing Roadmap.
- Regularly review and update the roadmap to ensure that dataset publishing goals are being met.
- Publish the Open Data Publishing Roadmap and progress updates.

3.2. Benefits for Users

Putting in place an Open Data Publishing Roadmap helps to embed Open Data publishing processes in an organisation. This in turn supports the ongoing publication of new datasets. If the roadmap itself is published, this allows users to understand what datasets will be published and when.

4. Associate Data with an Open Data Licence

All data published as Open Data should be associated with an Open Data Licence, so there is a clear understanding of under what conditions the data can be reused. The European Open Data Directive 2019 states that ‘any licences for the re-use of public sector information should, in any event, place as few restrictions on re-use as possible, for example limiting restrictions to an indication of source.’

The 2015 Open Data Technical Framework recommended the use of the use of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence². Under the CC BY 4.0 licence, a user is free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

CC BY 4.0 Licence Attribution Statement

Under the CC BY 4.0 Licence, users must acknowledge the source of the Information in their product or application. For data sourced on data.gov.ie, users can acknowledge the source by linking to the dataset URL on data.gov.ie., as well as including or linking to this attribution statement: “*Contains Irish Government Data licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence*”.

CC0 1.0 Universal (CC0 1.0)

Another licencing option for Open Data is CC0, which is even less restrictive for reusers. In contrast to CC’s licenses that allow copyright holders to choose from a range of permissions while retaining their copyright, CC0 empowers yet another choice altogether – the choice to opt out of copyright and database protection, and the exclusive rights automatically granted to creators – the “no rights reserved” alternative to our licenses³. By marking your work with a CC0 public domain dedication, you are giving up your copyright and allowing reusers to distribute, remix, adapt, and build upon the material in any medium or format, even for commercial purposes.

Multiple Attributions

If using data from several Information Providers and listing multiple attributions is not practical in a product or application, users may include a URI or hyperlink to a resource that contains the required attribution statements.

Disclaimer

All data linked to the Open Data portal is published “as is”. The information is licensed 'as is' and the Information Provider and/or Licensor excludes all representations, warranties, obligations and liabilities in relation to the Information to the maximum extent permitted by law.

² <https://creativecommons.org/licenses/by/4.0/>

³ <https://creativecommons.org/share-your-work/public-domain/cc0/>

The Information Provider and/or Licensor are not liable for any errors or omissions in the information and shall not be liable for any loss, injury or damage of any kind caused by its use. The Information Provider does not guarantee the continued supply of the Information.

Exemptions

This licence does not cover personal information, unless sufficiently anonymised and/or aggregated. Nor does it cover third party rights (including, but not limited to, patents, copyright, database rights or trademarks).

4.1. Recommendations for Publisher

- All data and metadata linked to data.gov.ie will be associated with the CC BY 4.0 Licence, at a minimum. Public bodies may waive copyright and associate datasets with CC0, if that is considered appropriate.
- The licence should be clearly identified in the metadata.
- If more specific usage guidelines are required, these should be clearly identified in the metadata, as part of the description or as a rights statement.

4.2. Benefits for Users

Users have a clear understanding as to how the data can be reused.

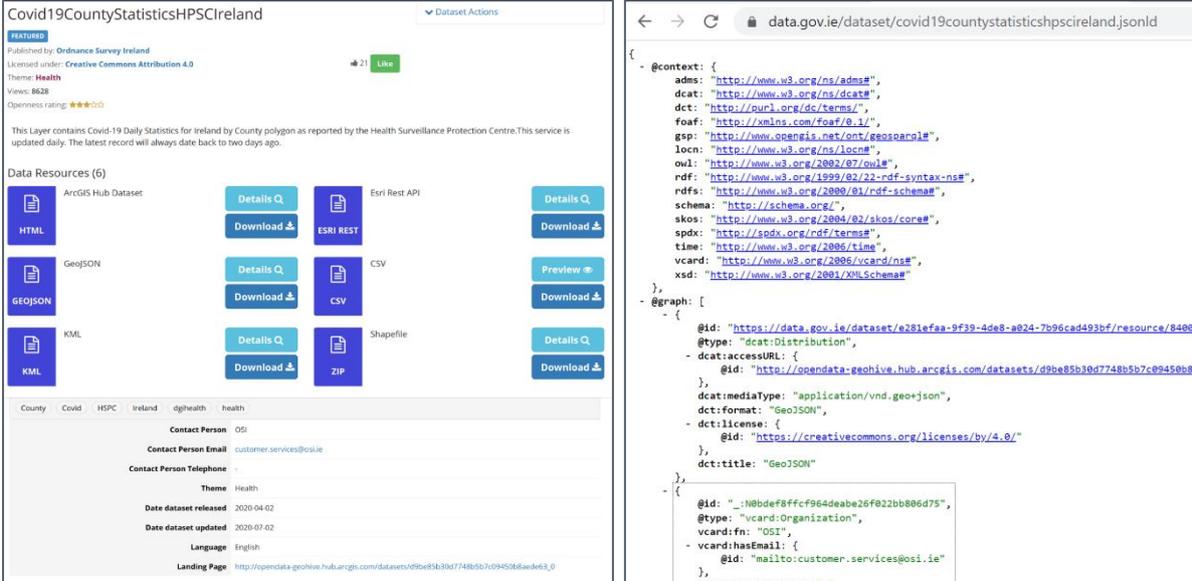
5. Provide Comprehensive Metadata

Metadata provides additional information that helps data consumers better understand the meaning of data, its structure, and to clarify other issues, such as rights and license terms, the organization that generated the data, data quality, data access methods and the update schedule of datasets⁴. Data.gov.ie is DCAT-AP compliant, meaning that all datasets on the catalogue include both a human-readable and a machine-readable DCAT metadata description. For example, the following screenshots are from the Covid-19 County Statistics dataset^{5,6}.

⁴ <https://www.w3.org/TR/dwbp/#metadata>

⁵ <https://data.gov.ie/dataset/covid19countystatisticshpscireland>

⁶ <https://data.gov.ie/dataset/covid19countystatisticshpscireland.jsonld>



However, providing more comprehensive and accurate the metadata improves the findability and usability of the data. The European Data Portal (EDP) analytical report on ‘The Future of Open Data’ highlights the need for metadata around three dimensions: *relevance* (is this the data I need?); *usability* (can I use it in practice?) and *quality* (how good is the data and how easy is it going to be do use it?)⁷. The ODI report on ‘Discovering the future of the London Datastore’ also highlights that poor and inconsistent metadata is one of the major issues highlighted throughout our findings⁸.

5.1. Recommendations for Publisher

- Provide DCAT-AP compliant metadata.
- Use meaningful naming conventions that a user not familiar with the dataset will understand.
- Provide a comprehensive description of the dataset.
- Provide a comprehensive list of keywords, including synonyms that a user may search for.
- Include information on the provenance of the dataset, i.e. the origins of the data and any changes that have been made⁹.
- Improve metadata based on a metadata quality assessment (MQA).

5.2. Benefits for Users

The provision of comprehensive metadata improves the **discoverability** of the data, helping users find the information that they are looking for.

6. Provide Accurate Timeframe Metadata

A number of timeframe elements are important for helping users understand what period of time a dataset covers and when the dataset was published and last modified.

⁷ https://www.europeandataportal.eu/sites/default/files/edp_analyticalreport_n8.pdf

⁸ <https://theodi.org/article/the-future-of-the-london-datastore-report/>

⁹ <https://www.w3.org/TR/dwbp/#DataProvenance>

6.1. Recommendations for Publisher

- Provide accurate metadata on what timeframe the dataset covers.
- Provide metadata on when the dataset was published / last modified.

6.2. Benefits for Users

The provision of accurate timeframe metadata allows users to perform time-based searches, while also helping users understand if this dataset is relevant or of use to them.

7. Ensure Data is Up-to-Date

One of the most common concerns with Open Data catalogues is outdated data⁸. The presentation of old data can damage users' trust in the catalogue, as they think it is not actively maintained. Up-to-date information is essential for users to perform effective decision-making and support innovations that are useful on a day-to-day basis, e.g. traffic applications.

What constitutes 'old' data is subjective and depends on how frequently the data should be collected and made available. For example, for annual datasets, the most recent year should be made available, whereas for real-time data, the most recent reading should be available to the nearest second/minute/hour. To ensure there is no human point-of-failure in keeping data updated, the publication of data should be automated using a programmatic publishing or harvesting process.

In order to help users understand how often a dataset is updated, information about update frequencies should be provided. (see Guideline 6). Also, related datasets should be linked, so that users can easily find a newer dataset if that is available (see Guideline 18).

7.1. Recommendations for Publisher

- Provide metadata on how frequently the data will be published.
- Ensure the data is kept up-to-date, in accordance with its update frequency.
- Automate the data publication/harvesting process.
- Identify related datasets in a time series.

7.2. Benefits for Users

Providing access to up-to-date data empowers effective decision-making and supports innovation. Users need to know how often the date will be updated before they can commit to its use.

8. Enable the Geolocation of Data

Geospatial data has been highlighted in the European Open Data Directive to be one of the thematic categories of High-Value Datasets. A lot of data can include geospatial information, such as GPS coordinates, Eircodes, addresses, etc., which can be used to overlay data from multiple sources, and also to filter data to a location that is relevant for the user.

8.1. Recommendations for Publisher

- Include standardised geospatial data if possible.

8.2. Benefits for Users

Enabling the geolocation of data allows a user to filter data by location, and to overlay information pertaining to a particular location.

9. Provide Granular Data

The more granular a dataset is, the more value it can deliver, as it allows for detailed analysis and more specific results. For example, if air quality is recorded at one-minute intervals, yet data published is on a daily basis, this ‘rounding up’ of the data can hide important information about what happens to air quality at different stages of the day.

A balance has to be struck between providing granular data and not breaching legal restrictions around personal, confidential or sensitive data. Open Data should never infringe upon a person’s right under data protection legislation, including, critically, the General Data Protection Regulation (GDPR) that came into force on the 25th May 2018.

Data anonymisation seeks to protect statistical data in such a way that they can be released without giving away confidential information that can be linked to specific individuals or entities. For example, statistical data published at a small geographic area, such as Electoral District (ED), may be rounded up so that individuals cannot be uniquely identified. The CSO supports PSBs in the publication of statistical data.

Another element of data granularity that should be observed by data publishers, is that data available at different levels of granularity should be accessible and modelled in a common way¹⁰. This allows for the merging or splitting of data.

9.1. Recommendations for Publisher

- Provide data in its most granular form.
- Ensure the provision of granular data does not breach legal restrictions, including infringing upon a person’s right under data protection legislation.
- Ensure data that is available at different levels of granularity is accessible and modelled in a common way.

9.2. Benefits for Users

Providing granular data allows users to carry out detailed analysis and provide fine-grained services and applications based on the data.

10. Co-Locate Documentation for the Dataset

As outlined in the EDP ‘Open Data Portal Assessment: Using User-Oriented Metrics’, the provision of documentation is fundamental, so that users do not need to be domain experts in order to understand the data¹¹. The documentation could be included within the metadata schema, for example, providing information on the provenance of the data or listing what standards the data uses. The documentation

¹⁰ <https://www.w3.org/TR/dwbp-ucr/#R-GranularityLevels>

¹¹ <https://www.europeandataportal.eu/en/highlights/sustainability-open-data-portals-infrastructures-reports-pt-1>

could also be provided as additional files or datasets, for example, the provision of a data dictionary that outlines what each component in the dataset represents, or the inclusion of a PDF document or website that details an API specification.

10.1. Recommendations for Publisher

- Provide as much documentation for the dataset that a user will require to fully understand and use the data, alongside the dataset.
- Where possible, include the documentation in the dataset's metadata.
- Provide reference data for each dataset, if applicable, e.g. data dictionary, standards used, etc
- Review dataset preview to see if satisfactory for a user.

10.2. Benefits for Users

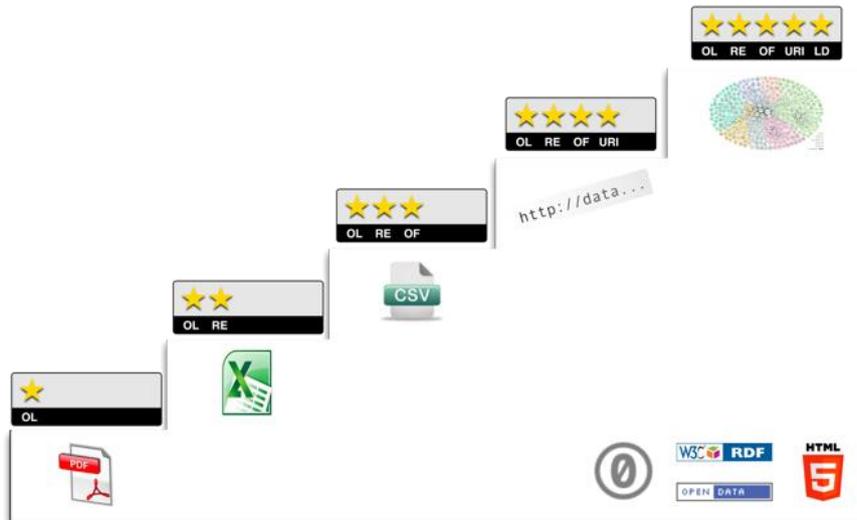
The co-location of documentation with a dataset helps users to understand the data and enable them to use it.

11. Publish Data in Open Formats

As outlined in the 2015 Open Data Technical Framework, data formats that can be readily used by a wide audience of people and by a variety of computer systems should be used for publishing Open Data. Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data¹². The greater the number of stars, the more reusable the data is, and the easier it is to reuse and interconnect data.

- ★ Publish data on the Web under an Open License
- ★★ Publish data in a machine-readable, structured format
- ★★★ Publish data in a non-proprietary format
- ★★★★ Use URIs to identify things, so that people can point at your stuff
- ★★★★★ Link your data to other data to provide context

¹² <http://5stardata.info/>



5-Star Open Data Scheme¹²

11.1. Recommendations for Publisher

- Publish data in Open, machine-readable formats (minimum 3-star), namely:

General Data	Geospatial Data	Domain-Specific Data
<ul style="list-style-type: none"> • CSV • JSON • XML • ODF • RDF • TSV 	<ul style="list-style-type: none"> • GeoJSON • GML • KML • WKT • LAS • IFC • Shapefile • WMS 	<ul style="list-style-type: none"> • PX • JSON-stat • NetCDF • BUFR • Datex II • GTFS • HDF5 • GRIB

- If available, publish data in multiple formats, with at least one format being open and machine-readable.
- Where applicable, publish data as 5-star Linked Data.

11.2. Benefits for Users

The use of open, machine-readable formats increases the reusability of the data. Having data available in a variety of formats supports a wider variety of users that are accustomed to or require the data in particular formats.

12. Publish Historical Datasets

As well as publishing up-to-date datasets, there is great value in publishing historical datasets. Providing access to historical data can support long-term analysis and research and facilitates comparative studies. If the frequency of data recording is high or the time period that the dataset covers is vast, historical datasets can be large. Therefore, publishers need to put in place suitable hosting and access mechanisms to serve the data. Some options include:

- Data can be made available for bulk download, and compressed to minimise the size of the dataset.
- Cloud data solutions could be used to optimise access to big datasets, e.g. AWS
- Big Data technologies could also be used to store and provide access to the data, e.g. Hadoop or Spark
- APIs could be built to provide access to subsections of the historical data at one time, e.g. pertaining to a particular time period or location.

12.1. Recommendations for Publisher

- Publish historical datasets if they are available.
- Optimise storage and access to the dataset to reduce costs and improve access time.
- If appropriate, assign datasets as key/current/archive data.
- Assign and indicate a version number or date for each dataset¹³.

12.2. Benefits for Users

Publishing historical datasets ensures users have access to large, historical datasets for analysis and comparison purposes, which can feed directly into policy and decision making.

13. Provide Access to Dynamic Data

The European Open Data Directive 2019 defines 'dynamic data' as "*documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data*"¹. Examples of dynamic data include environmental, traffic, satellite, meteorological and sensor generated data.

The availability of dynamic data can have significant economic, political, environmental and social impact, as its application (for example in web apps, data science models or decision-making) can take into account immediate measurements. For example, a homeowner can put in place flood mitigation measures based on weather forecast, while a local authority can provide appropriate response actions based on flood-sensor measurements.

As outlined in the Directive, to maximise the impact of the information, it is important that dynamic data is made available immediately after collection, or in the case of a manual update immediately after the modification of the dataset. Where this is not possible due to technical or financial constraints, public sector bodies should make the documents available in a timeframe that allows their full economic potential to be exploited. Ideally, dynamic data should be made available via an application programming interface (API), as outlined in Guideline 14, so as to facilitate the development of internet, mobile and cloud applications based on such data.

13.1. Recommendations for Publisher

- Publish dynamic data as soon after collection as possible.
- Make dynamic data available via an API

¹³ <https://www.w3.org/TR/dwbp/#VersioningInfo>

13.2. *Benefits for Users*

Providing access to dynamic data can create significant impact as it allows users to make decisions based on existing conditions.

14. Publish Data via APIs

Application Programming Interfaces (APIs) facilitate machine-to-machine communication and the seamless exchange of data, in a flexible and customised manner. APIs make it easy for developers to create new products and services using Open Data. Therefore, it is important that where possible, data is accessible via well-designed APIs, not only as downloadable files.

As outlined in the EU Open Data Directive, the set-up and use of APIs need to be based on several principles, including availability, stability, maintenance over lifecycle, uniformity of use and standards, user-friendliness, as well as security. Building an API is easy, but designing an API that meets business objectives, satisfies user needs, and is long-lived can be difficult. Therefore, APIs should be developed using a standard approach, and in collaboration with potential users.

As recommended by the W3C Data on the Web Best Practices, Web Standards should be used as the foundation of APIs¹⁴. APIs that are built on Web standards leverage the strengths of the Web. For example, using HTTP verbs as methods and URIs that map directly to individual resources helps to avoid tight coupling between requests and responses, making for an API that is easy to maintain and can readily be understood and used by many developers. An example of this in use is to use RESTful-JSON APIs.

APIs should be supported by clear technical documentation that is complete and available online. Where possible, open APIs should be used. The OpenAPI Specification is a broadly adopted industry standard for describing modern APIs¹⁵. The OpenAPI Specification (OAS) defines a standard, programming language-agnostic interface description for REST APIs, which allows both humans and computers to discover and understand the capabilities of a service without requiring access to source code, additional documentation, or inspection of network traffic. When properly defined via OpenAPI, a consumer can understand and interact with the remote service with a minimal amount of implementation logic. Similar to what interface descriptions have done for lower-level programming, the OpenAPI Specification removes guesswork in calling a service. At the time of writing, the latest version of the OpenAPI Specification is version 3.0.3, published on 20 February 2020¹⁶.

14.1. *Recommendations for Publisher*

- Provide access to high-value datasets via APIs.
- Provide access to dynamic datasets via APIs.
- Use Web Standards as the foundation of APIs (RESTful).
- Co-locate API documentation or a link to documentation, which specifies how the API can be used.
- Ensure the API document conforms to the OpenAPI Specification.

¹⁴ <https://www.w3.org/TR/dwbp/#accessAPIs>

¹⁵ <https://www.openapis.org/>

¹⁶ <http://spec.openapis.org/oas/v3.0.3>

14.2. Benefits for Users

The provision of data via APIs enables programmatic access to the data, and in turn facilitates the development of internet, mobile and cloud applications based on such data.

15. Use Data Standards

The intrinsic value of Open Data is not only in using datasets in isolation, but in integrating data sourced from different systems, different organisations, and even cross-border. To ensure data interoperability, data should be published in open, machine-readable formats, using recognised data standards.

Data standards are the rules by which data is described and recorded. Standards make it easier to create, share, and integrate data by making sure that there is a clear understanding of how the data are represented and that the data you receive are in a form that you expected¹⁷. This is especially important when data is being used by a third-party, being integrated from different sources, or when data is being shared across public bodies.

Data standards can refer to:

- *Reference Data*: standardised listings or taxonomies of terms or content from a particular domain, or
- *Data Vocabularies*: standard data models that are used to structure data in understandable and reusable ways.

When publishing Open Data, international standards defined by reputable standards organisations, such as ISO, the European Commission, W3C, IETF, OGC and OASIS should be used if possible. If international standards are unavailable or unsuitable, use national standards. For specific topics such as geospatial, statistics, or health, use national standards as defined by the responsible organisation (OSI, CSO, HIQA, etc.).

The Open Data Advisory Group (ODAG) reviewed the commonly used data standards by Irish Public Bodies. These are defined in the table below. This is not an exhaustive list and is designed to be a go-to point for data publishers. The list will be updated with new standards as they are adopted in general practice.

¹⁷ <https://www.usgs.gov/products/data-and-tools/data-management/data-standards>

Table 1: Recommended Data Standards for Publishing Open Data

Short Title	Title	Domain	Standardisation Body	URL
AR-DRG	Australian Refined Diagnosis Related Group	Health	Australian Government	http://www.aihw.gov.au/hospitals-data/ar-drg-data-cubes/
ATC/DDD	The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses	Chemical	WHO	http://www.who.int/classifications/atcddd/en/
COICOP	Classification of Individual Consumption According to Purpose	Consumption	UN Statistics Division	http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5
	CSO Standard Classification of Industrial Activity	Economic	CSO	https://www.cso.ie/en/methods/classifications/industrialactivity/
	CSO Standard Country Classification	Geographic	CSO	https://www.cso.ie/en/methods/classifications/standardcountryclassification/
	CSO Standard Currency and Funds Classification	Economic	CSO	https://www.cso.ie/en/methods/classifications/standardcurrencyfundsclassification/
	CSO Standard Days of Month Classification	Temporal	CSO	https://www.cso.ie/en/methods/classifications/standarddaysofmonthclassification/
	CSO Standard Days of Week Classification	Temporal	CSO	https://www.cso.ie/en/methods/classifications/standarddaysofweekclassification/
	CSO Standard Dwelling Type Classification	Housing	CSO	https://www.cso.ie/en/methods/classifications/standarddwellingtypeclassification/
	CSO Standard Employment Status Classification	Economic	CSO	https://www.cso.ie/en/methods/classifications/standardemploymentstatusclassification/
	CSO Standard Fields of Education Classification	Education	CSO	https://www.cso.ie/en/methods/classifications/fieldsofeducationclassification/
	CSO Standard Levels of Education Classification	Education	CSO	https://www.cso.ie/en/methods/classifications/standardlevelsofeducationclassification/
	CSO Standard Marital Status Classification	Population & Society	CSO	https://www.cso.ie/en/methods/classifications/standardmaritalstatusclassification/
	CSO Standard Months Classification	Temporal	CSO	https://www.cso.ie/en/methods/classifications/standardmonthsclassification/
	CSO Standard Principal Economic Status Classification	Economic	CSO	https://www.cso.ie/en/methods/classifications/standardprincipaleconomicstatusclassification/
	CSO Standard Quarterly Time Periods Classification	Temporal	CSO	https://www.cso.ie/en/methods/classifications/standardquarterlytimeperiodsclassification/

Short Title	Title	Domain	Standardisation Body	URL
	CSO Standard Sex Classification	Population & Society	CSO	https://www.cso.ie/en/methods/classifications/standardsexclassification/
	CSO Standard Weeks of the Year Classification	Temporal	CSO	https://www.cso.ie/en/methods/classifications/standardweeksoftheyearclassification/
	CSO Standard Years Classification	Temporal	CSO	https://www.cso.ie/en/methods/classifications/standardyearsclassification/
DataCube	Data Cube Vocabulary	Statistical	W3C	http://www.w3.org/TR/vocab-data-cube/
DCAT	Data Catalog Vocabulary	Metadata	W3C	http://www.w3.org/TR/vocab-dcat/
DCMI	Dublin Core Metadata Initiative	Metadata	Dublin Core	http://dublincore.org/documents/dcmi-terms/
Disadvantage Index	Disadvantage index		ERC	?
EUCAN	Common Cancers	Cancer	WHO	http://eco.iarc.fr/eucan/Default.aspx
IANA	IANA Media Types	Media/File Types	Internet Assigned Numbers Authority	http://www.iana.org/assignments/media-types/media-types.xhtml
IATI	International Aid Transparency Initiative	Aid	IATI	http://iatistandard.org/
ICCS	Irish Crime Classification System	Crime	CSO	http://www.cso.ie/en/media/csoie/releasespublications/documents/crimejustice/current/crimeclassification.pdf
ICD	International Classification of Diseases	Health	WHO	http://www.who.int/classifications/icd/en/
ISO 19100	19100 Geographic Information standard series developed by the International Organization for Standardization (ISO)	Geographic	ISO/OGC	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=54904&published=on&includesc=true
INSPIRE	Infrastructure for Spatial Information in the European Community	Spatial / Environmental	EC	http://inspire.ec.europa.eu/
ISO 3166-2:IE	Country codes and subdivisions		ISO	http://www.iso.org/iso/iso_3166-2_newsletter_ii-3_2011-12-13.pdf
ISO 639	Language codes	Language	ISO	http://www.iso.org/iso/home/standards/language_codes.htm
ISO 8601	Date and time format	Temporal	ISO	http://www.iso.org/iso/home/standards/iso8601.htm
ISO 4217	Currency codes	Economic	ISO	http://www.iso.org/iso/home/standards/currency_codes.htm
MDC	Major Diagnostic category	Health	Utah Department of Health	http://health.utah.gov/oph/IBIShelp/codes/MDC.htm

Short Title	Title	Domain	Standardisation Body	URL
NACE Rev.2	NACE Rev.2	Economic	Eurostat	https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2
NUTS	Nomenclature of territorial units for statistics	Geographic	EC	http://ec.europa.eu/eurostat/web/nuts/overview
SDMX	Statistical Data and Metadata eXchange	Statistical	SDMX	http://sdmx.org/
Total poverty index	Total poverty index		ERC	?
XBRL	eXtensible Business Reporting Language	Business	XBRL	http://www.xbrl.org/
	The 26 geographic counties, except that Tipperary NR and Tipperary SR are distinguished		CSO	?
	The 26 geographical counties		CSO	?
	The 32 geographic counties of Ireland and Northern Ireland		CSO	?
	The 34 administrative counties, except that Tipperary NR and Tipperary SR are combined		CSO	?

15.1. *Recommendations for Publisher*

- Publish Open Data using the data standards, outlined in Table 1.
- Reuse existing data standards as opposed to creating new data standards. For example, use data standards defined by international bodies such as ISO, the European Commission, W3C, IETF, OGC and OASIS, or national organisations such as the CSO, HIQA and the OSI.
- Contribute to data standard groups, e.g. the Local Authority Standards Group.

15.2. *Benefits for Users*

Using data standards when publishing allows the data to be easily understood and ensures interoperability of data from various sources.

16. Use Trusted Identifiers to Link Datasets

Research by the European Data Portal has revealed that datasets are often used with each other, with the most popular combination being that of population statistics, environmental datasets and regions and cities data¹⁸. Therefore, one of the recommendations in the EDP's Future of Open Data Portal's Analytical Report is to create links between datasets⁷.

Successful exploitation of datasets should be affected by the ability within portals to link to core reference data, such as classifications of common entities (e.g., types of places, organisations, products, assets), open address data and open geospatial data (mapping), units of measurement, temporal information etc. This will allow the cross-referencing and analysis of multiple datasets that are currently siloed or not interoperable on a non-personal basis. Creating links between datasets can be achieved by using a native approach such as Linked Data, which relies on universal Web URIs and domain vocabularies expressed in formal languages.

As outlined in the Public Service Data Strategy 2019-2023¹⁹, Irish Public Service Bodies (PSBs) have adopted a **National Data Infrastructure**, which seeks to establish the consistent use of unique trusted identifiers in Public Service administrative data focusing on citizen, business and address identification. Trusted identifiers that have already been implemented in Ireland or are under active consideration include:

- Personal Public Service Number (PPSN): A unique individual identifier that is used by individuals to engage with and access Public Services in Ireland.
- Individual Health Identifier (IHI): Provides for the introduction of unique Individual Health Identifiers for individuals, healthcare professionals and healthcare organisations, which is only accessible in a health context.
- Eircode: The National Postcode System for Ireland, Eircode benefits businesses and PSBs who can use it to plan delivery logistics or services to communities and to help drive geospatial analysis for policy formation and assessment.
- Unique Business Identifier (UBI) Government started an initiative in 2017 to assess the introduction of a UBI across the Public Service to act as a single standard identifier for business.

¹⁸ https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf

¹⁹ <https://www.gov.ie/en/publication/1d6bc7-public-service-data-strategy-2019-2023/>

- Unique Geographic Identifiers (UGIs: Ordnance Survey Ireland maintains a common set of unique geographic identifiers (UGIs) for over 50 million geographic objects in Ireland (including land parcels, buildings, road segments, etc.), which is available for use by all PSBs

16.1. Recommendations for Publisher

- Create inherent links between data through the use of common identifiers.
- Where applicable, use trusted identifiers as recommended in the National Data Infrastructure

16.2. Benefits for Users

The use of trusted identifiers unlocks the potential of PSB data holdings as they enable the linking of data across these holdings

17. Ensure Dataset URLs are Operational

Data.gov.ie is an Open Data catalogue that lists all published datasets from over a hundred Public-Sector Bodies. The majority of data is hosted elsewhere, with data.gov.ie simply providing a URL (link) to where the data can be accessed. It is therefore imperative that the URLs provided are operational and that if a user clicks on a dataset, they are able to download the dataset or have access via an API directly.

17.1. Recommendations for Publisher

- Ensure all dataset URLs are operational. Update the URL on data.gov.ie if the source URL changes.
- Remove the dataset from data.gov.ie if it is no longer available at source.

17.2. Benefits for Users

Ensuring dataset URLs are operational allows users to access the information whenever they want, and fosters trust in the catalogue and in the publishers.

18. Identify Related Datasets

A key way to help users navigate through datasets available on a data catalogue is to provide links to related datasets. This may be to datasets that are part of a common collection or datasets that pertain to a similar topic, or that are frequently used together. This can save a user valuable time and effort when searching for data.

There are a number of properties in DCAT-AP v2²⁰ to support defining the relationship between datasets, including `dct:source`, `dct:hasVersion`, `dct:relation`, `adms:sample`, etc.

²⁰ <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/201-0>

18.1. Recommendations for Publisher

- Include links to related datasets using properties defined in DCAT-AP v2.

18.2. Benefits for Users

Identifying related datasets allows a user to easily find other datasets that might be of use to them.

19. Support Non-Technical Users in Understanding the Data

Having access to raw data is of huge value to users who wish to view, understand, analyse and reuse the information. However, being able to work with raw data requires a certain level of technical expertise. As a result, there can be a digital divide between those data-savvy users and non-technical users. In order to address this, it is good practice to provide specific support to non-technical users to help them understand the data. This can be achieved by providing easy-to-understand data previews or co-locating tools that can be used to view and interact with the data. In *Discovering the Future of the London Datastore*, it is recommended to provide curated guides oriented around common use cases to increase findability of data⁸. For example, a curated guide could be oriented towards ‘here is the data you need for this challenge and how to use it’.

Another way to support non-technical users in understanding data is to provide tools to help them view and browse the data. For example, a map explorer could be used to help a user browse geospatial data²¹, or a filter query tool could be used to help a user browse multi-dimensional, statistical data²².

19.1. Recommendations for Publisher

- Ensure data previews are available for all datasets.
- Co-locate tools with the data, so that a wider range of users can be engaged with.
- Provide how-to guides for users on how the data can be used.

19.2. Benefits for Users

Providing specific supports for non-technical users will improve the accessibility of the information and help a wide range of users to understand the data.

20. Collaborate with User Community

The Open Data catalogue data.gov.ie is seen as a location to discover, access and use data, not just a technical listing of datasets. Engagement with users at every stage of data sharing is essential to encourage data reuse, and ultimately to drive economic, social, environmental and political impact. As outlined in *Discovering the Future of the London Datastore*, the most effective approach is a problem- or challenge-based one: working with data users and people affected to identify a problem which data can help solve, and then increasing access to the data required to solve it⁸. Focusing on challenges rather than simply creating inventories of data is more likely to yield reuse.

²¹ <https://data.smartdublin.ie/mapexplorer>

²² <https://data.cso.ie/>

20.1. Recommendations for Publisher

- Provide contact details for each dataset published. The contact details should include a generic email address associated with a role, as opposed to an individual.
- Engage in discussions around data requests.
- Engage in comments about datasets.
- Listen to feedback on datasets, and if appropriate, act upon this feedback
- Promote Open Data on all of your communication channels, including social media, newsletters, events, etc.
- Co-organise problem/challenge-based Open Data events with users and other publishers.

20.2. Benefits for Users

Collaborating with the user community facilitates the organic growth of ecosystems around data and paves the way for impact at scale.

21. Provide Examples of Data in Use

The EDP analytical report on ‘The Future of Open Data’ recommends promoting the use of Open Data to support sustainability and bring added value to portals⁷. Providing impact stories and examples of use demonstrate the potential of the data can facilitate wider access to the data, and can encourage further use of the data. Examples can be provided by the publisher directly, or contributed by other users with the support of the publisher.

However, providing detailed data insights or visualisations should not detract from the key focus of the catalogue, which is to improve the access, use and sharing of data⁸.

21.1. Recommendations for Publisher

- Encourage users / community to provide showcases and examples of how data is being used.

21.2. Benefits for Users

Providing examples of data in use can demonstrate the potential of the data, can facilitate wider access to the data, and can encourage further use of the data

22. Measure Data Use

There are many motivations for measuring Open Data. It is needed to maintain quality of data and support; to justify investment; to focus resources to most effect; to compare progress between countries, institutions and portals and to set benchmarks for countries, institutions and portals²³. If the aim of publishing Open Data is to help achieve positive environmental, economic, political and social impact, it is important that there is a clear way to measure how data is being used and what impact it is actually having. Measuring data use will help publishers identify if the goals set out in their

²³ https://www.europeandataportal.eu/sites/default/files/sustainability-data-portal-infrastructure_0_summary.pdf

organisation's Open Data Publishing Roadmap are being achieved. Ongoing review of data use can help publishers understand challenges users are facing, and potentially improve the data so that it is easier to use.

Defining metrics to measure data use can be difficult, especially if what is being measured can be subjective, for example, data quality. However, there are a number of indicators that can be automated and consistently reviewed to measure data access, such as dataset views and downloads on a data catalogue. For example, data.gov.ie uses Google Analytics to record dataset page views and downloads. If a publisher publishes data via an API, the number of API requests can be measured to better understand usage.

Domain-specific, quantitative metrics can be defined by the publisher, e.g. a Local Authority may use the 'number of inquiries as to where recycling centres are' as an indicator to measure if a 'recycling centre locations' Open Dataset is being used.

Qualitative methods and studies can also be used to garner a better understanding of how data is being used and the impact it is having. For example, organising surveys or focus groups with your user community. See Guideline 20 for more information on collaborating with the user community.

22.1. Recommendations for Publisher

- Review and compare your organisation's data access metrics on data.gov.ie on an ongoing basis.
- If data is published via APIs, use API analytics to measure API requests. Review and compare this data access information on an ongoing basis.
- Define domain-specific quantitative metrics to measure use of your organisation's particular datasets.
- Define qualitative methods to understand how your organisation's data is being used and the impact it is having.

22.2. Benefits for Users

Measuring data use helps publishers understand if and how their data is meeting user requirements, and should in turn lead to improved data publication and quality of Open Data.