

2015 EDITION

[illegible]

Manuscript completed in January 2015.

More information on the European Union is available on the Internet (<http://europa.eu>).

Luxembourg: Publications Office of the European Union, 2015

ISBN 978-92-78-41247-0

doi:10.2830/56485

© European Union, 2015

Reproduction is authorised provided the source is acknowledged.

Printed in Luxembourg

PRINTED ON ELEMENTAL CHLORINE-FREE BLEACHED PAPER (ECF)

EU open data

The basics for EU data providers

2015 EDITION

Contents

- Introduction..... 5**
- 1. Open data, the EU and the Open Data Portal 7**
 - 1.1. What is public open data?..... 7
 - 1.2. What makes data open? 8
 - 1.3. Why is open data important? 9
 - 1.4. The EU open data policy.....10
 - 1.5. Legal framework 11
 - 1.6. The G8 Open Data Charter and the EU implementation plan.....15
 - 1.7. Open data in the world 15
 - 1.8. The EU Open Data Portal..... 17
 - 1.8.1. What specifically is the role and the contribution of the portal? 17
 - 1.8.2. How to use the EU Open Data Portal and how it works 18
 - 1.9. Who are the users of the portal?20
 - 1.10. What are the benefits for your organisation?.....20
 - Recurrent excuses for data hugging* 21
- 2. Opening up your data via the EU Open Data Portal..... 23**
 - 2.1 Setting up the organisational structure.....23
 - 2.2 Setting up the process23
 - 1** Identifying data24
 - 1. What kinds of data are we looking for?25
 - 2. What kinds of data are interesting?25
 - 3. What data is available in your organisation?.....26
 - 4. Where is the data?26
 - 2** Analysing the identified data.....27
 - 1. Who owns the rights?27
 - 2. Does it contain any personal data?27
 - 3. What is the quality of the data?27
 - 4. In what format is data available?28
 - Tim Berners-Lee’s five stars scheme*30
 - Tim Davies’ five stars scheme*..... 31
 - 3** Creating metadata.....33
 - 1. What is metadata?33
 - 2. What is the difference between metadata, data and datasets?33
 - 3. Why is metadata important?33

4.	Why standardise metadata?	33
5.	Why is the quality of metadata important?	34
6.	Metadata on the EU Open Data Portal	34
7.	Legal notice.....	35
4	Publishing datasets and metadata.....	36
1.	Where to publish	36
2.	Publishing datasets on the Web	36
3.	Publishing metadata on the EU Open Data Portal.....	36
5	Curating.....	38
1.	Updating.....	38
2.	Deleting.....	39
6	Entering new datasets.....	39

ANNEXES

ANNEX 1 — Glossary	41
ANNEX 2 — Good practices in data management	46
References.....	53

Introduction

Objectives

By providing general information and guidelines on how to open up data, this document aims at supporting EU organisations (the data providers) in building up a data publishing process via the EU Open Data Portal. It thus contributes to the EU open data policy, the goal of which is to foster economic growth, job creation, transparency and administrative efficiency.

The document answers frequently asked questions and will undergo regular updates on the basis of data providers' feedback. It complements the 'Data provider user manual' ⁽¹⁾, which provides technical information on metadata and guidelines on how to publish metadata on the EU Open Data Portal.

Responsible services

The Publications Office is responsible for the operational management of the EU Open Data Portal. The Office also delivers technical and legal help through training courses, webinars, ad hoc meetings, video conferences, etc.

The implementation of the EU open data policy, a key pillar of the EU digital agenda, is the responsibility of DG Communications Networks, Content and Technology.

Structure of the document

This document comprises four parts.

- **Part I** explains the **basic concepts** of open data as well as its legal aspects. It also presents the EU open data policy and introduces the role of the EU Open Data Portal as one of its key elements.
- **Part II** gives **practical guidelines** on how to open data and on building up a regular data publishing chain from planning to dissemination.
- The annexes comprises:
 - a **glossary** where you can find an explanation of terms that come up regularly in the domain of data publishing, and a
 - section on **good practices**, collecting data providers' shared contributions. This part will keep growing through future contributions that are volunteered.

⁽¹⁾ The 'Data provider user manual' is available at the PubliCare website (<https://webgate.ec.europa.eu/multisite/publicare/en/page/open-data-355>).



Target audience

This document is addressed to anyone in the EU institutions, agencies and other bodies who has to implement open data policies or simply wishes to know more about how to put them into practice.

Is there an issue you would like to be covered in this document?

Please let us know!

Contact: op-odp-contact@publications.europa.eu

*This brochure can be downloaded and ordered from EU Bookshop:
<https://bookshop.europa.eu>*



1. Open data, the EU and the EU Open Data Portal

1.1. What is public open data?

Open data refers to the practice of publicly publishing raw data in a way that is accessible, reusable, machine readable and licensed permissively. It can be generated by a wide range of parties, from public authorities, the semi-public sector, businesses and the public. In the case of public authorities, such as European Union (EU) organisations, making their data available for public reuse supports economic development, openness and transparency.

As a matter of fact, public authorities produce, collect, centralise or pay for huge quantities of data, referred to as public data, public sector information or government data. Examples include national statistics, budget information, parliamentary records, geographical data, laws and data about various policies.

Most such data is publicly available, gathered at taxpayers' expense and for their ultimate benefit. Therefore it should be made freely available for both **access** and **reuse**, i.e. to use it for something different than its initial purpose (for example by combining it with data from other sources).

Generally, the **opening up of public data** helps foster governments' transparency and accountability, and it also has a strong leverage on the economy.

A transparent government provides open data on how it performs its duties, how taxpayers' money is spent, what that money is spent on and to what effect.

Interestingly, though originally based on the principles of transparency, participation and collaboration with citizens, the use of government data has evolved throughout time. In our digital era, data plays a central role in the economy, in particular thanks to recent technologies which make real-time data accessible to anyone and automatically processable by computers.

A recent study indicates that the overall economic gains from further opening up public sector information by allowing easy access could be up to EUR 40 billion a year for the EU-27 ⁽²⁾. According to an-



***What is the relationship between access and reuse?**
Access is all about obtaining the information. Reuse builds on existing access rules, but involves making use of the public information for purposes other than the original one for which it was created or collected.*

***Opening up of public data**
The implementation of open data shall be done in full respect of privacy legislation. Likewise, data protected by intellectual property rights of third parties is excluded from reuse.*

⁽²⁾ 'Review of recent studies on PSI reuse and related market developments', G. Vickery, August 2011 (http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?action=display&doc_id=1093).

other study, seven fields of the global economy alone could generate USD 3 trillion a year in additional value as a result of open data ⁽³⁾.

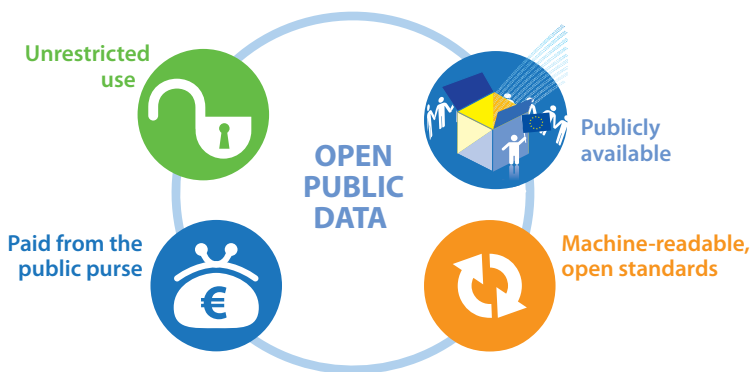
Areas such as geographical data, meteorological data and legal and business information form the basis of a steadily growing market. In the Netherlands, for example, the geo-sector accounted for 15 000 full-time employees in 2008 ⁽⁴⁾.

1.2. What makes data open?

According to <http://opendefinition.org>, 'A piece of data or content is open if anyone is free to use, reuse and redistribute it — subject only, at most, to the requirement to attribute and/or share alike.'

In short, this means that data should be easily accessible, with low entry barriers, widely available and in shareable formats. More specifically, it should comply with the following criteria.

- **Availability and access:** the data must be available as a whole and be downloadable over the Internet in a convenient and modifiable form.
- **Reuse and redistribution:** the data must be provided under terms that permit reuse and redistribution, including intermixing with other datasets.
- **Universal participation:** everyone must be able to reuse it, with no discrimination against persons or groups. For example, non-commercial restrictions that would prevent commercial use, compulsory registration or restrictions of use for certain purposes, are not in line with open data principles.



⁽³⁾ 'Open data: unlocking innovation and performance with liquid information', McKinsey & Company, 2013 (http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information). The seven fields are education, transportation, consumer products, electricity, oil and gas, healthcare and consumer finance.

⁽⁴⁾ Geomatics magazine, GIM International, 2011 (http://www.gim-international.com/news/id6308-Europe_Opens_Access_to_Public_Data.html).



1.3. Why is open data important?

Opening up government data for reuse can have major benefits for citizens, businesses and society, and for the governments themselves.

- **Allow the creation of new businesses** and new innovative added-value services, enabling the mash-up of existing data from different sources in unforeseen ways. Facilitating the reuse of this raw data can create jobs and thus stimulate growth.
- **Promote transparency:** open data is a powerful instrument to increase transparency and accountability by providing the public with previously inaccessible information about what the EU or a public body is doing.
- **Accelerate scientific progress:** access to raw data can accelerate scientific progress to meet the challenges of the 21st century in scientific discovery and learning.
- **Addressing societal challenges:** data can play a key role in our society. For example enhancing sustainability of healthcare systems or tackling environmental challenges.
- **Improve the quality of decision-making and civic participation:** the availability of solid public data provides the means for better evidence-based policymaking. This will help deliver better public services and administrative efficiency. It also improves public relations and attitudes towards governments as it makes it easier for citizens to contribute to the process of governance.

The importance of open data was underlined by the **European Council of 24/25 October 2013**, where the digital economy was part of its agenda.

'Open data is an untapped resource with a huge potential for building stronger, more interconnected societies that better meet the needs of the citizens and allow innovation and prosperity to flourish. Interoperability and the reuse of public sector information shall be promoted actively.'

*Extract from the conclusions
of the European Council of 24/25 October 2013, p. 4.*

What kinds of new services can public sector information give rise to?

Opportunities for reuse have multiplied in recent years as technological developments have spurred advances in data production as well as data analysis, processing and exploitation. Data can be integrated into new products and services that we use on a daily basis, such as car navigation systems, weather forecasts or other useful apps on smartphones.

Some examples are as follows.

- **TheyWorkForYou.** Based on governmental data, this website helps to track parliament activity. Originally created in the United Kingdom, versions of this website have been adopted in various countries around the world.
► <http://www.theyworkforyou.com/>
- **Wheelmap** helps people in Berlin to search, find and mark wheelchair-accessible places. They can access this online map using a computer or their mobile phone.
► <http://wheelmap.org/>
- **Europe's Energy** combines data from Eurostat and other agencies. The app helps users to compare how progress is being made in Member States towards Europe's target to reduce energy consumption.
► <http://energy.publicdata.eu/ee/vis.html>
- **OpenCorporates** is a database of companies. It aims to have a URL for every company in the world. It has grown from three territories and a few million companies to over 30 territories and 30 million companies, and is working with the open data community to add more each week.
► <http://www.opencorporates.com>
- **Tender time frame** is a website application. By collecting and analysing tenders, one can compare how many working days are necessary to complete a tender across Europe, and the average time to complete a tender by category of services and products.
► <http://tt.spendnetwork.com/index.html>



1.4. The EU open data policy

The European Union has for many years placed a strong emphasis on open data as a resource for innovative products and services and as a means of addressing societal challenges and fostering transparency.

As a result, it has conducted a proactive policy based on the idea that public information is a valuable asset that should be freely available for use and reuse. By opening up public information, the open data policy of the EU aims at creating a single EU digital information market whereby new products and services are developed and citizens' participation in political and social life is fostered.

The European Commission has put in place a series of actions consisting of dedicated legislation and accompanying non-legislative measures. By providing favourable framework conditions for the use and reuse of data, these actions aim to overcome existing barriers to the creation of this EU digital information market.



In December 2011, as part of the digital agenda for Europe, the Commission presented an open data package consisting of three elements: a communication on open data, presenting its vision (COM(2011) 882); a proposal to revise the 2003 directive on reuse of public sector information; and a revision of the 2006 decision governing reuse of the Commission's own information.

The Juncker Commission, in office from 1 November 2014, has identified a connected digital single market as a key driver for economic growth and job creation. By appointing a vice-president for the digital single market he clearly gives it a central role and sets out its ambition for progress in this field.

1.5. Legal framework

Directive on the reuse of public sector information (the PSI directive)

Directive 2003/98/EC on the reuse of public sector information, known as the PSI directive, sets out the general legislative framework at European level for government data. It provides for a minimum degree of harmonisation.

Though the directive was adopted before the emergence of the open data movement, it triggered a shift in the culture inside public administrations towards greater openness. The PSI directive is a key pillar of the open data policy.

In June 2013 a revision of the PSI directive was adopted. The revised PSI directive (Directive 2013/37/EU) brings about important improvements, as it introduces the default rule that public information will be available for free or at very low cost and it expands the scope to other institutions (museums, archives and libraries).

The reuse of public sector data, whether for commercial or non-commercial purposes, should fully respect EU and national privacy legislation as well as the intellectual property rights of third parties.

The Inspire directive

Apart from the reuse directive, which applies to public sector information in general, there are also European laws which apply to specific domains.

For example, in the area of geographical information, the Inspire directive entered into force in May 2007 and applies to geographical information. It requires EU Member States to make geo-information available in 34 themes, provide descriptions of them in the form of metadata and enable their reuse by means of open standards.



The directive has established a European spatial information infrastructure to support an integrated approach to European environmental policy.

Decision on the reuse of Commission documents

The EU is itself a producer and holder of public sector information. Not only does it practise what it preaches. It also goes even further with the reuse policy put in place by the European Commission for its own information resources.

Since 2006, under Decision 2006/291/EC on the reuse of Commission information, the Commission has allowed reuse of its documents for commercial and non-commercial purposes at no charge.

In December 2011, under the revised Decision 2011/833/EU, amending the abovementioned decision, it also engaged to work to provide documents in machine-readable formats where possible and included the research data produced by the Joint Research Centre (JRC) under the reuse regime. In the accompanying communication other EU institutions are invited to adopt a similar reuse policy.

Moreover, it set up an open data portal to promote accessibility and the reuse of data from Commission services. This portal is also open to data from other institutions and bodies. In fact, 'the Commission will explore how similar rules could be taken up by other EU Institutions and key agencies' ⁽⁵⁾.

According to Article 13, 'the decision shall be reviewed 3 years after its entry into force'.

The decision is complementary to the Commission policy on access to Commission documents ⁽⁶⁾ (Regulation (EC) No 2001/1049).

Recommendation on access to and preservation of scientific information

In July 2012 the Commission issued a package of measures on scientific information that shows the EU's strong support for opening up scientific data to boost the benefits of public investment in research.

A recommendation to Member States ⁽⁷⁾ provides a policy framework for access to and preservation of scientific information.

The recommendation was complemented by a communication ⁽⁸⁾ on better access to scientific information. It sets out open access as a general principle for the scientific publications funded through Horizon

⁽⁵⁾ Commission communication — Open data — An engine for innovation, growth and transparent governance (COM(2011) 882).

⁽⁶⁾ http://ec.europa.eu/transparency/access_documents/index_en.htm

⁽⁷⁾ OJ L 194, 21.7.2012, p. 39.

⁽⁸⁾ Commission communication — Towards better access to scientific information: boosting the benefits of public investments in research (COM(2012) 401).

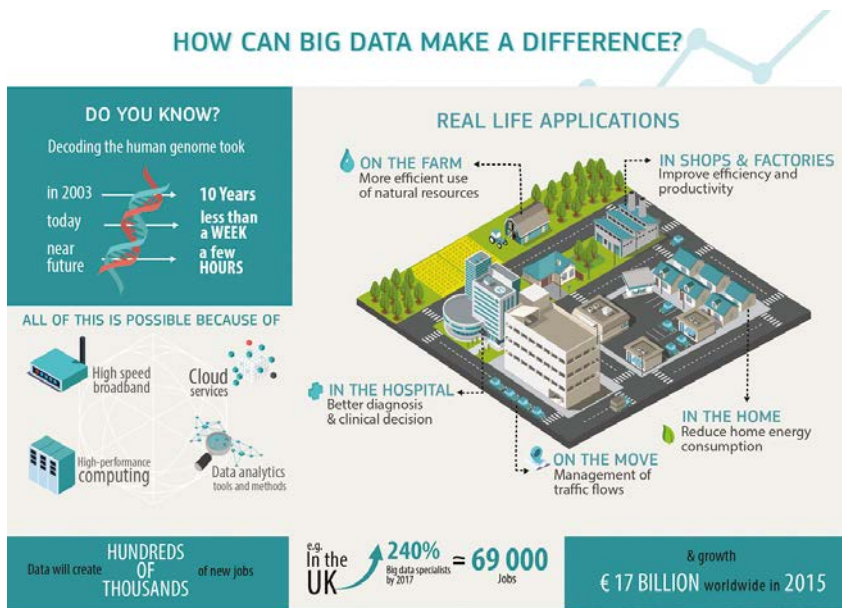


2020, the EU's research and innovation funding programme for 2014–20. The Commission has also set up a pilot scheme on open access to the data generated by projects in selected areas of Horizon 2020 (e.g. the numerical results of experiments).

Communication on data-driven economy

On 2 July 2014, the European Commission adopted a communication⁽⁹⁾ entitled 'Towards a thriving data-driven economy' as a response to the European Council's conclusions of October 2013 (see page 9). This communication outlines that the big data market is a great opportunity to create new jobs and growth. However, it also points out that Europe has been slow in embracing this revolution, compared to other competitors like the United States.

The communication describes the features of the data-driven economy of the future and sets out operational conclusions to support and accelerate the transition towards it. Indeed, according to the Commission, a public–private partnership (PPP) would be the most effective way to implement the Horizon 2020 research programme in the field of big data. It also supports small and medium-sized enterprises (SMEs) to embrace the new commercial opportunities for exploiting big data.



⁽⁹⁾ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Towards a thriving data-driven economy', (COM (2014) 442).

OPEN DATA CHARTER PRINCIPLES

PRINCIPLE 1

Open data by default: all government data will be published openly by default.

PRINCIPLE 2

Quality and quantity: data should be released as early as possible in its original form and fully described in clear language. This principle denotes the importance of metadata and user feedback to improve quality.

PRINCIPLE 3

Usable by all: data will be published in open formats for humans and machines wherever possible and will be free.

PRINCIPLE 4

Releasing data for improved governance: governments will share their technical expertise with each other and document their own open data initiatives.

PRINCIPLE 5

Releasing data for innovation: G8 governments will promote open data literacy and the provision of data in machine-readable formats.



G8 Summit at Lough Erne, United Kingdom, 17 June 2013.
© EU



1.6. The G8 Open Data Charter and the EU implementation plan

As a testament to the growing importance of the role of data, the G8 Summit released the Open Data Charter in June 2013. The charter recognised the central role open data can play in improving governance and in stimulating growth through innovation in data-driven products and services. It endorses the principle of 'open by default', which establishes that all government data should be published openly.

The charter sets out five principles to be acted on (see table opposite), with the nations committing to national plans for free government information flows. The charter also recognises a list of high-value datasets to be released as a priority.

Thanks to the many initiatives already adopted at EU level, compliance with the charter is fully consistent with existing EU policy. In its plan ⁽¹⁰⁾, published in autumn 2013, the EU is committed to:

- identifying and making available both core (budget, election and statistics) and high-value datasets held at EU level by 2015 (most of them have already been published);
- publishing data on the EU Open Data Portal;
- promoting the charter's principles in the 28 EU Member States (ensuring implementation of the PSI directive and preparing guidelines);
- supporting activities, outreach, consultation and engagement;
- sharing experiences on open data initiatives and best practices.

1.7. Open data in the world

The phenomenon of open up data is relatively new but it is gaining ground worldwide. Many public authorities from around the world have embraced the principles of open government and established open data initiatives. In the United States, the opening of government data accelerated after President Obama, in 2009, signed an executive order stating that all government information should be made public. The Obama administration also launched the US open government initiative to publish government data and created the website <http://www.data.gov> to distribute the data.

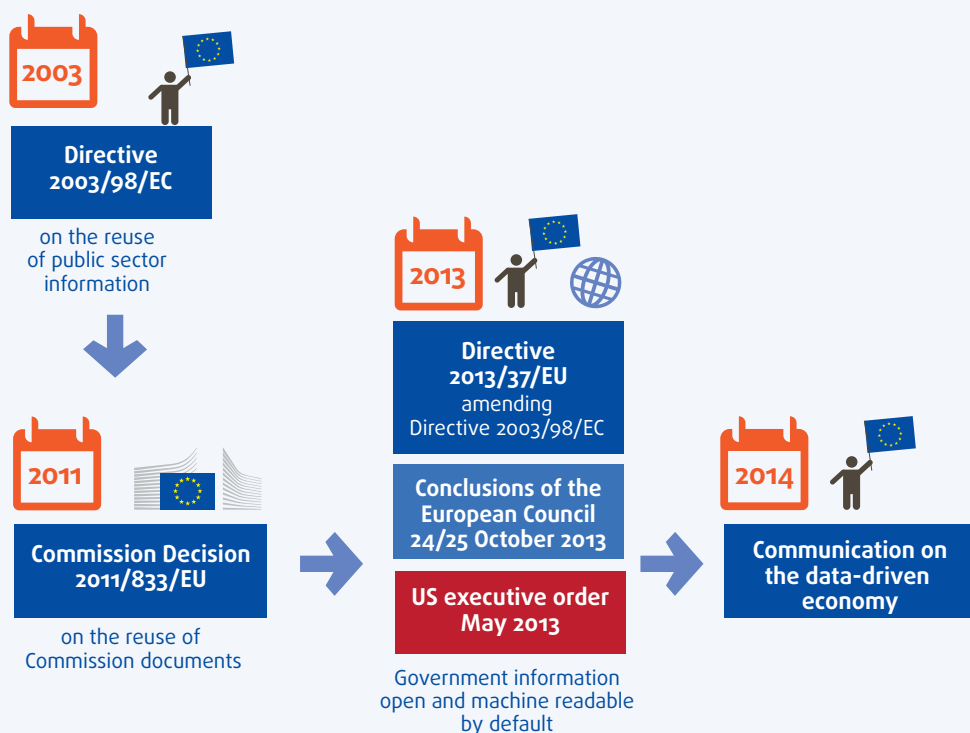
In the same year the New Zealand government launched its open data portal. Following this other projects mushroomed, and open data initiatives have been implemented widely in the world. Finally, in May 2013, as the latest of a series of actions in support of increasing access

⁽¹⁰⁾ The EU implementation plan is available at: <http://ec.europa.eu/digital-agenda/en/news/eu-implementation-g8-open-data-charter>

to and transparency of government information, President Obama's executive order ⁽¹⁾ laid out the general principles of open, machine-readable data as the 'new default'.

As of 2013 more than 40 countries had a government open data platform and more than 1 million datasets had been opened by governments around the globe ⁽¹²⁾.

And in those countries where national governments have not taken the lead by establishing clear open data policies, it is citizens, non-governmental organisations (NGOs) and urban government leaders driving the movement for more data.



⁽¹⁾ 'Executive order — Making open and machine readable the new default for government information', May 2013.

⁽¹²⁾ Source: Open data: Unlocking innovation and performance with liquid information, McKinsey & Company, 2013.



1.8. The EU Open Data Portal

(<http://open-data.europa.eu>)



To be used as open data, the data must not only be published but also findable. Data should be published proactively, with EU organisations not only providing information on request, but on their own initiative. Only in this way is data opened up for society as a whole.

Data portals have acquired a key role in providing a single point of access to governments' data. Their role is to help citizens and businesses, who sometimes find it difficult to know what kind of information exists and which authority holds this information.

At EU level, an open data portal was formally established by Decision 2011/833/EU, which in its Article 5 states:

'The Commission shall set up a data portal as a single point of access to its structured data so as to facilitate linking and reuse for commercial and non-commercial purposes.

Commission services will identify and progressively make available suitable data in their possession. The data portal may provide access to data of other Union institutions, bodies, offices and agencies at their request.'

The portal has now become the European Union Open Data Portal. Based on open-source software components, the portal was opened in its beta version in December 2012 and publicly launched in April 2013. Acting as the single access point to EU information, it provides access to data held by the Commission services and other EU institutions and bodies (the 'data providers').

The pan-European data portal

The pan-European data portal has been developed as part of the EU open data policy actions for Member States.

The pan-European data portal gives access to European open data assets in a harmonised way. As one of the instruments implementing the EU open data policy, its aim is to provide search and download capabilities for accessing freely reusable datasets from local, regional and national public bodies across Europe.

A pilot was released in June 2011 and is available online (<http://www.publicdata.eu>). A fully fledged portal, part of a comprehensive digital service infrastructure for open data deployed under the Connecting Europe Facility will be operational in the fourth quarter of 2015.

1.8.1. What specifically is the role and the contribution of the portal?

- **Provide a single point of access,** making it easier to find and discover EU data dispersed across the various websites of the EU organisations, the structure of which may not be known by citizens.
- **Facilitate and promote reuse:** in line with the EU open data strategy and reuse principles, anyone can easily search, explore, link, download and reuse the data for commercial or non-commercial purposes. This is made easier thanks to a metadata catalogue that is based on international standards, common encoding rules and a set of specific metadata agreed at EU interinstitutional level (within the Interinstitutional Metadata Maintenance Committee (IMMC)).
- **Facilitate interoperability and linking:** data linking is made easier thanks to the portal's functionalities enabled by semantic technologies. The portal includes a Sparql endpoint and an application programming interface (API), allowing automatic download from machines.



- **Help to overcome technical and legal barriers:** a challenge for reuse is that most data providers publish mostly or only in closed formats like PDFs. As it is very complicated to access the data behind it, the promotion of open formats is one of the portal's aims. In addition, if across the many EU websites there are various terms of use, on the portal by default there is one term in line with decision on reuse (see 'Metadata on the EU Open Data Portal', page 34).
- **Stimulate improved data management, across EU organisations:** this will foster standardised use of metadata, find common solutions for publishing open data and publish it in machine-readable formats at interinstitutional level. This will eliminate duplication of data, reduce costs and increase interoperability.
- **Engage with the user community,** by offering the public:
 - a way to request additional data not yet published on the portal;
 - the possibility to give feedback on the data, thus increasing its quality;
 - a showcase with graphical representations of data in order to better understand it;
 - the opportunity to publish results from the use made of the EU data from the portal.

Engaging with user communities is also done via social media and EU information multipliers (Commission representations, Parliament information offices, delegations, etc.).

1.8.2. How to use the EU Open Data Portal and how it works

The EU Open Data Portal is a catalogue of metadata, describing data which remains stored on the data providers' websites. The portal is run by the Publications Office and is based on open source technologies, namely:

- Comprehensive Knowledge Archive Network (CKAN) platform, for data management,
- Virtuoso triple store database, for management of linked open data (RDF format),
- Drupal, for web content management,
- SOLR, for indexation of metadata in CKAN.

Thanks to these functionalities, users search for information in the portal's metadata catalogue and in different ways via its various entry tabs as explained below.

Data tab, allowing searches in the portal database (CKAN) in a conventional way for a human user. The search can be launched on the basis of various criteria such as publishers, titles or themes. The results will return one or more dataset, each giving a description of the data

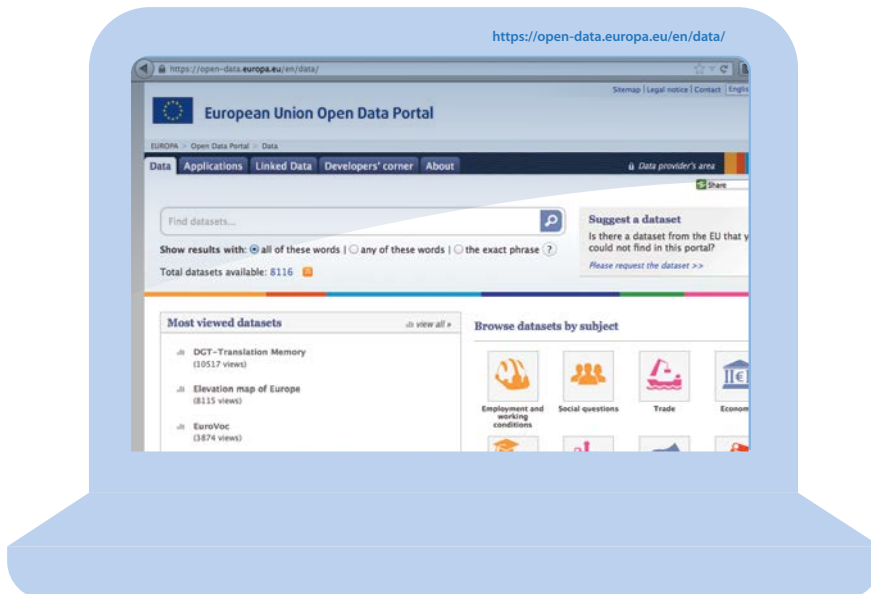


(metadata), including the link to the website where the data is stored. By clicking on this link the user is redirected to the data provider's website.

Applications tab, displaying examples of graphical representations of data. Since data in raw format can be difficult for the human eye to read, applications are developed to visualise them. The examples displayed are authored by data providers, and in the future also by external parties who have reused data from the EU Open Data Portal.

Linked data tab, allowing searches for data stored in the portal's triple store database (Virtuoso) via a Sparql endpoint. Sparql is the query language designed specifically to query RDF databases and the Sparql endpoint interface is constructed so that the queries are launched, provided the human user has a certain knowledge of RDF and Sparql. The search returns answers in a meaningful way, in a format chosen by the user (RDF, JSON, HTML, XML, CSV). Yet, for those users who are not at all familiar with the Sparql endpoint a series of sample queries are made available on the same page. These queries correspond to the most popular searches that users could launch on the basis of the content of the Virtuoso database.

Developers' corner, addressed to technical users. It explains how to access the portal's database programmatically by machines in two ways: via a CKAN-generated API or via the Sparql endpoint.



1.9. Who are the users of the portal?

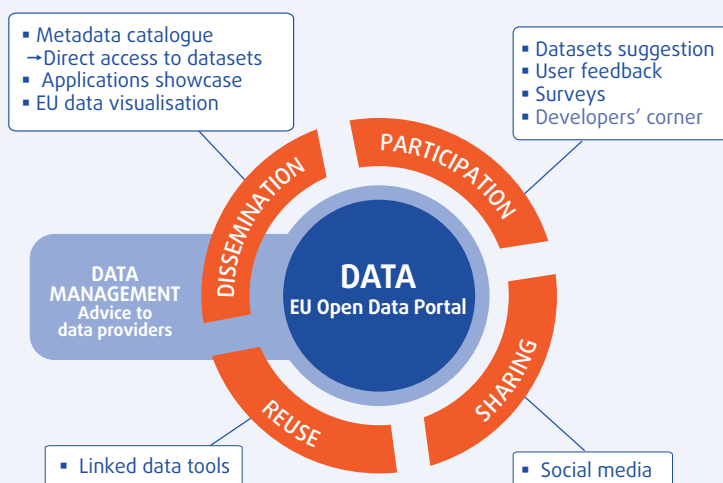
In general, users can be identified in the category types below:

- IT developers and businesses, to create new products and services;
- academia, researchers, statisticians and students, to use for their papers;
- journalists, for fact-finding;
- public administrations, lobbyists, interest groups and NGOs, for monitoring and supporting arguments and decision-making processes;
- open data specialists, who follow up open data issues and specialise in specific fields including research, academic, science, civil society, etc.;
- aggregators, who aggregate data with other sources to create new knowledge;
- citizens, who wish to be better informed.

1.10. What are the benefits for your organisation in disseminating data via the portal?

Making data accessible via the portal will provide an opportunity to:

- enhance the accessibility, reuse, visibility and discoverability of your data;
- enable the interoperability of data for machine reuse;
- make your data easy to find across languages;
- identify your open data at the outset;
- give your organisation an image of openness and transparency;
- comply with Commission Decision 2011/833/EU, when applicable.





Recurrent excuses for data hugging

The usefulness and potential of open data is not always understood by everyone. Below is a sample of 'data hugging' excuses, for which we try to provide an answer or clarify the misconception.

My data is not very interesting.

Let others judge how interesting or useful it is. Even niche datasets have people that care about them.

We are not sure we own the data.

When data has not been produced by the EU organisation's staff, the contract with the third party providing the information can clarify the ownership issues. It could be the case that the contract is not clear enough because data openness is a relatively recent concept. In this case, you can contact the contractor and, most importantly, plan specific clauses on the technical (formats) and legal openness of the data for the next contract specifications.

Ultimately, this is a worthwhile exercise in order to have an overview of your organisation's output, of what data can be made openly available and of what data can be made available only under specific conditions.

It will generate an increased number of user requests on the interpretation of data.

On the contrary, when you provide raw data, end-users will extract and combine the data according to their needs. This will potentially reduce the number of requests. This is, for example, the case for CORDIS datasets which are accessible in bulk download files via the EU Open Data Portal. Interestingly, making these files accessible has resulted in a decrease in the number of requests from users for specific data.

I am afraid people will draw superficial conclusions from the data without understanding the wider picture.

In any form of publication the risk is there. Generally, the data is accompanied by supplementary documentation which put the data into context.

I do not think my data is sufficiently accurate to be shared and reusable.

You cannot keep data always to yourself, just as you cannot keep your children always with you. When children are mature enough parents let them go. The same with data. There is a time of maturity; a time when data must be made available for reuse.

It will cost too much to put the data into a standard format.

When planned during the production phase, the production of or conversion to an open standard format requires little effort. Keep in mind that the same information in various formats can make a huge difference for the users compared to the invested efforts from the data provider.

For whom are we opening up data?

For society as a whole and for different reasons. You will find a typology of users and the use which can be made out of data on page 20 of this brochure.

Will there be a negative effect of putting data on the Open Data Portal for our own visibility and branding?

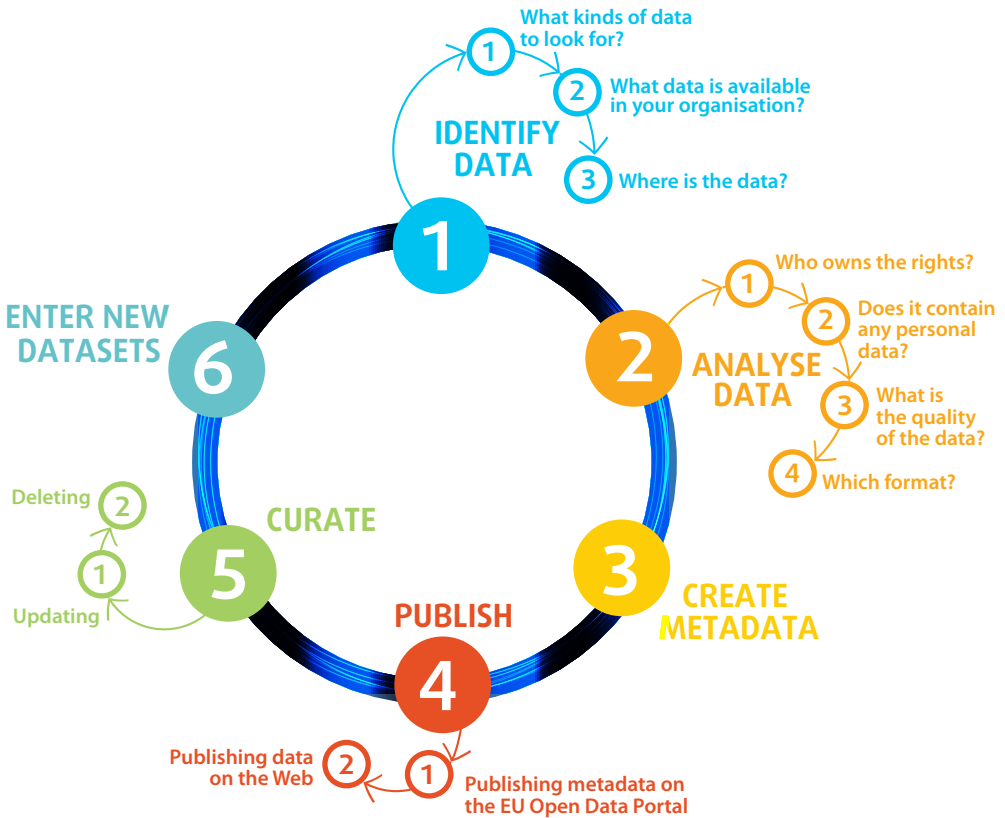
One of the main objectives of the EU open data policy is that it leads to the wider use and to the spread of Union information, enhancing the image of openness and transparency of the institutions.

Contributing to the EU Open Data Portal, one building block of this policy, helps foster this image, which is in our now digital time well scrutinised by the community.

In addition, the portal is a catalogue of metadata, describing data stored on the data providers' websites. The user is each time directed to the data provider's website. This potentially increases the traffic to that website.



SETTING UP THE PROCESS



Some questions to ask yourself about the data

- ▶ Are the URLs stable?
- ▶ Can you provide documentation about your data?
- ▶ What kind of release: a one-off release or several updates?
- ▶ Is it a single file or several linked pieces of data? If several pieces, can I merge them into one?
- ▶ How often are you planning to publish the data?
- ▶ Are you planning to publish as soon as the information is made public?
- ▶ In what format(s) is your data available?
- ▶ Is your data also available in machine-readable format?
- ▶ Is your data openly licensed ?

2. Opening up your data via the EU Open Data Portal

2.1. Setting up the organisational structure

Opening up your data is a continuous exercise rather than a one-off action. As data changes over time, it needs to be updated and new data has to be made regularly available.

Setting up the organisational structure starts with identifying and appointing a contact point who will liaise with staff that have the right IT, legal and data-management knowledge. It will also involve looking at the overall information management process and deciding where to establish and include data production and dissemination.

Data is not static, therefore your strategy might need to evolve. It could be the case that, as a result of monitoring and analytical tools, such as statistics and user feedback, an information need could arise in a specific area.

As a result, production of new data could be envisaged or made available on the Web, if not previously done for practical reasons.

2.2. Setting up the process

Moving from a general view to an operational analysis, we can identify the following main steps in opening up data.

- 1 Identifying data
- 2 Analysing the identified data
- 3 Creating metadata
- 4 Publishing datasets and metadata
- 5 Curating
- 6 Entering new datasets



1 Identifying data

Having a full insight into what data is available is essential in order to decide what should be opened up and what should not. Where is the data, what kind of data you can find and what kind of data to give priority to are questions to ask during this phase.

1. What kinds of data are we looking for?

Data can be any piece of content, in whatever medium, which is stored electronically and made accessible on the Web. By a piece of content we mean any objective, factual, non-personal and non-aggregated information that we collect or produce during our activities.

Some examples are provided below.

- Statistics and time series.
- Geographic, geospatial and meteorological data.
- Reference data (thesauri, vocabularies, ontologies, etc.).
- Sets of bibliographic records of:
 - documents (publications, legal documents, cases, opinions, reports, recommendations, studies, requests, approvals/rejections, patents, calls for tender, archives, registries, appeals, speeches, interviews, staff directories, organisational charts, CVs, etc.);
 - speeches, interviews, documentaries, reports, charts, diagrams, etc.
- Sets of data about projects, programmes, actions, measures, operations, procedures, grants, etc.
- Sets of resource management data, e.g. financial data, service contracts, logistical data or facility management data.

In the framework of its implementation plan for the Open Data Charter, the EU's commitments included identifying core datasets (budget, elections and statistics) at EU level, as well as preparing guidelines for EU Member States relating to publishing core and high-value datasets. The guidelines ⁽¹³⁾ were adopted by the College of the Commission on 17 July 2014. They singled out a series of data for priority release, namely: geospatial data, environment and earth observation, transport, statistics and company data.

Moreover, the charter recognises 14 areas as high-value data the release of which should be prioritised. These are companies, crime and justice, earth observation, education, energy and environment, finance and contracts, geospatial, global development, government accountability and democracy, health, science and research, social mobility and welfare, statistics, and transport and infrastructure. For more details, see the table below.

⁽¹³⁾ Guidelines are available in all official languages (except Irish) and published on the Official Journal C 240, 24.7.2014, Commission notice — Guidelines on recommended standard licences, datasets and charging for the reuse of documents (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2014.240.01.0001.01.ENG)



DATA CATEGORY (ALPHABETICAL ORDER)	EXAMPLE DATASET
Companies	Business register
Crime and justice	Crime statistics, safety
Earth observation	Meteorological/weather, agriculture, forestry, fishing and hunting
Education	Lists of schools, performance of schools, digital skills
Energy and environment	Pollution levels, energy consumption
Finance and contracts	Transaction spend, contracts let, calls for tender, future tenders, local budget, national budget (planned and spent)
Geospatial	Topography, postcodes, national maps, local maps
Global development	Development aid, food security, extractives, land
Government accountability and democracy	Government contact points, election results, legislation and statutes, salaries (pay scales), hospitality/gifts
Health	Prescription data, performance data
Science and research	Genome data, research and educational activity, experiment results
Social mobility and welfare	Housing, health insurance, unemployment benefits
Statistics	National statistics, censuses, infrastructure, wealth, skills
Transport and infrastructure	Public transport timetables, access points, broadband penetration

Source: G8 Open Data Charter, technical annex.

2. What kinds of data are interesting?

What data is interesting for civil society? The answer is not always straightforward. As general rule, when data is available with no restrictions, open it.

Asking questions could help get an idea of the data's potential economic and social interest. For example: what is the quality of bathing water in the EU compared to South America? What is the fastest public transport available in a town for a person in a wheelchair? How has the EU spent money on buildings? How have MEPs cast their votes on a given topic? What is the gender balance in the EU public sector? How do prices compare for similar services?

Generally, it is the data community that will identify its added value, whether via an application developer to create a new product, data journalists to understand and inform or researchers, analysts or academia to monitor and analyse. Ultimately, the reuse of data may result in new uses not initially foreseen by its author.



3. What data is available in your organisation?

The large amounts of data produced or collected in the natural course of EU organisations' activities can be broadly classified into two data types.

- **In-house output:** which is the production by internal staff of data related to annual activities and management reports. This includes budget figures, number of staff, number of visits to a website, voting records, statistics, etc. This type of data, when opened up, enhances your organisation's image of transparency and openness.
- **External entities' deliverables:** we often collect data from various sources, both inside our organisation and from third parties, such as Member States, national and international organisations, consultants, external companies or academic institutions.

This data is then analysed, selected and validated in order to be, in most cases, published in official reports, publications, studies, databases, etc. Where contractually possible, you should make sure you receive from the external entity all the files, possibly in various formats, and the right to reuse this data (an example of contract clauses can be found in Annex 2).

All the cases above can include both data that has been published and that which has not been published, for economic or other practical reasons, such as studies and reports.

4. Where is the data?

Some of our organisations manage a large amount of data daily, but because of the huge quantity it is not always centralised or immediately traceable. So, first of all, you can identify and outline which units in your organisation produce potentially suitable data.

Then the place where data is stored should be found. Today, all our information is available electronically, but it could be stored in different locations, in various databases or in archives with restricted access.

In the end this is a worthwhile exercise, not only in order to define a strategy of open data but also to review the output of your organisation.

Where is my data? A fictitious example of data outlining

Unit/person responsible	Source of data	Subject of data	Publication/storage
OP C.1 — Common Portal and Open Data Portal Unit — Jean Dupont	Unit C.1 — Open Data Portal section	Eurovoc	On Europa website (http://eurovoc.europa.eu/drupal)
OP C.2 — EUR-Lex and TED — John Smith	Contractor ABC	TED historical archives	H Drive XYZ



2 Analysing the identified data

Once the data outline has been defined, it is necessary to analyse the identified data and assess the feasibility of opening it up. The analysis covers legal and technical aspects. For this purpose, you should answer the following questions.

1. Who owns the rights?
2. Does it contain personal data?
3. What is the quality of the data?
4. In what format is the data available?

1. Who owns the rights?

On the basis of the reuse policy principles, information can be reused, whatever its medium, for commercial or non-commercial purposes, free of charge and without having to request permission to do so.

The right of reuse does not cover those elements for which the EU organisation does not hold the copyright.

This is why it is essential to properly identify those pieces of information belonging to third parties and to distinguish them from those that can be freely reused. If you are not sure about the copyright owner or the clauses in the contract are not clear as regards reuse, then you need to contact the owner to clarify these issues before you can open up.

2. Does it contain any personal data?

Reuse is only allowed in full compliance with the personal data protection rules. Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data applies.

3. What is the quality of the data?

After establishing the list of data that can be disseminated from a legal point of view, you can then analyse the quality of the data from a technical perspective. Good quality will enable an easier discoverability, access, searching and machine-readability process and easier reuse.

Data quality has multiple dimensions and is more than just the correctness of data. The main facets to be taken into account for delivering good quality data are as follows.

- **Accuracy:** does data correctly represent an object, situation or event? For example, information about weather conditions can vary in specificities such as 'partly rainy' or a 'rainfall probability of 70 %'. Accuracy has a certain cost and therefore must be measured against the target audience, whether specialist or not.



- **Availability:** can the data be accessed over time? Best practice for open data recommends the assignment of URIs (unique resource identifiers) for the long-term availability of data.
- **Completeness:** is all relevant data part of the dataset that is included? For example, if the dataset is a list of universities in the European Union at a given period, the list should not cover only a part of them.
- **Conformance:** does data follow a set of standards or rules for publication and description? For example, a description of a dataset according to the DCAT application profile.
- **Consistency:** does data contain contradictions? This could be the case when the same dataset has different licences, or where the date of last modification is before the creation date.
- **Correct semantics and syntax,** otherwise, the search could be affected. There could be a mistake in the description of the data, such as 'Climat quality' instead of 'Climate quality'.
- **Timeliness:** is the data published in its latest version and updated frequently? For example, timeliness is crucial for those datasets that contain real-time traffic data. It also bears risks: for example, an error on a roadmap could be fatal. Because of the risks entailed, it is good practice to provide information about the timeliness of the data so that users can decide if the data is relevant for their purpose.

You can find more information on the quality of data and metadata in the online training material published on the Europa website ⁽¹⁴⁾.

4. In what format is data available?

To answer this question, you should be aware of the possibilities of processability provided by file formats. For this purpose, it is useful to know about openness and how to distinguish open formats from closed formats, or their degree of openness.

From closed formats to open formats

In data, we distinguish between information published in open and closed formats ⁽¹⁵⁾.

An **open** format is one where the specifications for the software are available to anyone, free of charge, so that anyone can use these specifications in their own software without any limitations on reuse imposed by intellectual property rights.

If a file format is **closed**, however, this may be either because the file format is proprietary and the specification is not publicly available or

⁽¹⁴⁾ <https://joinup.ec.europa.eu/community/ods/document/tm22-open-data-metadata-quality-en>

⁽¹⁵⁾ For the definition of open and closed formats see the Open Knowledge Foundation handbook (<http://opendatahandbook.org/en/appendices/file-formats.html>).



because the file format is proprietary and, even though the specification has been made public, reuse is limited.

If information is released in a closed file format, this can create significant obstacles to the reuse of the information encoded in it.

The preference from the open government data perspective therefore is that information be released in open file formats that are machine readable.

From open data to linked data

Each time information is linked — combined with other information — it increases in value because it creates new knowledge, in most cases with economic, intellectual and social relevance.

However, reusing, integrating and making sense of data from different sources is a real challenge. In many cases, in fact, data is published without adhering to a standard, thus preventing the combination of data. Data should be structured in a standard way so that software applications can easily identify, recognise and extract specific data (machine-readable format), and so that the file format is platform independent and made available with no restrictions (open format).

As shown in the five-star scheme below there are various degrees of openness. So-called linked data (LD) is the highest level of openness and can respond to these challenges by providing a set of principles for sharing machine-readable data on the Web. This way data can communicate (interoperate), be understood and thus be combined into a larger system. As a result, this can entail the generation of new products and/or services that have a direct or indirect economic or social impact.

Tim Berners-Lee's five-star scheme

To distinguish between various formats and their degree of openness, the World Wide Web Consortium (W3C) ⁽¹⁶⁾, the main international standards organisation for the World Wide Web, has proposed a classification ranking from one to five stars.

- It is a good practice to provide the same information in various formats, including one for human readability and at least one in three-star format. Different users require varying modes of access to information. Providing multiple ways of access, including data as a bulk download, will help ensure variation of access and reuse regardless of the user's technological capability.

Often this is a small effort for data providers, but makes a lot difference for the user. Ultimately, this will increase both the number of users and ease of access and reuse.



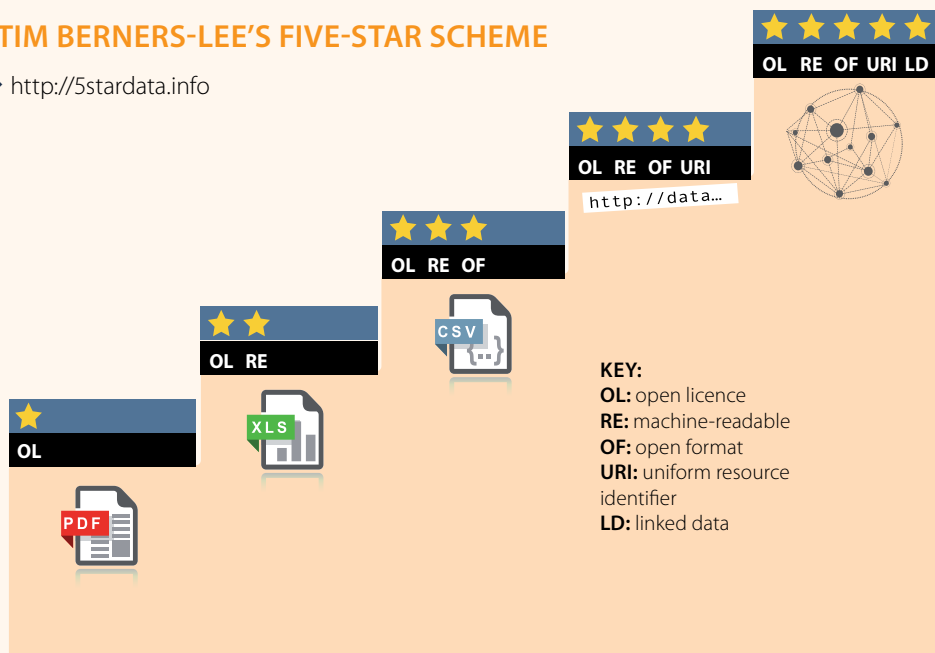
Good practices for publishing high-quality data and metadata

- Provide appropriate descriptions of data (metadata).
- Use standard vocabularies for metadata (i.e. Eurovoc, named authority lists, international standards).
- Specify the licence.
- Adhere to legal requirements concerning protection of personal data.
- Provide information about the source of the data.

⁽¹⁶⁾ W3C was founded and is currently led by Tim Berners-Lee, the creator of the World Wide Web. It is an international community that develops open standards to ensure the long-term growth of the Web.

TIM BERNERS-LEE'S FIVE-STAR SCHEME

► <http://5stardata.info>



★ ONE STAR

- Available on the Web whatever the format.
- Non-structured format.
- With an open licence.
- Readable and printable online, but processing is not possible.

Example: For an image: GIF, JPG, PNG.

For a document: Microsoft Word, Adobe PDF.

★★ TWO STARS

- Structured data but in a proprietary format.
- Unlike one star data can be elaborated, but under proprietary software conditions.

Example: Microsoft Excel instead of a scanned image of a table.

★★★ THREE STARS

- Same as for two stars, though data is also structured and codified in a non-proprietary format.

Example: CSV (comma-separated values) instead of Microsoft Excel.

★★★★ FOUR STARS

- All the above plus open standard (RDF and Sparql) to identify things.
- Data is structured and codified in a non-proprietary format with a URI (uniform resource identifier) that makes the data usable directly online and in a structure based on the RDF model.

In practice it means that a single piece of data from a dataset is available online in an open format (generally XML or RDF) and can be accessed through a specific URL. This allows it to be accessed by an application or from a programme that can then process it.

Example: A codified dataset consisting of a map plus a list of bus stops: from any piece of software it is possible to connect to a URL indicating a single bus stop, and then reference it on a map.



★★★★ FIVE STARS

- All the above plus linking data to other sources of data to provide context.
- Datasets are linked to others, combining information from various sources, thanks to the RDF descriptive model.

Example: If we continue with the example of the map with the bus stops, it is possible in addition, when clicking on the bus stop, to tag in the map extra information from other sources. This could

be, for example, the price of the bus ticket, places to eat close to the bus stop, the number of bus lines, their timetables, etc.

This information is given by linking to other sources (in this case, these could be the website of the ministry of transport, the tourist office, the local bus company, etc.). As these primary sources regularly update this information as part of their activity, it becomes more useful to link rather than aggregate this dynamic information on the map.

TIM DAVIES' FIVE STAR OPEN DATA ENGAGEMENT

► <http://www.opendataimpacts.net/engagement/>

Publishing open data is not just about technology. Tim Berners-Lee's five stars of linked open data set out a series of approaches that open data initiatives can take to publish data on the web. The following five-star model seeks to add to this, highlighting key steps that open data initiatives can take to engage with data users. Each star includes a set of questions and actions needed to unpack what might be involved in taking that step towards engagement.

★ ONE STAR ► *Be demand driven*

- Data structure, tools and support based on community needs and demands.
- Ways of listening to people's requests for data, and responding with open data.

★★ TWO STARS ► *Put data in context*

Provide metadata (e.g. frequency of updates, data formats and data quality).

- Include qualitative information such as data creation, or data manuals.
- Link from data catalogue pages to analysis of the data.

★★★ THREE STARS ► *Support conversation around data*

Can people comment on datasets, or create structured conversation around data?

- Do you join the conversations?
- Easy ways to contact the individual 'data owner'.
- Offline opportunities to have conversations that involve your data.

★★★★ FOUR STARS ► *Build capacity, skills and networks*

Tools for people to work with your datasets.

- 'How to' guidance on using open data analysis tools.
- Do you go out into the community to run skill-building sessions?
- Do you sponsor or engage with capacity building?

★★★★★ FIVE STARS ► *Collaborate on data as a common resource*

Have feedback loops so people can help you improve your datasets.

- Collaborate with the community to create new data resources.
- Provide support to people to build and sustain useful tools and services.
- Work with other organisations to connect up your data sources.

This is a RECORD

European Union Open Data Portal

[Legal notice](#) | [Contact](#) | [Search](#) | [English \(en\)](#)

[EUROPA](#) > [Open Data Portal](#) > [Data](#) > [Publisher](#) > [Publications Office](#) > [CORDIS - EU research projects un...](#)

[Data](#) | [Applications](#) | [Linked Data](#) | [About](#)

[Data provider's area](#)

CORDIS - EU research projects under FP7 (2007-2013)

Publisher
 Publications Office

Framed in orange, the METADATA

Licence:
 Europa Legal Notice

Description

This dataset contains projects funded by the European Union under the seventh framework programme for research and technological development (FP7) from 2007 to 2013. Grant information is provided for each project, including reference, acronym, dates, funding, programmes, participant countries, subjects and objectives. A smaller file is also provided without the texts for objectives.

Keywords

Controlled thermonuclear fusion
 FI1
 FP7
 International co-operation activities
 Management of radioactive waste
 Marie Curie Actions - Human resources and mobility
 New and emerging science and technology - NEST -
 Radiation protection
 Research and Innovation
 Research for policy support
 Research infrastructures
 Science and society
 Specific SMF activities
 Strengthening the foundations of FRA experts
 health
 ict
 list
 research

Resources

Download	CORDIS FP7 Programmes	Excel
Download	CORDIS FP7 Programmes	CSV
Download	CORDIS SIC codes	CSV
Download	CORDIS SIC codes	Excel
Download	CORDIS activity codes	Excel
Download	CORDIS activity codes	CSV
Download	CORDIS country codes	Excel
Download	CORDIS country codes	CSV
Download	FP7 CORDIS project participants part 1	Excel
Download	FP7 CORDIS project participants part 2	Excel
Download	List of all FP7 projects (excluding Abstracts) Extra	
Download	List of all FP7 projects (excluding Abstracts) Extra	
Download	List of all FP7 projects (including Abstracts) Extra	
Download	List of all FP7 projects (including Abstracts) cordis	

THIS IS A DATASET

Catalogue record

Added to open-data.europa.eu
 2013-12-04
 Updated on open-data.europa.eu
 2014-02-04

URI
http://cordis.europa.eu/fp7/home_en.html

Alternative Title
 FP7 research projects

Status
 Completed

Interoperability Level
 Legal

Release Date
 2013-11-27

Modified Date
 2013-11-27

Geographical Coverage
 EU27, Croatia

Language
 English

Contact
 CORDIS helpdesk
 2 rue Mercier, Luxembourg
 Tel: +352 2929 42210
cordis-helpdesk@publications.europa.eu
http://cordis.europa.eu/guidance/helpdesk/home_en.html

Ref	Project title	Acronym	Start date	End date	Duration	Coordinator	Participant countries	Programme	Activity	Subject	Keywords	Abstract	Objectives	Results	Impact	Dissemination	Other
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1	CORDIS	EU27	FP7	Research and Innovation	Research for policy support	Controlled thermonuclear fusion						
213757	Support to		2013	2013	1												



3 Creating metadata

1. What is metadata?

'Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data'⁽¹⁷⁾.

In a few words, metadata provides information about other information to make it easier to find and manage. Basic metadata elements include, for example, the author, title, year of publication, etc.

2. What is the difference between data, metadata and datasets?

In the world of open data you will come across three key recurring terms: data, datasets and metadata.

- **Data** can be any piece of electronically stored content.
- A **dataset** is a collection of data grouped according to certain criteria. The collection of one or many data files is referred to as a 'resource'. For instance, in the example of the CORDIS dataset in the graph opposite, there are 14 resources for the same dataset. This makes it easier for users to locate related data and easier for data providers to manage and maintain it.
- **Metadata** describes each dataset with properties such as title, author, description of the content, URL, etc.

You will also find their graphical representation on the basis of a dataset taken from the portal (opposite).

3. Why is metadata important?

Metadata is important for many reasons, most specifically to:

- facilitate discovery;
- enable high ranking of search results;
- enable refinement of a search;
- help organise electronic datasets;
- provide digital identification;
- support archiving and preservation;
- facilitate interoperability, that is the ability of systems to exchange information and use the exchanged information.

4. Why standardise metadata?

Over the years a myriad of metadata elements and vocabularies have been created in different systems, often to identify a single concept such as a country, a document or an organisation. As a consequence, the exact meaning and purpose of a metadata property can be ambiguous between different systems. Also, this variation hinders data exchange between applications and the receipt of trustworthy search results.

This is why standards organisations have developed and maintain common vocabularies to make metadata understandable across various sys-

⁽¹⁷⁾ Source: National Information Standards Organisation (NISO).

tems and across disciplines (for example metadata specific to publications, geospatial information, audiovisual material, people, etc.).

In doing so, metadata is mapped to a common system and, when different systems communicate, the sender and the receiver share a common understanding of the meaning of the data.

5. Why is the quality of metadata important?

By quality of metadata we mean information that is accurate and complete. Comprehensive metadata affects the discoverability of your data and users know it can be trusted.

Taking into account the user's perspective and behaviour helps enhance the quality of metadata. For example, one tip when giving information about the datasets: the description should include all the important keywords in the first sentence and not exceed 160 characters, according to search engine optimisation (SEO) best practices. In this way, data has a greater chance of being discovered and having a high rank in an Internet browser.

6. Metadata on the EU Open Data Portal

Metadata on the portal is based on international standards, principally the following:

- data catalogue vocabulary (DCAT) ⁽¹⁸⁾;
- the Dublin core metadata initiative (DCMI), for metadata attributes ⁽¹⁹⁾;
- asset description metadata schema (ADMS), for interoperability assets ⁽²⁰⁾;
- the Friend of a Friend (FOAF), to describe people and organisations ⁽²¹⁾.

The portal will also implement the DCAT application profile (DCAT-AP), the specification for describing public-sector datasets in Europe, after it has been revised in summer 2015.

In addition, you will find metadata created to answer specific requirements of the EU organisations, particularly their multilingual needs. The Publications Office has identified and developed a list of common vocabularies for labelling the themes of datasets (Eurovoc) ⁽²²⁾ and named authority lists (such as countries, corporate bodies, interinstitutional procedures, languages, etc.) ⁽²³⁾.

To ensure that the same metadata and encoding rules are used by all data providers, the EU institutions work together in the Interinstitutional Metadata Maintenance Committee (IMMC) established for this purpose.

⁽¹⁸⁾ DCAT : <http://www.w3.org/TR/vocab-dcat/>

⁽¹⁹⁾ DCMI: <http://dublincore.org/>

⁽²⁰⁾ ADMS: <https://joinup.ec.europa.eu/asset/adms/home>

⁽²¹⁾ FOAF: <http://xmlns.com/foaf/0.1>

⁽²²⁾ Eurovoc, the multilingual thesaurus of the EU, <http://eurovoc.europa.eu>

⁽²³⁾ The Metadata Registry registers and maintains data definitions used by EU institutions, such as named authority lists (<http://publications.europa.eu/mdr/>).



The metadata schema set up for the portal is published on the same portal (linked data tab) ⁽²⁴⁾.

7. Legal notice

An important metadata element is the legal notice for your data. Reuse of content catalogued on the EU Open Data Portal is permitted free of charge for commercial or non-commercial purposes, provided that the source is acknowledged, unless otherwise stated.

According to its copyright notice, 'Reuse is authorised provided the source is acknowledged.'

This follows the principles of the reuse policy implemented through Directive 2013/37/EU and Decision 2011/833/EU. In the same way, reuse should fully respect privacy legislation and does not apply to data subject to the intellectual property rights of third parties. In limited cases it can be subject to conditions (Article 2(2) of Decision 2011/833/EU).

For specific needs in this area the Publications Office will give you advice and support.

⁽²⁴⁾ <http://open-data.europa.eu/en/linked-data>.



4 Publishing datasets and metadata

1. Where to publish

To facilitate searching and access across the various EU websites, the EU Open Data Portal provides a single point of access to EU data.

The portal is based on a decentralised architecture: datasets remain stored on the data provider's website and their metadata is stored and published on the portal. The portal provides a catalogue of metadata that directs the user via a link to the data on the data provider's website.

2. Publishing datasets on the Web

Your dataset should be normally published on the website of your organisation. Your website was probably not designed to have a data corner or a place for extracted data relating to PDF publications or to databases, charts, graphics, etc.

You might envisage, for future developments, identification of the right strategy to locate this data, for example by offering database export tools in structured formats or by providing subcategories of final reports with tables and graphics as separate files in their original format.

As a temporary solution you might also place data on the server with a URL that is not accessible via your website but can still be linked to and accessible via the portal.

3. Publishing metadata on the EU Open Data Portal

By making metadata available on the portal you make your data more easily discoverable and allow searches to be more refined.

On the EU Open Data Portal metadata is stored both in a conventional database, readable by humans (search box), and in a triple store, readable by machines (Sparql endpoint and API), to facilitate interoperability.

Metadata can be entered on the portal in two ways, as follows:

- via the website user's interface, by filling in a predefined form;
- via an FTP server, sending a zipped package of RDF files. This is advisable for frequent and regular publications with the possibility of metadata being published in 24 languages.

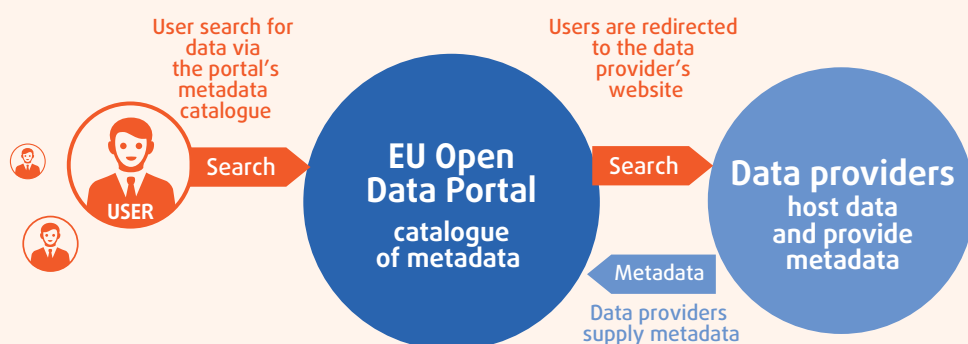
Some metadata is compulsory (e.g. title, publisher, URL, licence) but other metadata is optional (e.g. file format, file size, geographical coverage).



You will find the step-by-step procedure for entering metadata on the portal in the user manual, electronically available on the PubliCare website ⁽²⁵⁾.

HOW DOES IT WORK?

1. Data providers publish online data on their website.
2. Data providers create the related metadata and supply it to the portal for publication.
3. Users can search for data via the portal's metadata catalogue. They are redirected to the data provider's website to access it.



⁽²⁵⁾ <https://webgate.ec.europa.eu/multisite/publicare/en/page/open-data-355>

5 Curating

After their creation and publication, both metadata and datasets have to be maintained. They are always linked, even if they are not published in the same location. Updating and deleting data has an effect on the metadata published on the portal.

Moreover, if you have started with a pilot project to be discovered via the EU Open Data Portal, the identification and prioritisation of data should continue. The early experience will help to speed up the process for later projects. It will help you to review and open other channels in your output's production process and to think of future strategies for the dissemination of information on your website.



If the data is uploaded automatically and regularly updated, include its frequency in the description of your data (e.g. 'this data is updated on a daily basis'). This way you do not need to modify the metadata each time the data is updated, and the user will be informed how recent the data is no matter what its publication date.

1. Updating

When your data is modified, whether manually or automatically, the link to its location generally remains the same. For this reason, you might not think of the impact of this change on the portal.

However, each version of a dataset has different metadata, which requires an update for each modified dataset. For example, if the responsible organisation changes as result of a reorganisation this might affect the publisher's name, its website address or the temporal coverage.

If the data is uploaded automatically and regularly updated, it is advisable to include its frequency in the description of your data (e.g. 'this data is updated on a daily basis'). This way you do not need to modify the metadata each time the data is updated, and the user will be informed how recent the data is no matter what its publication date.

Apart from the description, you could update a dataset if any of the following happen.

- A new file format of previously published information becomes available.
- The contact email has changed.
(As for the contacts, in order to avoid unnecessary updates, it is strongly recommended not to use an individual's email address. It is better to use a shared inbox as this provides for continuity, should the named data custodian change.)
- The organisation name has changed. This is often the case following a reorganisation.
- Your organisation has migrated to a new website with a new URL. Consequently the link previously provided for your data is no longer active or brings the search to a blank page.



2. Deleting

If data is no longer available because it is no longer valid or for other reasons, the metadata should be kept with the information that the data has been deprecated. And in fact we use the term deprecation instead of deletion because it is not possible to completely delete all mention of the data. Use of 'deprecated' informs the enduser that the data is now obsolete, because it has been either changed or suppressed.

Metadata needs to be kept over time. In general, the metadata life cycle is longer than the data life cycle, as it is created before the data is published and needs to be kept after data has become obsolete.

6 Entering new datasets

After the first project it will be easier and faster to identify new data to make accessible via the portal. As open data is a relatively recent requirement and there are expectations with regard to output, start thinking of new elements to be included in the production processes of your organisation on the basis of past experience.

And with a different mindset ...

As with every change made to a process, at the beginning new challenges such as organisational issues, release of open (machine-readable) formats, fear of losing control of your data and copyright issues are faced.

However, visitors to your website have different objectives, from simple consultation or download to reuse and linking.

With the digital revolution not only is technology changing but also culture is being shaped: new businesses that were inconceivable only a few years ago have come into being; a new generation of employees and new working patterns have changed workplaces; media and communication are largely influenced by users, who can promote themselves with no need for intermediaries; and politicians interact directly with voters, while censorship is becoming more difficult. The list could go on.

Citizens are now more attentive regarding how their public administrations work and are keen on having a say in decision-making. Opening up public data is part of adapting to this new reality. The data revolution is still at an early stage!



Here is a list of general actions we suggest be kept in mind for data publishing and curation.

- **Appoint a contact person** in charge of data coordination.
 - **Think of disaggregated data at the outset:** raise awareness with your authors, whether internal staff or contractors, to identify information before it is combined to produce a report or other form of aggregated output. The identified data should remain available in the raw disaggregated format and may have more than one workflow for dissemination.
 - **Check copyright issues** in order to know the terms of use of your data: when you publish a work delivered by a contractor, be sure you sign sound contracts with appropriate clauses concerning copyright or other intellectual property rights allowing its online reuse and linking.
 - **Think of other deliverables in your contracts:** if you produce data via a contractor, think of requesting deliverables in open formats with rich metadata descriptions accompanying them.
 - **When publishing information graphics** or other types of visualisations, add also the source file of the same visualisation.
 - **Publish in open formats**, not just formats such as PDF, images, databases with no export tools or HTML pages. You might consider, for example, providing alongside a report in PDF format tables and graphics extracted in an open format (e.g. CSV, XML, JSON), or source data in a tabular format alongside graphics and interactive applications. If you publish a database, offer export tools in various formats.
-



ANNEX 1

Glossary

API	<p>Application programming interface.</p> <p>A way computer programmes talk to one another. Can be understood in terms of how a programmer sends instructions between programmes.</p> <p>Source: http://schoolofdata.org/handbook/appendix/glossary</p>
Big data	<p>Big data refers to large amounts of data produced very quickly by a high number of diverse sources. Data can either be created by people or generated by machines, such as sensors gathering climate information, satellite imagery, digital pictures and videos, purchase transaction records, GPS signals, etc. It covers many sectors, from healthcare to transport and energy.</p> <p>Source: http://ec.europa.eu/digital-agenda/en/making-big-data-work-europe http://ec.europa.eu/digital-agenda/en/big-data</p>
(Data) Crawling	<p>A crawler is a program that visits Web sites and reads their pages and other information in order to create entries for a search engine index. The major search engines on the Web all have such a program, which is also known as a 'spider' or a 'bot'.</p> <p>Source: http://searchsoa.techtarget.com/definition/crawler</p> <p>When extracting data from the Web, the term 'crawling' is often also referred to as 'data scraping' or 'harvesting'. There is a difference between these terms: crawling refers to dealing with large datasets where one can develop his or her own crawlers (or bots), which crawl to the deepest parts of the web pages. Data scraping on the other hand refers to retrieving information from any source (not necessarily the Web).</p> <p>Source: https://www.promptcloud.com/blog/data-scraping-vs-data-crawling/</p>
Dataset	<p>A collection of related sets of data that is composed of separate elements but that can be manipulated as a unit and accessed or downloaded in one or more formats.</p>
DCAT	<p>Data catalogue vocabulary.</p> <p>RDF vocabulary for interoperability of data catalogues.</p> <p>See also: W3C — http://www.w3.org/TR/vocab-dcat</p>
DCAT-AP	<p>DCAT application profile.</p> <p>A common vocabulary for describing datasets hosted in data portals in Europe, based on the data catalogue vocabulary (DCAT).</p> <p>See also: https://joinup.ec.europa.eu/asset/dcat_application_profile/description</p>



(Data) Dump	<p>A large amount of data transferred from one system or location to another.</p> <p>Source: http://www.oxforddictionaries.com</p>
Interoperability	<p>The ability of systems to exchange information and use the exchanged information.</p>
Linked data	<p>Linked data describes a method of publishing structured data so that it can be interlinked. It builds upon standard Web technologies such as HTTP and URI, but rather than using them to serve Web pages for human readers it extends them to share information in a way that can be read automatically by computers.</p> <p>Source: http://www.wikipedia.org</p>
Linked data principles	<p>Linked data principles provide a common API for data on the Web that is more convenient than many separately and differently designed APIs published by individual data suppliers. Tim Berners-Lee, inventor of the Web and initiator of the linked data project, proposed the following principles upon which linked data is based.</p> <ol style="list-style-type: none">1. Use URIs to name things.2. Use HTTP URIs so that things can be referred to and looked up (dereferenced) by people and user agents. When someone looks up a URI provide useful information using open Web standards such as RDF or Sparql.3. Include links to other related things using their URIs when publishing on the Web. <p>Source: W3C — http://www.w3.org/TR/ld-glossary/#linked-data-principles</p>
Machine readable	<p>Formats that are machine readable are ones which are able to have their data extracted by computer programs easily. PDF documents are not machine readable. Computers can display the text nicely but have great difficulty understanding the context that surrounds the text. Common machine-readable file formats are CSV and Excel files.</p> <p>Source: http://schoolofdata.org/handbook/appendix/glossary</p>
Mashup	<p>Combination of multiple datasets from multiple sources to create a new service, visualisation or information.</p>
Metadata	<p>Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data.</p> <p>Source: NISO — http://www.niso.org/publications/press/UnderstandingMetadata.pdf</p>



(Data) Mining	<p>The practice of examining large pre-existing databases in order to generate new information.</p> <p>Source: http://www.oxforddictionaries.com</p> <p>‘For example, one midwest grocery chain used the data mining capacity of Oracle software to analyse local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer display closer to the diaper display. And, they could make sure beer and diapers were sold at full price on Thursdays.’</p> <p>Source: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm</p>
Ontology	<p>A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms).</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#ontology</p>
Open standards	<p>Generally understood as technical standards which are free from licencing restrictions. Can also be interpreted to mean standards which are developed in a vendor-neutral manner.</p> <p>Source: http://schoolofdata.org/handbook/appendix/glossary</p>
Open government data	<p>Refers to content that is published on the public Web by government authorities in a variety of non-proprietary formats.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#open-government-data</p>
(Data) Parsing	<p>Breaking a data block into smaller chunks by following a set of rules, so that it can be more easily interpreted, managed or transmitted by a computer.</p> <p>Source: http://www.businessdictionary.com/definition/parsing.html</p>
Raw data	<p>An expression that refers to data in its original state that has not been processed, aggregated or manipulated in any other way. It is also defined as ‘primary’.</p>
RDF	<p>Resource description framework.</p> <p>A family of international standards for data interchange on the Web. RDF is based on the idea of identifying things using Web identifiers or HTTP URIs, and describing resources in terms of simple properties and property values.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#rdf</p>



Resource	<p>The physical representation of a dataset. Each resource can be a file of any kind, a link to a file elsewhere on the Web or a link to an API. For example, if the data is being supplied in multiple formats or split into different areas or time periods, each file is a different 'resource' which should be described individually.</p>
Semantic Web	<p>An evolution or part of the World Wide Web that consists of machine-readable data in RDF and an ability to query that information in standard ways (e.g. via Sparql).</p> <p>Source: W3C — http://www.w3.org/TR/ld-glossary/#semantic-web</p>
(Data) Scraping	<p>The process of extracting data in machine-readable formats of non-pure data sources e.g.: webpages or PDF documents. Often prefixed with the source (Web-scraping, PDF-scraping).</p> <p>Sources: http://en.wikipedia.org/wiki/Data_scraping http://schoolofdata.org/handbook/appendix/glossary</p>
Sparql	<p>Sparql protocol and RDF query language (Sparql) defines a query language for RDF data, analogous to the structured query language (SQL) for relational databases.</p> <p>Source: W3C — http://www.w3.org/TR/ld-glossary/#sparql</p>
Sparql endpoint	<p>A service that accepts Sparql queries and returns answers to them as Sparql result sets. It is a best practice for dataset providers to give the URL of their Sparql endpoint to allow access to their data programmatically or through a Web interface.</p> <p>Source: W3C — http://www.w3.org/TR/ld-glossary/#sparql-endpoint</p>
Structured data	<p>Data that resides in fixed fields within a record or file. Relational databases and spreadsheets are examples of structured data. Although data in XML files is not fixed in location like traditional database records, it is nevertheless structured, because the data is tagged and can be accurately identified.</p> <p>Source: PC Magazine encyclopaedia — http://www.pcmag.com/encyclopedia/term/52162/structured-data</p>
Triplestore	<p>A triplestore is a purpose-built database for the storage and retrieval of triples through semantic queries. A triple is a data entity composed of subject-predicate-object, like 'Bob is 35' or 'Bob knows Fred'.</p> <p>Much like a relational database, one stores information in a triplestore and retrieves it via a query language. Unlike a relational database, a triplestore is optimised for the storage and retrieval of triples. In addition to queries, triples can usually be imported/exported using RDF and other formats.</p> <p>Source: Wikipedia.org http://en.wikipedia.org/wiki/Triplestore</p>



URI	<p>Uniform resource identifier.</p> <p>A string that uniquely identifies virtually anything including a physical building or more abstract concepts such as colours. It may or may not be resolvable on the Web.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#uniform-resource-identifier</p>
URL	<p>Uniform resource locator.</p> <p>A global identifier commonly called a 'Web address'. A URL is resolvable on the Web. All HTTP URLs are URIs: however, not all URIs are URLs.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#uniform-resource-locator</p>
Vocabulary	<p>A collection of 'terms' for a particular purpose. Vocabularies can range from simple, such as the widely used RDF schema, FOAF and Dublin core metadata element set, to complex vocabularies with thousands of terms, such as those used in healthcare to describe symptoms, diseases and treatments. Vocabularies play a very important role in linked data, specifically to help with data integration. The use of this term overlaps with 'ontology'.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#vocabulary</p>
Web 1.0	<p>The first generation of the World Wide Web, characterised by separate static websites rather than continually updated weblogs and social networking tools.</p> <p>Source : http://en.wiktionary.org/wiki/Web_1.0</p>
Web 2.0	<p>A colloquial description of the part of the World Wide Web that implements social networking, blogs, user comments and ratings and related human-centred activities.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#web-2.0</p>
Web 3.0	<p>A colloquial description of the part of the World Wide Web that implements machine-readable data and the ability to perform distributed queries and analysis on that data. It is considered synonymous with the phrases 'semantic web' and 'the Web of data'.</p> <p>Source: W3C — http://www.w3.org/TR/Id-glossary/#web-3.0</p>

ANNEX 2

Good practices in data management

The JRC open data project

(Kindly shared by the Joint Research Centre)

The Joint Research Centre (JRC) has recently adopted a corporate data policy with the aim of making the maximum of its data available for further reuse.

The work on the JRC data policy follows the challenge identified in the JRC management plan 2014 to develop a strategy for open access to its data to complement the JRC policy on open access to scientific publications and supporting guidance and to promote open access to research data in the context of Horizon 2020.

An important policy commitment includes the Commission Decision on the reuse of Commission documents (2011/833/EU of 12 December 2011), and the JRC data policy has also been developed in line with the principles of the G8 open data charter and the Commission recommendation on access to and preservation of scientific information (COM(2012) 4890).

The objectives of the JRC data policy include the following.

- To share and use data on the basis of open data: fully, freely, openly and in a timely manner.
- To ensure optimal conditions for the use of purchased third party data.
- To facilitate consistent data management at the JRC.
- To enable data which is used in publications to be accessed and cited.
- To broaden access to and use of the JRC data.
- To continuously support evidence-based decision-making and research.
- To ensure reproducibility of research.
- To ensure that JRC data is made available through the EU Open Data Portal.
- To reinforce the goals of Horizon 2020.
- To support the EU implementation of the G8 open data charter.

Contacts: Anders Friis-Christensen, Fabrizio Bonato



Guidelines for publishing open datasets and studies

(Kindly shared by the Directorate-General for Communications Networks, Content and Technology)

DG Communications Networks, Content and Technology applies transparency policy and promotes reuse of the relevant documents, reports and datasets. For this purpose, in May 2014 it adopted guidelines for publishing open datasets and studies.

The objective of the guidelines is to set out common procedures when publishing study results and open datasets. Particularly, it aims to improve their searchability through homogeneous tagging and by maintaining easy access to inventories (see <http://ec.europa.eu/digital-agenda/en/digital-agenda-you/our-studies-and-open-datasets>).

The relevant resources concerned are as follows.

- Reports and studies procured by the directorate.
- The open datasets maintained by the directorate in its administrative activity (i.e. research grants, reports from Member States) or created via contracted studies (i.e. raw data from surveys, inventories and statistics). Datasets are published with their metadata — needed for data interpretation and reuse (metadata include codebooks, methodology and definitions, original questionnaires and administrative reports).

Publication of open datasets is announced on the EU institutions' Open Data Portal. Selected key indicators are published on the <http://digital-agenda-data.eu> semantic repository. They can be easily consulted, and downloaded via interactive charts.

Contact: Stefano Abruzzini

Eurostat big data task force

(Kindly shared by Eurostat)

Objective and timeline ⁽²⁶⁾

The European statistical system (ESS) big data action plan and roadmap 1.0 ⁽²⁷⁾ was produced by a task force composed of representatives of national statistical offices, the Organisation for Economic Cooperation and Development (OECD), the United Nations Economic Commission for Europe (UNECE), DG Communications Networks, Content and Technology, the JRC, academic experts and Eurostat. It is the response to the Scheveningen memorandum ⁽²⁸⁾, which called for the adoption of an ESS action plan and roadmap by mid 2014.

⁽²⁶⁾ <http://www.cros-portal.eu/content/big-data>

⁽²⁷⁾ <http://www.cros-portal.eu/content/ess-big-data-action-plan-and-roadmap-10>

⁽²⁸⁾ <http://www.cros-portal.eu/content/scheveningen-memorandum>



The European statistical system committee (ESSC) endorsed the ESS roadmap and action plan 1.0 at its meeting on 26 September 2014. In addition the ESSC agreed on the continuation of the work of the ESS big data task force on the implementation of the roadmap and action plan.

The long-term goal of the roadmap and action plan is to prepare the ESS for integration of big data sources into the production of official statistics across the ESS. The action plan and roadmap are part of the implementation strategy of the ESS Vision 2020 ⁽²⁹⁾, in particular with regard to how the ESS will respond to the data revolution and integrate new data sources into official statistical production.

List of use cases

The actions identified in the action plan are grouped around the following topics that are relevant for the development of the overall big data domain:

- policy
- communication
- big data sources
- applications/pilots
- methods
- quality
- IT infrastructure
- skills
- experience sharing
- legislation
- governance.

Among the big data sources that were found most promising were:

- the Internet as a source for information society statistics ⁽³⁰⁾,
- the use of mobile phone data for tourism statistics ⁽³¹⁾,
- the use of publicly available online shopping data for price statistics.

Copyright clauses for contracts

<http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies>

Including data-related clauses in our tenders will be beneficial in many ways.

- In addition to the commissioned deliverable(s), we would also receive one or more datasets that can be processed in ways not provided for by the original deliverable.

⁽²⁹⁾ <http://ec.europa.eu/eurostat/documents/10186/756730/ESS-Vision-2020.pdf/8d97506b-b802-439e-9ea4-303e905f4255>

⁽³⁰⁾ <http://www.cros-portal.eu/content/analysis-methodologies-using-internet-collection-information-society-and-other-statistics-0>

⁽³¹⁾ <http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies>



- Citizens would have access to datasets that have already been paid for.
- These datasets would be of presumed high quality and could be interlinked with other data, for example from organisations such as the World Bank and linked data versions of Wikipedia or DBPedia. This could increase the potential reuse of EU data.
- Consultants would bring expertise on linked data principles and connections with a large number of entities with whom they interact. This added value could contribute to the creation of a European data ecosystem to the benefit of everyone.

Below you will find extracts from clauses about the provision of data from a tender specifications document. This tender is meant to be used when procuring services like studies, evaluations, impact assessments and other data provision-based services for open procedures. However, clauses about linked open data can also be used in administrative arrangements, and in negotiated, low-value framework contracts and/or specific contracts under framework contracts.

Text in *italics* refers to optional text to be adapted, while text **highlighted in yellow** refers to instructions to be deleted.

The deliverables listed below must be provided by the contractor.

Deliverables

...

Deliverables can be reports, datasets, presentations, minutes or any other type of service to be defined in this section.

...

Relevant data shall be available for reuse ⁽³²⁾, thus the deliverables paragraph should include their provision no later than the final report, and describe the technical requirements to facilitate their publication and data linking. The 'Objective' and 'Methodology' chapters should describe the raw datasets the study is expected to collect and the indicators it is expected to compute from the raw data or derive from existing sources.

Tenderers should be encouraged to provide the datasets in a format suitable for publication as linked open data. They should be invited to describe the approach they will adopt for that purpose and such aspects should be assessed as part of the technical quality award criteria.

⁽³²⁾ According to the Commission decision of 12 December 2011 on the reuse of Commission documents (2011/833/EU)

The deliverables listed below must be provided by the contractor.

- ...
- The final datasets, as described under task [complete], should be provided as structured data in a machine-readable format (e.g. in the form of a spreadsheet and/or an RDF file) for Commission internal use and for publishing on the EU Open Data Portal, in compliance with Commission Decision 2011/833/EU. If third parties' rights do not allow their publication as open data, the tenderers should describe in the offer the subpart that will be provided to the Commission free of rights for publication and the part that will remain for internal use. The raw datasets and/or the derived indicators shall be made available to the Commission's services within [number of months] months after signature of the contract by the last contracting party.
- ...

Data formats

The raw datasets should comply with the following provisions.

- The data delivered should include the appropriate metadata (e.g. description of the dataset, definition of the indicators, label and sources for the variables, notes) to facilitate reuse and publication.
- The data delivered should be linked to data resources external to the scope of the study, preferably data and semantic resources from the EU Open Data Portal or from the upcoming pan-European portal. The tenderer should describe in the offer the approach they will adopt to facilitate data linking. **Specify if this should remain a generic requirement to be evaluated in the award phase or if the study requires the mandatory adoption of specific semantic assets, vocabularies or taxonomies.**
- Additional example that could be introduced in the text: Companies mentioned in the datasets should be cross referenced with the company identifiers made available by Open Corporates ⁽³³⁾.
- If the study is analysing legislation: Use standard European legislation identifiers ⁽³⁴⁾. For a list of shared data interoperability assets see the Joinup catalogue ⁽³⁵⁾ from DG Informatics' ISA programme.
- In the case of statistical data that could be used to derive/compute indicators (e.g. for benchmarking national policies), the

⁽³³⁾ <http://opencorporates.com>

⁽³⁴⁾ http://en.wikipedia.org/wiki/European_Legislation_Identifier

⁽³⁵⁾ <http://joinup.ec.europa.eu/catalogue/all>



contractor should use templates provided by (or agreed with) Commission services (such as those available at <http://ec.europa.eu/digital-agenda/en/download-data>) on the data cube vocabulary.

- Specify details about data granularity. The final data deliverable should include an aggregated dataset with all the information on the variables listed in section ... [i.e. for comparison over time or with other sources]. The spreadsheet should also include a pivot table of the variables listed in section ... x.
- Define rules about the anonymisation of microdata. This could be in the sense of requesting clear identification of the units analysed/interviewed, for example in view of future data collections concerning the same units, or in the sense of requiring an anonymised version of the raw data for publication purposes. The contractor will have to provide a non-anonymised raw dataset with clear identification of the units analysed/interviewed [specify].

The final datasets should be considered when setting the terms for the approval of deliverables.

The final datasets should be described in the timetable.

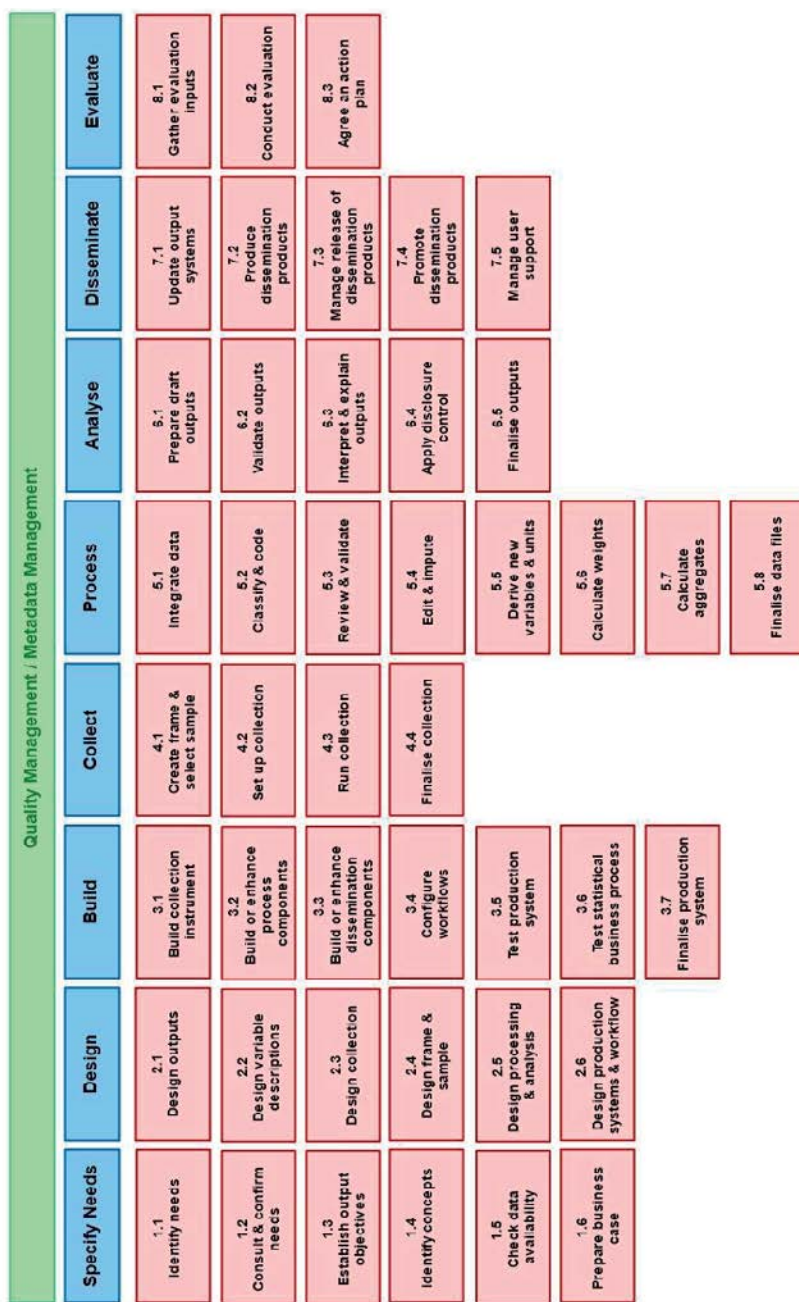
Technical award criteria

The tenders will be qualitatively assessed on the basis of the technical award criteria and respective scores listed below.

Technical award criterion	Maximum score/ weighting	Threshold
1.Understanding of the tasks required (examples of optional sub-criteria):	30	15
...		
2. Technical quality of the tender (examples of optional sub-criteria):		
2.1. ...		
2.2. Quality of the methodological approach/methodology used to collect and analyse data/soundness and appropriateness of the proposed analysis tools and data gathering techniques.	50	25
2.3. Quality of proposed deliverables in- cluding the final datasets for publication and data linking.		
2.4. ...		

Generic statistical business process model (GSBPM)

(Kindly shared by Eurostat)



Source: <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

References

- Digital agenda for Europe (<http://ec.europa.eu/digital-agenda/en/content-and-media/data>; http://ec.europa.eu/information_society/digital-agenda) European Commission, 2014.
- 'Digital agenda, the Commission's open data strategy, questions and answers' (MEMO/11/891), European Commission, December 2011.
- 'Scientific data: open access to research results will boost Europe's innovation capacity', European Commission press release, 17 July 2012.
- Shadbolt, N., 'Towards a pan EU data portal', December 2010.
- Commission communication — Open data — An engine for innovation, growth and transparent governance (COM(2011) 882), December 2011
- Vickery, G., 'Review of recent studies on PSI reuse and related market developments', (study carried out for the European Commission), August 2011.
- Online training material on how to publish open data, European Commission, September 2013 (<http://ec.europa.eu/digital-agenda/en/news/online-training-material-how-publish-open-data>).
- 'G8 Open Data Charter and technical annex', June 2013 (<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>).
- 'EU implementation of the G8 Open Data Charter', European Commission, October 2013 (<http://ec.europa.eu/digital-agenda/en/news/eu-implementation-g8-open-data-charter>).
- 'Open government', Wikipedia (http://en.wikipedia.org/wiki/Open_government).
- 'Open data', Wikipedia (http://en.wikipedia.org/wiki/Open_data).
- 5-star model on open government data, Michael Hausenblas (<http://5stardata.info>) (page last updated April 2012).
- World Wide Web Consortium (W3C) website (<http://www.w3.org/DesignIssues/LinkedData.html>).
- 'Eight principles of open government data', December 2007, (<http://www.opengovdata.org/home/8principles>).
- 'Publishing open government data' Open Government Working Group, (<http://www.w3.org/TR/gov-data>).
- 'Open government data for citizen engagement in managing development', United Nations, 2013.

- 'Linee guida per i siti web della PA', Presidenza del Consiglio dei Ministri, 2011.
- Gilles, W., 'Open government and democracy: the French case', Fourth International Conference on Democracy as Idea and Practice, University of Oslo, 10/11 January, 2013.
- 'Open data handbook', Open Knowledge Foundation, November 2012.
- Bauer, F., and Kaltenböck, M., 'Linked open data: the essentials', edition mono/monochrome, Vienna, January 2012.
- 'Open data: Unlocking innovation and performance with liquid information', McKinsey & Company, October, 2013.
- Linked open data glossary, W3C, June 2013 (<http://www.w3.org/TR/ld-glossary>).
- Trend report open data, Algemene Rekenkamer, March 2014
- Open data toolkit, Australian government (https://toolkit.data.gov.au/index.php?title=How_to_use_data.gov.au).



Keep up to date with EU open data via:

Twitter @EC_opendata

Facebook and Google+: 'EU law and publications'

LinkedIn: Publications Office of the European Union

Yammer: 'EU Open Data Portal + your data'

Joinup collaborative platform: joinup.ec.europa.eu

***This brochure can be downloaded and ordered from
EU Bookshop: <https://bookshop.europa.eu>***

