

Report on licence usage on the data.europa.eu portal

An analysis of trends, practices, and recommendations for
data providers

data.europa.eu
The official portal for European data



This study has been prepared as part of data.europa.eu, an initiative of the European Commission. The Publications Office of the European Union is responsible for the management of data.europa.eu contracts.

For more information about this paper, please contact the following.

European Commission

Directorate-General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

data.europa.eu

Email: info@data.europa.eu

Author:

Hans Graux

Last update: 14 January 2025

<https://data.europa.eu/>

DISCLAIMER

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission’s behalf may be held responsible for the use that may be made of the information contained herein.

Luxembourg: Publications Office of the European Union, 2024

© European Union, 2024



The Commission’s reuse policy is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39, ELI: <http://data.europa.eu/eli/dec/2011/833/oj>). Unless otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licences/by/4.0/>). This means that reuse is allowed, provided appropriate credit is given and any changes are indicated.

ISBN: 978-92-78-44725-0

doi: 10.2830/7448965

Catalogue number: OA-01-24-397-EN-N

Contents

Abbreviations	4
What are licences and how are they used?	5
Why are licences important?	6
Objectives of this report	7
Introduction	8
Characteristics of open licences on the portal.....	9
Permission-related characteristics.....	10
Reproduction	10
Distribution	10
Derivative works	11
Sublicensing	11
Use patent claims.....	12
Obligation-related characteristics.....	13
Notice	13
Attribution.....	14
Share-alike.....	15
Copyleft	15
Lesser copyleft	16
State changes	16
Prohibition-related characteristics	17
Commercial	17
Scoping-related characteristics.....	18
Database rights	18
Use trademark.....	18
Methodological note.....	19
General statistics and trends	21
Zooming in on the most common licences	26
Licence groups	26
International group.....	26
National and regional group	29
EU sectoral group – Inspire	39
Observations on key trends	40

Abbreviations

CC	Creative Commons
CC0	Creative Commons Zero
CC BY 4.0	Creative Commons Attribution 4.0 International
DL-DE-Zero-2.0	Data licence Germany – Zero – Version 2.0
INE	Spanish Statistical Office
Inspire	Infrastructure for Spatial Information in Europe
ISTAC	Canarian Institute of Statistics
ODbL	Open Data Commons Open Database Licence
OGL	Open Government Licence 3.0
UK	United Kingdom

1. Introduction to licences and open data

What are licences and how are they used?

Within the open data community, and especially on open data portals, licences can play a critical enabling role for data reusers. An open data approach implies that available datasets can be used as easily and as freely as possible, by any interested party. As the [open data directive](#) notes in its recitals, open data is ‘generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose. Open data policies which encourage the wide availability and re-use of public sector information for private or commercial purposes, with minimal or no legal, technical or financial constraints, and which promote the circulation of information not only for economic operators but primarily for the public, can play an important role in promoting social engagement, and kick-start and promote the development of new services based on novel ways to combine and make use of such information’.

To achieve this goal of openness, it must be clear what exactly data reusers are allowed to do with specific datasets. From a legal perspective, this is clarified by licences. A licence [generally refers to a document that denotes the terms of use](#) under which a specific work – such as a dataset – may be used. This is a very broad concept, which covers both traditional contracts (agreements between two or more parties) and unilateral declarations (a statement from the data holder that they will permit usage under certain conditions, or even waivers abandoning their ownership rights as far as possible).

When datasets are made available, they are generally accompanied by such a licence. By way of example, [on the official portal for European data](#), aspiring reusers can search for relevant data, and when selecting a dataset, the relevant licence is shown (see Figure 1).



Figure 1. Screenshot from a [randomly selected dataset](#), showing the applicable licence, in this case Creative Commons Attribution 4.0 International

Users can click on the licence name in order to access a copy of the full licence text, via which they can learn about any applicable constraint.

Why are licences important?

Intuitively, from an open data perspective, licences initially may not seem that essential. After all, open data is about making data freely available for reuse. If that goal is achieved, one might expect that licences would contain no restrictions at all. However, this is an oversimplification. When a dataset is referred to as ‘open data’, this implies that it is made available for any lawful use with only ‘minimal or no legal, technical or financial constraints’, as the aforementioned recital from the open data directive notes. Restrictions should thus be **minimal** but not non-existent, and they can vary depending on the licence.

While there is no official EU-level definition of specific legal requirements for an open data licence, common characteristics include a broad permission to access, reuse and redistribute a dataset, free of charge, and typically without any limitation on use cases or potential beneficiaries of the licence. Beyond these broad permissions though, it is also common – as we shall examine below – for licences to include an attribution requirement, i.e. requiring reusers of the datasets to indicate the source of the data, in various ways, indicated in the licence itself. Other constraints, such as a limitation to non-commercial use, or to make modified versions openly available for free reuse as well, are less common, but can also be a part of a licence’s constraints.

In that sense, licences may seem like a complicating factor for a reuser, since it requires them to review and consider permissible terms of use. However, licences can play an enabling role as well, as they can reduce doubts about the application and impact of intellectual property rights. Depending on the situation, a dataset could be subject to copyrights and/or database rights, whereas in other cases no intellectual property rights apply at all. It would be extremely difficult and risky for a data reuser to make such an assessment.

The presence of licences removes this problem, since they contain a clear statement of the applicable conditions, applied by the data provider. While any constraints in the licences may not be particularly welcome to a data reuser, the use of licences still removes at least a first layer of difficulty: the permissions and conditions become available to a reuser. In that sense, licences create transparency and remove doubt about usage rights.

Of course, in order to achieve this effect, it is important that simple and accessible licences are used by data providers wherever possible. Using obscure licences – which are complex in phrasing or only known to a limited community – will remove the benefits in terms of transparency and understandability: aspiring data reusers would in principle need to read them and reflect upon them, and by using uncommon or complex licences, this effort is needlessly increased.

Especially from an EU perspective, it is important for data providers to apply popular and commonly used international licences wherever possible, rather than opting for national (or even unique) licences, since they usually facilitate cross-border reuse. In this report, we will examine current trends and practices using empirical data, as will be explained in the following sections.

Objectives of this report

It can be challenging for both data providers (who, in principle, choose the licences that will apply to their datasets) and data reusers (who are required to respect the terms of these licences) to retain a clear understanding of their rights and obligations.

- Data providers have a broad range of licences to choose from, and their choice significantly impacts the possibilities of reuse, and thus the value of the datasets for reusers, and for the European market as a whole.
- Data reusers are required to identify the limitations imposed by licences in the datasets that they use, and to abide by them. This means that they must consult and understand these limitations. Moreover, when combining multiple datasets in a single service, application or product, they may need to respect multiple cumulative conditions if those datasets apply different licences, thus creating a problem of licence compatibility: all conditions from all licences then need to be respected.

This report aims to help resolve some of these challenges, first by explaining how open licences work (Section 2 of the report, explaining what their main characteristics are); and then examining which licences are commonly used in practice (Section 3, analysing this topic empirically based on data from the official portal for European data, <http://data.europa.eu>). Finally, in Section 4 we will provide recommendations to data providers on selecting an appropriate licence.

The overall goal of this report is to help European data providers to make beneficial choices, particularly by using the tools available on the portal, including its [Licencing Assistant](#). In that way, the providers can contribute to the creation of an efficient open data ecosystem in the European Union by making it easier for their data to be reused.

2. Characteristics of open licences

Introduction

As was already noted in the introduction, there is no official definition in the EU of what constitutes an ‘open data licence’, other than the broad description that open data should allow free use, reuse and sharing by anyone for any purpose. In reality, there is a very extensive range of open licences in common use in Europe, each of which vary in the permissions and restrictions that they contain. On 27 November 2024 (when the dataset used for this report was last updated), the data.europa.eu contained 1 815 243 datasets, which used more than 400 different named licences. In Section 3 we will examine in more detail what the most common licences and their features are in practice, since the popularity of these licences is of course very uneven: the most commonly used licence is applied by almost 280 000 datasets, whereas only 204 (i.e. around half) of the named licences are used by more than 50 datasets each.

However, in this section of the report, we will first examine a different issue: what are the characteristics of open licences in general, and how should they be understood? The simplest answer is of course that each licence should be read in detail, and appropriate legal counsel should be sought before a data provider selects a licence and before a data reuser decides to use a dataset. This can be an arduous task though, particularly when there are already more than 400 licences in use on a single portal.

For that reason, the data.europa.eu provides a Licencing Assistant, which can be used to more quickly search through known licences and their characteristics. The assistant allows a data provider or a data reuser to search for a specific licence by name (which is mainly relevant for data reusers who want to get a brief overview of what a specific licence entails), or to filter licences based on a predefined and standardised set of characteristics (which is mainly relevant for data providers who want to choose an appropriate licence for their dataset).

It is worth noting that the Licencing Assistant operates on the basis of the [EU vocabulary of licences](#), a list made available by the Publications Office of the European Union, described as ‘a controlled vocabulary listing the European Commission Reuse Notice and the European Union Public Licences. It also provides a list of standard licences available internationally, including the Creative Commons licences’. The vocabulary contains 164 named licences, most of which are also included in the more than 400 named licences used on the European data portal. Because of this difference, it is clear that very uncommon licences on the data.europa.eu are not present in the vocabulary, and thus also not in the Licencing Assistant. Nonetheless, all common licences (which, for reasons that will be explained below, comprises the licences used by more than 20 000 datasets on the portal) are also findable via the Licencing Assistant.

In the following sections, we will explain how the Licencing Assistant describes licences based on a standardised set of common characteristics. These characteristics allow a data reuser to quickly understand what a licence entails, and allow data providers to select licences that meet their needs for datasets that they want to publish.

For completeness, it should be noted that the Licencing Assistant does not provide legally binding or authoritative information, and that appropriate diligence should always be exercised before relying on its analysis.

Characteristics of open licences on the portal

In this section of the report, we will explain each of the characteristics of the licences that can be found with the Licencing Assistant, using the standard descriptors that the website of the assistant uses. There are four categories of descriptors:

- **permission-related characteristics**, i.e. the licence explicitly makes a statement on a permission that a reuser of the dataset receives from the data provider, allowing it to use the dataset for a specific purpose;
- **obligation-related characteristics**, i.e. the licence explicitly makes a statement on an obligation that a reuser of the dataset will have, which is a precondition for the lawful use of a dataset;
- **prohibition-related characteristics**, i.e. the licence explicitly makes a statement forbidding a reuser of the dataset to use it for a specific purpose;
- **scoping-related characteristics**, i.e. the licence explicitly makes a statement on the scope of the permissions and prohibitions; in practice, this relates to whether the licence also makes statements in relation to trademarks and/or database rights.

Each subsection below will describe a characteristic, explaining in each case:

(1) what it **means**, i.e. how the characteristic is interpreted by the Licencing Assistant; this is necessary because these concepts do not have official legal definitions;

(2) what the potential **challenges** are in interpreting and applying that characteristic in relation to specific datasets; this is mainly relevant for reusers who want to understand better how they are affected if they use a dataset;

(3) what the **impacts** are of applying a licence with that characteristic to a dataset; this is mainly relevant for data providers who want to understand better what the implications are of imposing certain characteristics via their licence choices.

Permission-related characteristics

Reproduction

Meaning	When this characteristic is present, the licence includes a permission to make copies of the dataset.
Challenges	No particular challenges in practice.
Impacts	This is the most trivial characteristic, which is present in all 164 of the licences covered by the Licencing Assistant. The explanation is simple: all datasets available on the portal allow downloading (either of an entire dataset, or of individual records of a dataset via application programming interfaces), which implies that a copy is made. Thus, a licence without the reproduction characteristics would not be lawfully available, and thus should not be published on the portal.

Distribution

Meaning	When this characteristic is present, the licence includes a permission to share the dataset with third parties. This includes the right to distribute copies of the dataset, and to display it publicly (sometimes also referred to as 'public performance'), for example, by showing it in an online service or app.
Challenges	<p>The characteristic can apply to the entire dataset (all of the dataset), or be limited to a subset. Moreover, the right to display the data publicly (e.g. by integrating it in an online service or app) is not always explicitly stated in the licence itself. The right to distribute data however normally implicitly comprises the right to display it.</p> <p>It is also possible that distribution is coupled with a specific prohibition or limitation, such as allowing distribution only towards an academic or educational audience; this is however a rare constraint in practice (existing only in 2 of the 164 licences covered by the Licencing Assistant).</p>
Impacts	This is an important characteristic, since it must be present before the dataset (or parts of it) can be made accessible to third parties (such as the users of an application or services); in the absence of this characteristic, the data can only be used internally (within the reuser's organisation or environment).

Derivative works

Meaning	When this characteristic is present, the licence includes a permission to modify the data, creating a new work (the 'derivative work'), and to distribute that derivative work to third parties.
Challenges	<p>The derivative work can be a modified dataset (a subset of the original), or an enriched dataset (combining the original with one or more other sources), or an application or service containing the dataset.</p> <p>It is worth recognising that the characteristic is sometimes interpreted more narrowly, namely as just the right to create a derivative work, without any permission to distribute it. In the Licencing Assistant however, it is only used when the resulting derivative work may also be distributed.</p>
Impacts	This characteristic must be present before the dataset can be modified, including by enriching or updating it (other than by downloading a new set from the original data provider). Without it, the original dataset must remain intact and unchanged.

Sublicencing

Meaning	When this characteristic is present, the licence includes a permission to also allow a third party (other than the reuser) to use the data, for example, by allowing that third party to distribute or create their own derivative works.
Challenges	<p>Sublicencing in principle allows the data reuser to become a data provider of the dataset, in a certain sense. The characteristic can be hard to evaluate in some licence templates, especially when sublicencing is not explicitly mentioned in the text of the licence, but the contents of the licence do make it clear that third parties other than the reuser are permitted to receive rights from the data reuser, without further interactions with the original data provider.</p> <p>It is worth noting that the Licencing Assistant does not consider this characteristic to be present when only the right for the reuser to distribute the dataset to a third party is covered (otherwise the sublicencing characteristic would be identical to the distribution characteristic, and thus largely meaningless in practice).</p>
Impacts	<p>This characteristic creates significantly greater flexibility in the reuse of datasets, since it allows the dataset to be made available for reuse by all reusers. In other words, it creates a 'network' of reusers who can build on the dataset without necessarily having to obtain it from the original data provider. This is important in creating an open data community.</p> <p>Inversely however, it also removes a certain amount of control from the data provider by design, since the intention and effect is precisely that the datasets (and certain rights to use them) can be obtained elsewhere.</p>

Use patent claims

Meaning	When this characteristic is present, the licence includes a permission from the data provider to use one or more patents .
Challenges	<p>Patent grants are very uncommon in the open data community, since raw data itself is in principle not patentable (i.e. patents cannot apply to the data itself). However, patent rights can still be granted via the licence, which then usually convey the right to use the data in ways that would otherwise infringe on a patent.</p> <p>In order for this characteristic to be applicable in the Licencing Assistant, it is sufficient that the licence mentions patent grants. One patent is enough; it is not necessary that the licence guarantees that all possible patents are licenced. It is also not required by the Licencing Assistant that the patent is identified, and the phrasing in the licence text can even be hypothetical ('if a patent exists, permission is granted'). Finally, it is not necessary that the data provider is itself the holder of the patents, or that it makes any claims on that point; it is sufficient that the data provider states via the licence that at least one patent licence will be granted if it exists.</p>
Impacts	<p>This characteristic does not appear to have a significant impact in practice in the European open data community; its presence in various licences is largely a legacy from the fact that many open data licences are also used in the open-source software development community, where patents are significantly more important (because software can, conceptually, be qualified as a patentable computer-implemented invention).</p> <p>The impact thus seems limited – the presence of this characteristic is beneficial for open data, but not crucial.</p>

Obligation-related characteristics

Notice

Meaning	When this characteristic is present, the licence includes an obligation to retain a reference to the licence in every use of the dataset.
Challenges	<p>As the description shows, the obligation must relate to the licence itself, not to the dataset or the data provider (see the attribution characteristic below, which does refer to the provider) – i.e. the licence terms themselves must be disclosed.</p> <p>In order for this characteristic to be applicable in the Licencing Assistant, the licence can require a reference to itself by name, or a URL, or an obligation to provide a copy of the entire licence itself, or any other situation mentioned in the licence where the licence has to be shared, including at the request of a data user. If no such requirement exists, then the characteristic is said in the Licencing Assistant not to be present.</p>
Impacts	<p>The inclusion of this characteristic is mainly important when the data provider wants to communicate under which terms a recipient of the data can receive the data from the data provider – i.e. it informs a recipient that it can potentially gain a copy of the data under those terms.</p> <p>It is a relatively light obligation that focuses on transparency only. It is very important in copyleft licences (see 'Copyleft' below).</p> <p>The data provider should keep in mind though that there are operational implications on the reuser as well: they will only be able to use the dataset if there is a way for them to communicate with the recipients of the data. This is not normally a problem, but if there is a use case that involves no display and no documentation to a user (e.g. a physical device without a display that stores and uses the data internally), it can be difficult or even impossible to meet the obligation.</p>

Attribution

Meaning	<p>When this characteristic is present, the licence includes an obligation to identify one or more sources from where the dataset originated.</p>
Challenges	<p>In order for this characteristic to be applicable in the Licencing Assistant, the licence must require the disclosure of the source of the dataset. Sources can be persons or organisations; this may be an actual name of a data provider, its contact info or a pseudonym.</p> <p>Since the Licencing Assistant provides templates of licence text, the characteristic is present if the licence has a generic requirement without specifying details (e.g. ‘Any use must disclose the name of the data provider’ without identifying the actual name of the data provider in the text). The characteristic is also present if the licence has a placeholder section where the copyright holder must be named (e.g. ‘Name: (add the name of the copyright holder)’).</p> <p>If no such requirements are present, then the characteristic is said in the Licencing Assistant not to be present.</p>
Impacts	<p>As the following sections will show, attribution is the most common obligation available in open data licences. It is arguably the easiest obligation to meet for reusers, and creates a benefit for the data providers as well.</p> <p>However, here too the data provider should keep in mind that there are operational implications on the reuser. As with the notice characteristic above, they will only be able to use the dataset if there is a way for them to communicate with the recipients of the data, since this is necessary to attribute a source.</p> <p>Moreover, the attribution requirement is specified in very different ways across the licences. Depending on the licence, attribution can cover (non-exhaustively):</p> <ul style="list-style-type: none"> • the identity (name) of the public administration that creates and maintains the data; • the identity (name) of the public administration that operates the portal via which the data is made available; • the technical source (URL) where it can be found. <p>Especially when a reuser wants to combine datasets with different attribution requirements, this creates some degree of administrative overhead.</p>

Share-alike

Meaning	When this characteristic is present, the licence includes an obligation to apply the same licence to any derivative works .
Challenges	<p>Share-alike is a relatively well known concept in open licencing, which aims to ensure that reusers of open data have to grant the same benefits to others that they have received themselves: since they have received a dataset under specific terms, they have to allow others to use that dataset or any result that they create using that dataset (a derivative work) under the same licence.</p> <p>For the purposes of the Licencing Assistant, the ‘same licence’ means ‘under substantively identical terms’ – not necessarily the same name, or using identical phrasing, but the rights must be substantively the same.</p>
Impacts	Share-alike is a substantial obligation, and relatively strict limitation upon reusers. It is arguably ‘fair’ in the sense that reusers must grant to others what they have received themselves, but this also eliminates some business choices and commercial options. The most obvious example is that the use of share-alike licences may stop a reuser from commercialising a derivative dataset, since they must offer it under the (free) licence that they have been granted. This may seem intuitively reasonable, but is not particularly conducive to enabling optimal exploitation of open data. For that reason, as we shall see below, it is not commonly used on the portal in practice.

Copyleft

Meaning	When this characteristic is present, the licence includes an obligation for reusers to grant at least equal permissions to their own users, which must include at least the right to reproduction and distribution and the right to create derivatives , in any use of the dataset and of derivative works.
Challenges	<p>Copyleft is a concept with widely varying descriptions in various communities, but the Licencing Assistant approaches it in a very structured way: in very practical terms, copyleft is treated as a form of share-alike that focuses on a broad set of rights, which must include at a minimum the right to reproduce, distribute and modify into derivative works (but may of course also include other obligations such as attribution and notice).</p> <p>In basic logical terms, for the Licencing Assistant, copyleft equals reproduction + distribution + derivative + share-alike (+ other rights or obligations).</p>
Impacts	Much like share-alike, copyleft is a significant obligation for reusers. It has a strong background in the open-source software community where the notion originated, but much less so in the open data community. Like share-alike, it is not particularly conducive to enabling optimal exploitation of open data, due to the constraints it imposes on further exploitation. For that reason, as we shall see below, it is not commonly used on the portal in practice.

Lesser copyleft

Meaning	This is identical to copyleft (i.e. the licence contains an obligation for reusers to grant at least equal permissions to their own users, which must include at least the right to reproduction and distribution and the right to create derivatives, in any use of the dataset and of derivative works), but with an explicit exception in the licence for combined works that merely interact with the original work – these do not need to be licenced under the lesser copyleft licence.
Challenges	Lesser copyleft is a concept that, much like copyleft, originated in the open-source software community, where it was created to mitigate the downsides of copyleft licences. Essentially, copyleft requires all of a derivate work to be licenced under the copyleft licence; whereas lesser copyleft limits this obligation to changes in the original work alone. Applied to datasets, this would mean that changes (updates, enriched databases, etc.) in an original lesser copyleft dataset would need to be made available to end users under a lesser copyleft licence, but not the entire product or service (e.g. an app or online service) into which that dataset was integrated.
Impacts	Lesser copyleft originated in the open-source software community, where the notion is easier to apply: software is usually composed of multiple modules, and it is thus easier to determine whether and how changes were made in specific lesser copyleft modules, to which the impact of a lesser copyleft module can be limited. In an open data concept, the notion of lesser copyleft is much less obvious. For this reason, as we shall see below, it is not commonly used on the portal in practice.

State changes

Meaning	When this characteristic is present, the licence includes an obligation to disclose changes made to the original dataset (if any).
Challenges	<p>In order for this characteristic to be applicable in the Licencing Assistant, the licence must require the disclosure of changes in the dataset (updates, omissions, extensions, etc).</p> <p>For the purposes of the Licencing Assistant, the characteristic is present if the licence requires that changes must be disclosed; it is not relevant whether the licence requires disclosure of who made changes, or what the changes were.</p> <p>If no such requirements are present, then the characteristic is said in the Licencing Assistant not to be present.</p>
Impacts	A state changes obligation is another transparency duty, comparable to attribute and notice, and thus relatively easy to meet for reusers. It has the benefit of making it easier to determine whether any problems that occur in the use of a dataset are likely due to defects in the original data, or due to changes made by the reuser.

	<p>Nonetheless, the obligation does create administrative burdens to keep track of such changes, so it should not be introduced lightly. Moreover, the state changes requirement is specified in very different ways across the licences. Depending on the licence, state changes can cover (non-exhaustively):</p> <ul style="list-style-type: none"> • the date of the latest update of the data on the source location; • a description of any changes made by the user; • the date of any changes made by the user. <p>Especially when a reuser wants to combine datasets with different attribution requirements, this creates some degree of administrative overhead.</p>
--	--

Prohibition-related characteristics

Commercial

Meaning	When this characteristic is present, the licence includes a prohibition against commercial use of the dataset.
Challenges	<p>In order for this characteristic to be applicable in the Licencing Assistant, the licence must prohibit commercial use, either using the expression ‘commercial use’, or reasonable variants such as ‘use for profit’, or ‘for any gain or benefit other than strictly personal or societal’.</p> <p>In practice, this notion can be difficult to interpret, since what constitutes commercial use is not always obvious. Selling datasets directly and for profit would of course be considered commercial use, but evaluations can quickly become complex (e.g. selling access to a service built around the dataset at-cost by a non-profit-making association, research or testing conducted by a company that may build a commercial product later based on that research or testing). The licences do not address such complex scenarios, which are left to users of the licence to resolve.</p>
Impacts	<p>When licences with this characteristic are applied to a dataset, the objective is to eliminate commercial exploitation, or at least to significantly reduce it. While this can be legitimate, European open data policy explicitly recognises the value of commercial exploitation of data, so choosing licences that only allow non-commercial uses can cause significant downsides. This must be carefully considered before accepting them.</p> <p>Moreover, the interpretation of the notion of ‘commercial use’ is not as obvious in all instances as one might initially think, again causing additional overhead that must be borne by the reusers of a dataset.</p> <p>For these reasons, as we will see below, licences that prohibit commercial use are rare on the portal.</p>

Scoping-related characteristics

Database rights

Meaning	When this characteristic is present, the licence explicitly grants permissions in relation to database rights .
Challenges	<p>Database rights are uniquely EU-based intellectual property rights, which generally do not have a clear equivalent in other regions of the world. For this reason, the characteristic is normally only found in licences created in the EU, or tailored towards an EU context.</p> <p>Database rights, also known as <i>sui generis</i> rights under the database directive, only apply when there has been a qualitatively and/or quantitatively substantial investment in the creation of the database. This criterion is generally not satisfied if the database is created naturally as a side-effect of another activity. For example, when a government has a specific mandate, and as a result of that work, it creates a database, then that database is not covered by database rights since no substantial investment was made in the database (separate from the substantial investment that was made in performing the mandate). For that reason, database rights seldom apply in an open data context.</p>
Impacts	Given the limited and often unclear relevance of database rights, and the fact that they are only known and used in the EU, the characteristic is fairly neutral in importance: it generally does not create significant barriers or incentives.

Use trademark

Meaning	When this characteristic is present, the licence includes a prohibition against the use of a protected name or designation .
Challenges	<p>In order for this characteristic to be applicable in the Licencing Assistant, the licence must contain a statement that no trademark permission is granted, or that permission to use a specific name or trademark must be requested. This characteristic is interpreted broadly in the Licencing Assistant: it applies of course when a specific trademark is mentioned by name, but also when it references trademarks in general, and/or when it references trade names or organisation names (irrespective of whether these are formally trademarks).</p> <p>The main difficulty in interpretation occurs when a licence notes that the dataset may not be used in a way that suggests endorsement, approval or support of the data provider – this will not be considered to fall under a trademark use constraint, unless it mentions use of a name more broadly.</p>
Impacts	In practice, the impact of this characteristic in an open data context is fairly limited: trademarks generally do not apply to most datasets, and when a data provider wishes to be credited, the preference is generally to rely on an attribution requirement that requires the provider to be mentioned by name.

3. Licences on the data.europa.eu portal

Having now established a conceptual overview of the role and importance of licences in Section 1, and a description of the most common characteristics of open data licences in Section 2, Section 3 of the report will provide some empirical data on what licences are actually used for open datasets in Europe and what they mean for data providers and reusers in practice.

Methodological note

It would of course not be possible to analyse the more than 1.8 million datasets that are [searchable via data.europa.eu](#) one by one, but thankfully this is also not necessary to obtain a good overview of the licences that are actually used. It is possible to gain access to usage statistics, grouped per licence on the portal, via the [Sparql Search](#) functionality on the portal. Using the following specific Sparql query, a list can be generated that indicates, licence by licence, how many datasets apply that licence.

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
```

```
PREFIX odp: <http://data.europa.eu/euodp/ontologies/ec-odp#>
```

```
PREFIX dct: <http://purl.org/dc/terms/>
```

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
PREFIX foaf: <http://xmlns.com/foaf/spec/>
```

```
SELECT DISTINCT ?distributionLicence (COUNT(?dataset) AS ?totalDatasets)
```

```
WHERE { ?dataset a dcat:Dataset;
```

```
dcat:distribution ?dist.
```

```
?dist dct:license ?distributionLicence.
```

```
}
```

```
ORDER BY DESC (?totalDatasets)
```

It was already noted above that, based on this data, more than 400 licences are applied across those datasets ⁽¹⁾. That could lead one to infer that, on average, each licence is used by about 4.538 datasets (around 18 million datasets divided by around 400 licences). But this conclusion assumes perfect availability of licence information – i.e. it assumes that each of the more than 1.8 million datasets is linked to one licence. Adding up all the numbers of datasets from the usage statistics, 1 230 495 datasets have at least one licence linked to them – for the other datasets, the licence was not entered at all, or not in a manner that allows it to be linked to a specific licence. This data too can be extracted via the aforementioned Sparql functionality, using the following query.

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>

PREFIX odp: <http://data.europa.eu/euodp/ontologies/ec-odp#>

PREFIX dct: <http://purl.org/dc/terms/>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX foaf: <http://xmlns.com/foaf/spec/>

SELECT (COUNT (DISTINCT (?dataset) AS?totalDatasets))

WHERE {?dataset a dcat:Dataset;

dcat:distribution?dist.

?dist dct:license?distributionLicence.

}
```

On the portal, missing licence information can be found easily. When clicking on a random recently added dataset, the following overview appears.



Figure 2. Screenshot from a [randomly selected dataset](#), showing a ‘No Licence Provided’ message

Thus, comprehensive analysis is not possible, and the analysis must be done on the basis of the 1.2 million datasets for which licence information is available (representing around two thirds of all datasets on the portal).

⁽¹⁾ The exact number cannot be determined via the query, since the list contains too many data errors. It generates more than 50 000 results; however, after around 300 results, virtually all results refer to data errors where a random string is erroneously identified as a licence. Bearing this in mind, the estimation of just over 400 licences is reasonable.

General statistics and trends

When looking only at the 1.2 million datasets for which licence information is available, it is clear that there are vastly popular licences and very rare ones: the most commonly referenced licence (Data licence Germany – attribution – Version 2.0) is used 278 958 times. The rarest ones are each used only once.

To optimise the validity of the analysis without having to be comprehensive, we will focus first on licences which are used more than 20 000 times on the portal. Looking at the raw data first – which contains several data errors that will be corrected below – we get the following result.

Licence usage count	Licence name as indicated on the portal (i.e. the formal title used in the metadata on the portal when a URL is shown; the URL is the formal title in the dataset)	Short description of the licence
278 958	Data licence Germany – attribution – Version 2.0 http://dcat-ap.de/def/licenses/dl-by-de/2.0	German licence (created by GovData , the German open data portal). Open licence with an attribution requirement (data provider, data source and licence name). Changes must be notified.
264 498	Creative Commons Attribution 4.0 International https://creativecommons.org/licenses/by/4.0/	International licence. Open licence, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
200 932	Licence of the Spanish Statistical Office https://www.ine.es/aviso_legal	Licence of the Spanish Statistical Office (<i>Instituto Nacional de Estadística</i> – INE). The URL is not correct; but the licence is available at https://www.ine.es/dyngs/AYU/en/index.htm?cid=125 . Open licence, with an attribution requirement (data portal, data provider, date of latest update).
188 650	Etalab Licence Ouverte / Open Licence https://www.etalab.gouv.fr/licence-ouverte-open-licence	French licence (created by Etalab , a department of the Ministry of Digitisation, but used more broadly in the French public sector). Open licence, with an attribution requirement (data provider, data source, date of latest update). References data protection, and affirms compatibility with the Creative Commons Attribution 4.0 International (CC BY 4.0) and the Open Government Licence 3.0 (OGL), among others. Explicitly references applicable French legislation.
115 243	Open Government Licence 3.0 https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/	United Kingdom (UK) licence (created by the United Kingdom’s National Archives , but used more broadly in the UK public sector). Open licence, with an attribution requirement (data provider, data source, licence). Explicitly states what is not covered by the licence (personal data, trademarks, identity documents, etc).
106 815	Licence of the Canarian Institute of Statistics http://www.gobiernodecanarias.org/istac/aviso_legal.html	Licence of the Canarian Institute of Statistics (<i>Instituto Canario de Estadística</i> – ISTAC) ⁽²⁾ Open rights statement with an attribution requirement (data portal, data provider, date of latest update).

⁽²⁾ ISTAC is the regional statistical authority and research center of the Government of the Canary Islands, created and regulated by Law 1/1991 of January 28, Statistics of the Autonomous Community of the Canary Islands.

77 958	http://dcat-ap.de/def/licenses/cc-by/4.0	German language translation of the CC BY 4.0 licence mentioned above (which is identical in substance; only the language differs). Open licence, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
77 391	https://creativecommons.org/licenses/by/4.0/deed.de	German language translation of the CC BY 4.0 licence mentioned above (which is identical in substance; only the language differs). Open licence as defined above, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
71 673	https://op.europa.eu/en/web/eu-vocabularies/concept/-/resource?uri=http://publications.europa.eu/resource/authority/licence/CC_BY_4_0	Standardised EU-level reference to the CC BY 4.0 licence mentioned above (which is identical in substance; only the language differs). Open licence as defined above, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
54 976	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/conditionsUnknown	Data error – no structured usage conditions.
44 241	Data licence Germany – Zero – Version 2.0 (DL-DE-Zero-2.0) http://dcat-ap.de/def/licenses/dl-zero-de/2.0	German licence (created by GovData). Open licence, granting all rights unconditionally.
37 012	Norway Digital Licence https://www.geonorge.no/Geodataarbeid/geografisk-infrastruktur/Norge-digitalt/Avtaler-og-maler/Norge-digitalt-lisens/	Norwegian data licence, specifically created to enable open reuse of geographic data. Allows internal use of the data only, and requires attribution. Explicitly references applicable Norwegian legislation.
36 961	Open Data Commons Open Database Licence (ODbL) https://opendatacommons.org/licenses/odbl/summary/	Data licence maintained by the Open Knowledge Foundation . Open licence, containing an attribution obligation and a share-alike obligation (requiring data to be made available under this licence, even when integrated into an application or service).
32 412	https://www.etalab.gouv.fr/wp-content/uploads/2014/05/Licence_Ouverte.pdf	Identical to the French licence mentioned above; the only distinction is the URL (referencing the web page, respectively a PDF version). Open licence, with an attribution requirement (data provider, data source, date of latest update). References data protection, and affirms compatibility with CC BY 4.0 and the OGL (among others). Explicitly references applicable French legislation.
32 021	http://creativecommons.org/publicdomain/zero/1.0/deed.nl	Dutch language translation of the Creative Commons Zero (CC0) licence (identical in substance; only the language differs).
30 602	http://creativecommons.org/publicdomain/zero/1.0/	English language translation of the CC0 licence (identical in substance; only the language differs).
30 596	http://dcat-ap.ch/vocabulary/licenses/terms_by	Swiss licence (created by opendata.swiss , the Swiss open data portal). Open licence with an attribution requirement only.
29 512	http://www.opendefinition.org/licenses/cc-by	Identical to the CC BY 4.0; it is referenced separately because a different identifier is used as the title in the dataset.

27 895	https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/	Same as the OGL: UK licence (created by the UK's National Archives, but used more broadly in the UK public sector). Open licence, with an attribution requirement (data provider, data source, licence). Explicitly states what is not covered by the licence (personal data, trademarks, identity documents, etc.).
27 029	Licence of the Geological and Mining Institute of Spain http://www.igme.es/Ayuda/CondiUso.htm	Spanish data licence, specifically created to enable open reuse of geographic data. Allows use of the data with attribution and notice. Explicitly references applicable Spanish legislation. NB: the URL is incorrect; the licence is available at https://info.igme.es/media/Pdfs/LicUsoIGME_GENERICA_2022.pdf .
24 597	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/noConditionsApply	Inspire (Infrastructure for Spatial Information in Europe) licence (geographical information), indicating that no conditions apply (a 'licence' in the broadest sense of the word).
24 392	http://data.vlaanderen.be/id/licentie/moделlicentie-gratis-hergebruik/v1.0	Belgian (Flemish Region) licence, created by the Flemish legislator. Open licence, with an attribution requirement (as indicated by the data provider; or otherwise data provider and year of use; with an alternative if cumulative references are not practical). Explicitly references applicable Flemish legislation.
Total number of licences counted: 1 814 364		

It should be acknowledged that there is a country bias in the analysis: since the selection criterion is based only on the frequency of use of a specific licence across the data portal, the sample will inherently favour Member States that provide a high number of datasets under a single licence – i.e. there is a bias towards Member States that provide large numbers of datasets (generally larger Member States) and/or that consistently opt for only a small number of licences. In simpler terms: the only **national/regional licences in this table are those from France, Germany, Spain and the United Kingdom** as larger EU Member States (or in the case of the United Kingdom, former Member State), **and Belgium (Flemish Region), Norway and Switzerland** as consistent users of a single regional licence.

As indicated, the list above shows that there are **1 814 364 licences counted**, across **22 named licences**. That number may seem intuitively impossible, given that we indicated above that there are only 1.2 million datasets for which licence information is available. However, it should be noted that a single dataset can be linked to multiple named licences – either because it is available under multiple licences, or because a single licence is listed under two or more names. By way of example, the table above shows the licence name 'Creative Commons Attribution 4.0 International – <https://creativecommons.org/licenses/by/4.0/>' and the licence name '<http://dcat-ap.de/def/licenses/cc-by/4.0>'. In reality, both of these refer to the exact same licence, but point to two different locations where the licence text can be found.

Moreover, as the table shows, there are **several data errors** in the table, at least for the purposes of this report.

- **One** of the licence names is shown as '<http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/conditionsUnknown>'. As the name clearly suggests, this is **not an actual licence**, but rather indicates datasets for which the terms of use are unknown. It should thus be eliminated from our analysis.

- There are no less than **five** separate references to the **CC BY 4.0** licence (under the names <https://creativecommons.org/licenses/by/4.0/>, <http://dcat-ap.de/def/licenses/cc-by/4.0>, <https://creativecommons.org/licenses/by/4.0/deed.de>, https://op.europa.eu/en/web/eu-vocabularies/concept/-/resource?uri=http://publications.europa.eu/resource/authority/licence/CC_BY_4_0, and <http://www.opendefinition.org/licenses/cc-by>). These are all merely different ways to reference the same licence. Thus, in reality, **CC BY 4.0 is the most commonly referenced licence on the portal**. Adding up the numbers of the five names under which this licence is referenced, there is a total count of **521 032 references** – more than a **quarter** of the references on the portal. As indicated above, this does not necessarily show that a quarter of the datasets on the portal are available under the CC BY 4.0 licence, since a single dataset might use multiple licence names.
- Similarly, there are **two** separate references to the **CC0** licence (under the names <http://creativecommons.org/publicdomain/zero/1.0/> and <http://creativecommons.org/publicdomain/zero/1.0/deed.nl>). Adding up the numbers for both of these names, **references to the CC0 licence are included 62 623 times**.
- There are also **two** separate references to the **French Open Licence** (under the names <https://www.etalab.gouv.fr/licence-ouverte-open-licence> and https://www.etalab.gouv.fr/wp-content/uploads/2014/05/Licence_Ouverte.pdf). Again, this is a parsing issue, with the first reference pointing to a web version of the licence terms (in English and French), and the second to a PDF version in French only. Substantively they are identical, so the **French licence actually is referenced 221 062 times**, making it the second most popular national licence (after the ‘Data licence Germany – attribution – Version 2.0’).
- Finally, there are **two** separate references to the **OGL** (under the names <http://reference.data.gov.uk/id/open-government-licence>, and <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>), which are also substantively identical. **The OGL is therefore actually referenced 143 138 times**.

Thus, a cleaned up and simplified table of the main licences on the portal would look as follows.

Licence count	usage	Licence name
521 032		Creative Commons Attribution 4.0 International
278 958		Data licence Germany – attribution – Version 2.0
221 062		Etalab Licence Ouverte / Open Licence
200 932		Licence of the Spanish Statistical Office
143 138		Open Government Licence 3.0
106 815		Licence of the Canarian Institute of Statistics
62 623		Creative Commons Zero
44 241		Data licence Germany – Zero – Version 2.0
37 012		Norway Digital Licence
36 961		Open Data Commons Open Database Licence (ODbL)
30 596		Swiss data licence
27 029		Licence of the Geological and Mining Institute of Spain
24 597		Inspire licence without conditions
24 392		Flemish attribution licence
Total number of licences usages counted: 1 759 388		

Based on these 14 licences, several interesting observations can be made.

- **Firstly, 10 out of the 14 relate to licences used only in a single European national or subnational territory (national or regional):** 3 from Spain, 2 from Germany and 1 each from Belgium (Flemish Region), France, Norway, Switzerland and the United Kingdom.
- **Only 1 out of the 14 relates to an EU-level licence,** specifically for the Inspire datasets. This is a particular case, since the Inspire statement is only a licence in the broadest sense – it is actually a statement that no conditions apply).
- **The remaining 3 out of the 14 relate to two [Creative Commons licences](#) (CC BY 4.0 and CC0), and the ODbL.**

Generally, based on an examination of these most popular licences on the data.europa.eu portal, the **Creative Commons licences (notably CC BY 4.0) are the only cross-border licence templates in common use today**, with the ODbL taking third place. Moreover, the **CC BY 4.0 licence is by some margin the most popular licence on the data.europa.eu portal**, accounting for 521 032 references (roughly as much as the second and third most popular licences added together).

This summary provides a detailed picture of the licences most commonly in use today. In the following section, we will examine each of these 14 licences in more detail, mapping them to the characteristics discussed above.

Zooming in on the most common licences

Licence groups

To facilitate the reading, the 14 licences identified above will be discussed in three groups:

- the international group, comprising the CC BY 4.0 and CC0 licences, and the ODbL;
- the national and regional group, comprising the eight licences mentioned above;
- the EU group, comprising Inspire.

For each licence, we will very briefly describe its origins, its characteristics based on the standard characteristics described above and the impacts (pros and cons) of using this licence in practice. In this way, we can provide guidance to data reusers in assessing and understanding these licences, and to data providers in deciding when to choose a particular licence.

International group

This group consists of three licences: the CC BY 4.0 licence, the CC0 licence and the ODbL.

Name and source – Creative Commons Attribution 4.0 International licence (CC BY 4.0) https://creativecommons.org/licenses/by/4.0/deed.en	
Origins	The Creative Commons licences (often abbreviated as ‘CC licences’) are established and maintained by Creative Commons , an international non-profit-making organisation established in 2001, which is dedicated to the creation of tools that ‘ enable people to grow and sustain the thriving commons of shared knowledge and culture ’. One of these tools is the Creative Commons licence family , containing six different licences (including the CC BY 4.0 licence) and one public domain dedication tool (the CC0 licence). The licences were first made available to the public in 2002, but have been revised repeatedly since that time.
Characteristics	The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicensing. It has notice and attribution obligations, and an obligation to state changes made. It does not prohibit commercial use. It does not contain patent grants, but does cover trademarks and database rights.
Impacts	As noted in the overview above, the CC BY 4.0 licence is highly popular, since it is relatively accessible: all CC licences are published accompanied by a so-called ‘deed’, a summary that can be more easily understood by non-legal professionals. The licence is well known internationally,

	<p>and tailored to the EU market (as witnessed by the coverage of database rights and the availability of official translations in multiple EU languages).</p> <p>Moreover, CC BY 4.0 is a fairly balanced licence, since it principally requires standardised attribution of the source, and a reference to the applicability of the licence itself. No other constraints apply, thus granting fairly broad flexibility to aspiring reusers.</p> <p>The principal benefits are undoubtedly recognition and uptake: the licence is well known internationally, thus facilitating use and reducing cross border compatibility concerns.</p>
--	--

Name and source – Creative Commons Public Domain Dedication (CC0) https://creativecommons.org/publicdomain/zero/1.0/	
Origins	<p>The CC licences are established and maintained by Creative Commons, an international non-profit-making organisation established in 2001, which is dedicated to the creation of tools that ‘enable people to grow and sustain the thriving commons of shared knowledge and culture’. One of these tools is the Creative Commons licence family, containing six different licences (including the CC BY 4.0 licence) and one public domain dedication tool (the CC0 licence). The licences were first made available to the public in 2002, but have been revised repeatedly since that time.</p>
Characteristics	<p>The licence grants all possible permissions.</p> <p>It contains no obligations.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants, but does cover trademarks and database rights.</p>
Impacts	<p>CC0 is not technically a licence, but rather a statement (a ‘dedication’) from the data provider indicating that it waives, to the maximum possible extent of the law, all of its rights to the work worldwide under copyright law, including all related rights such as database rights.</p> <p>As noted in the overview above, the CC0 licence is popular, due to its extreme simplicity: as with other CC licences, it is published accompanied by a ‘deed’, a summary that can be more easily understood by non-legal professionals. The licence is well known internationally, and tailored to the EU market (as witnessed by the coverage of database rights and the availability of official translations in multiple EU languages).</p> <p>CC0 is highly recommended when data providers wish to relinquish their claims to the data to the maximum extent possible, with the sole caution that this also includes a waiver to claim attribution.</p> <p>The principal benefits are simplicity, recognition and uptake: the licence is intuitive and well known internationally, thus facilitating use and reducing cross border compatibility concerns.</p>

Name and source – Open Data Commons Open Database Licence (ODbL) https://opendatacommons.org/licenses/odbl/summary/	
Origins	<p>The ODbL was created and is maintained by the Open Knowledge Foundation, a non-profit-making organisation. Unlike the aforementioned CC licences, the ODbL is specifically created to focus on databases (whereas the CC licences can be used for any work).</p> <p>It is an open licence, containing an attribution and notice obligation, and a share-alike obligation (in case of public use of any adapted version of the database, or works produced from an adapted database, you must also offer that adapted database under the ODbL, or under a compatible licence).</p>
Characteristics	<p>The licence grants all possible permissions.</p> <p>It has notice and attribution obligations, and an obligation to state changes made.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants or trademarks, but does cover database rights.</p>
Impacts	<p>As the overview above shows, the ODbL is the third most popular international licence (after both of the aforementioned CC licences). They explicitly cover database rights by name, thus being well suited to the EU context where such rights are recognised. Like the CC licences, the ODbL too is published accompanied by a summary that can be more easily understood by non-legal professionals.</p> <p>The ODbL is somewhat more demanding than the otherwise comparable CC BY 4.0 licence, due to its share-alike approach, requiring the ODbL (or a compatible licence) to also be applied to derivative works.</p> <p>The principal benefits are its accessibility and uptake: the licence is known internationally, thus facilitating use and reducing cross border compatibility concerns. The principal constraint is the share-alike approach, which introduces a small additional layer of complexity.</p>

National and regional group

The national and regional group comprises ten licences from seven European nations or regions: 3 from Spain, 2 from Germany and 1 each from Belgium (Flemish Region), France, Norway, Switzerland and the United Kingdom. They will be discussed hereunder.

Name and source – Licence of the Spanish Statistical Office (<i>Instituto Nacional de Estadística</i>)	
https://www.ine.es/dyngs/AYU/en/index.htm?cid=125	
Origins	<p>The general approach of the licence is to declare the application of CC BY 4.0, unless otherwise indicated on the website. It specifies the manner in which the attribution must be made to the INE website, and clarifies the responsibilities and liabilities of INE, along with a statement on the applicability of Spanish public sector information reuse law.</p> <p>The official terms are available in English and Spanish.</p>
Characteristics	<p>The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicencing.</p> <p>It has notice and attribution obligations, and an obligation to state changes made.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants, but does cover trademarks and database rights.</p>
Impacts	<p>Since the general approach is based on CC BY 4.0, the impacts will generally be the same as those of the CC BY 4.0 licence. As noted however, the right is reserved to impose additional licencing terms, and the application of Spanish law is confirmed. With this in mind, the approach of the INE can best be implemented by directly adopting the CC BY 4.0 licence.</p>

Name and source – Licence of the Canarian Institute of Statistics (*Instituto Canario de Estadística*)

http://www.gobiernodecanarias.org/istac/aviso_legal.html

Origins

The licence was developed to support the dissemination activities of the Canarian Institute of Statistics. However, it should not be understood as a single licence applicable to all datasets; rather, it references general terms of use for public sector information under Spanish law, and notes that such use is ‘subject to the conditions established in standard licences’. Thus, it is not so much a licence, as an anchor point to determine usage rights.

The central usage rights described in these terms will be summarised in the following characteristics section, but as noted, additional terms may apply. Moreover, the website also contains additional high-level specific conditions of use of the institute’s application programming interfaces, widgets and connectors, and a specific ISTAC-bot.

The official terms are available in Spanish only.

Characteristics

The licence grants all possible permissions.

It contains attribution obligations, using a specific phrasing imposed by the licence.

It does not prohibit commercial use.

It does not contain patent grants and does not cover trademarks or database rights.

Impacts

As stressed above, the statement is not necessarily a comprehensive licence, but rather a general intellectual property rights policy statement that can explicitly be amended by additional terms. It is designed to support the functioning of the Canarian Institute of Statistics, and focuses on this context. From a cross-border reuse perspective, the approach may be difficult to work with, since the source material is only available in Spanish, references additional possible constraints that can vary depending on the dataset, and is subject to Spanish law. For that reason, it may not be easily used in other contexts.

Name and source – Licence of the Geological and Mining Institute of Spain (*Instituto Geológico y Minero de España*)

https://info.igme.es/media/Pdfs/LicUsoIGME_GENERICA_2022.pdf

Origins	<p>The Geological and Mining Institute of Spain is an autonomous public sector body, tasked with providing support to public administrations, specifically by providing the necessary knowledge and information in relation to geological sciences and technologies for any activity on the territory.</p> <p>It applies a bespoke licence to support its knowledge dissemination tasks, which applies to all materials on its website, including its datasets. It is an open licence, permitting commercial and non-commercial use, including the reuse and dissemination of its data. It does require that applicable Spanish law is respected, and imposes a specific attribution requirement of the source, and of the date of the latest update of the source material. Finally, it notes that additional constraints may apply if they are communicated by specific datasets.</p> <p>The official terms are available in Spanish only.</p>
Characteristics	<p>The licence grants all possible permissions.</p> <p>It contains attribution obligations, using a specific phrasing imposed by the licence, and an obligation to indicate the date of the last update of the source material, if this information is available.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	<p>The licence (labelled ‘Terms of Use’) applies horizontally for all materials made available by the institute (datasets, but also texts, images and other materials on its website), and its provisions are thus relatively flexible and high level. This is a benefit since they are easy to understand; but the downside is that the licence repeatedly indicates that additional conditions may apply, depending on the material being reused.</p> <p>As with the licence above, from a cross-border reuse perspective, the approach may be difficult to work with, since the source material is only available in Spanish, references additional possible constraints that can vary depending on the dataset, and is subject to Spanish law. For that reason, it may not be easily used in other contexts.</p>

Name and source – Data licence Germany – attribution – Version 2.0	
https://www.govdata.de/dl-de/by-2-0	
Origins	<p>This licence was created by GovData, the German open data portal, and was intended to be tailored to the German legal environment.</p> <p>It was established in the context of the German Open Government project, aiming to promote the use of fewer, simpler and uniform terms of use. It was developed in cooperation with the federal government, the German states (<i>Länder</i>) and municipal associations.</p> <p>The licence family is referred to as the ‘Data licence Germany’ family, which is available in two variants: an ‘attribution’ variant which obliges the data user to name the data provider, and a ‘Zero’ variant which enables unrestricted further use. In practice, they can be considered as the German equivalents of CC BY 4.0 and CC0, respectively.</p> <p>Both of these ‘Data Licence Germany’ licences are commonly used, since there is a closed list of licences that are permitted for delivery to GovData, and since GovData recommends using the Data licence Germany 2.0 (Zero or attribution) or CC BY 4.0.</p>
Characteristics	<p>The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicencing.</p> <p>It has notice and attribution obligations, and an obligation to state changes made.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	<p>As noted in the overview above, the German attribution licence is highly popular, surpassed only by CC BY 4.0 on the data.europa.eu portal in terms of usage. It is available both in German and in English, which can increase its popular appeal.</p> <p>It is a fairly balanced licence, since it principally requires standardised attribution of the source, and a reference to the applicability of the licence itself. No other constraints apply, thus granting fairly broad flexibility to aspiring reusers. It does formalise the attribution obligation to some extent, by prescribing the use of a source note containing the name of the provider, the annotation ‘Data licence Germany – attribution – Version 2.0’ or ‘dl-de/by-2-0’ and a reference to the online location (the URI) where the dataset can be found. Changes, editing, new designs or other amendments must be marked as such in the source note. This approach is arguably quite prescriptive, and may be less intuitive to non-German reusers.</p>

Name and source – Data licence Germany – Zero – Version 2.0

<https://www.govdata.de/dl-de/zero-2-0>

Origins	<p>This licence was created by GovData, the German open data portal, and was intended to be tailored to the German legal environment.</p> <p>It was established in the context of the German ‘Open Government’ project, aiming to promote the use of fewer, simpler and uniform terms of use. It was developed in cooperation with the federal government, the German states (<i>Länder</i>) and municipal associations.</p> <p>The licence family is referred to as the ‘Data licence Germany’ family, which is available in two variants: an ‘attribution’ variant which obliges the data user to name the data provider; and a ‘Zero’ variant enables unrestricted further use. In practice, they can be considered as the German equivalents of CC BY 4.0 and CC0, respectively.</p> <p>Both of these ‘Data licence Germany’ licences are commonly used, since there is a closed list of licences that are permitted for delivery to GovData, and since GovData recommends using the Data licence Germany 2.0 (Zero or attribution) or CC BY 4.0.</p>
Characteristics	<p>The licence grants all possible permissions.</p> <p>It contains no obligations.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	<p>As can be observed in the overview above, the DL-DE-Zero-2.0 licence is fairly popular, being more or less used as frequently as CC0. It is available both in German and in English, which can increase its popular appeal.</p> <p>Like CC0, the appeal undoubtedly stems from its extreme simplicity: the licence consists of just five short sentences. DL-DE-Zero-2.0 is highly recommended when data providers wish to relinquish their claim over data to the maximum extent possible, with the sole caution that this also includes a waiver to claim attribution.</p> <p>The principal benefit is its simplicity; the principal downside is that it is more or less equivalent to CC0, which is better known internationally.</p>

Name and source – Flemish licence for free reuse 1.0

<http://data.vlaanderen.be/id/licentie/modellicentie-gratis-hergebruik/v1.0>

Origins	<p>The licence was created by the region of Flanders in Belgium, specifically by the Flemish legislator. The laws promote the use of either CC0, the Flemish licence for free reuse (which is the present licence), or a less commonly used licence for reuse against a fee.</p> <p>The Flemish licence for free reuse allows any commercial or non-commercial reuse, free of charge, with an attribution requirement. The attribution must be indicated by the data provider; or otherwise the reuser must by default identify the data provider and year of use; with an alternative if cumulative references are not practical. The licence contains multiple explicit references to applicable Flemish legislation and is only available in Dutch.</p>
Characteristics	<p>The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicensing.</p> <p>It has notice and attribution obligations, but no obligation to state changes made.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	<p>The licence is flexible in use, mainly having an explicit and fairly detailed attribution requirement. However, its use may negatively impact cross border usability of a dataset, since it is officially only available in Dutch, and contains multiple references to Flemish administrative law, which are not likely to be easily understandable outside of the region.</p>

Name and source – French Open Licence

<https://www.etalab.gouv.fr/wp-content/uploads/2018/11/open-licence.pdf>

Origins	<p>This licence was created by Etalab, a department of the Interministerial Digital Directorate, but is promoted and used more broadly in the French public sector. The French Open Licence has an attribution requirement for the identification of a data provider, data source and the date of the latest update of the dataset. The licence also references data protection compliance obligations, and affirms its compatibility with CC BY 4.0 and OGL (among others). It also explicitly references applicable French legislation.</p>
Characteristics	<p>The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicensing.</p> <p>It has attribution obligations, but no obligation to state changes made (other than the date of the last update of the data, which seems to refer to updates from the data provider, not from the data reuser).</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks, but it does cover database rights.</p>
Impacts	<p>The licence is flexible in use, mainly having an explicit but fairly simple and intuitive attribution requirement. It is available in French and English, and includes a standardised logo, which can facilitate its use. It also appears to be relatively commonly used, since it is the second most common national licence. Moreover, it clearly considers the EU context, based on its references to data protection law and database rights.</p> <p>However, its use may be less appropriate for non-French data providers, since it contains multiple references to French administrative law, which are not likely to be easily understandable outside of France.</p>

Name and source – Norway Digital Licence

<https://www.geonorge.no/Geodataarbeid/Norge-digitalt/Avtaler-og-maler/Norge-digitalt-lisens/>

Origins	<p>The licence is used by Geonorge, the national Norwegian website for geographic information. The website was created in the context of <i>Norge digital</i> (Digital Norway), a partnership involving public enterprises that are responsible for establishing and managing map data and other location information. Geonorge is developed and run by Kartverket, the Norwegian Mapping Authority, on behalf of members of the Norge digital partnership.</p> <p>The licence is short, and grants usage rights to members of Digital Norway, and to businesses in the EU and the European Economic Area, in accordance with Norwegian law (the Geodata Act). The right of use applies only to the parties' employees and contractors, and includes analogue and digital copying, processing, value addition and use in internal networks and services.</p> <p>The data may be integrated in external services, but these must then use up to date data, and be free of charge in principle. There is also an attribution requirement.</p> <p>The licence is only available in Norwegian.</p>
Characteristics	<p>The licence grants permissions to reproduce and distribute the data, subject to constraints (obligation to use updated data, and no commercial exploitation in principle).</p> <p>It has attribution obligations (name of source/agency is required, name and version of dataset / data source is recommended).</p> <p>It prohibits commercial use (or more accurately charging).</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	<p>The licence is very contextual: beyond being only available in Norwegian and referencing the Norwegian Geodata Act repeatedly, it has constraints on the use of the data and on charging. It was designed to support the activities of Geonorge, but is not easy to apply in cross-border situations, due to its strong national focus.</p>

Name and source – Swiss attribution licence

<https://opendata.swiss/en/terms-of-use>

Origins	Following the Creative Commons approach, the Swiss open data portal opendata.swiss has created a family of licences, including the Swiss attribution licence, which is highly comparable to CC BY 4.0, but significantly simpler in its phrasing and approach, making them more accessible to understand.
Characteristics	<p>The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicensing.</p> <p>It has basic attribution obligations (author, title and link to the dataset), but no obligation to state changes made.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	The licence is very easy to use, since it is very short and phrased in an accessible manner. The sole obligation is a pragmatic attribution obligation. It is available in English, German, French and Italian, and includes a standardised logo, which can facilitate its use.

Name and source – Open Government Licence 3.0

<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Origins

The OGL was created by the UK's National Archives, but is used more broadly in the UK public sector. It is a part of the UK Government Licencing Framework, which provides a policy and legal overview of the arrangements for licencing the use and reuse of public sector information, both in central government and the wider public sector.

The framework mandates the OGL as the default licence for Crown bodies and recommends OGL for other public sector bodies.

The OGL is an open licence, and explicitly states what topics are not covered by the licence (personal data, trademarks, identity documents, etc.).

Characteristics

The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicencing.

It has attribution and notice obligations (data provider, data source, licence), but no obligation to state changes made.

It does not prohibit commercial use.

It does not contain patent grants or cover trademarks, but it does cover database rights (specifically 'Crown database rights').

Impacts

The licence is relatively short and accessible, and thus easy to use. The sole obligation is a pragmatic attribution obligation. However, it does reference UK law (including the application of 'Crown copyright and Crown database rights'), which may complicate cross-border application of the licence.

EU sectoral group – Inspire

Finally, there is one licence included in the EU sectoral group: the Inspire licence. Within the portal, the licence consists only of a brief statement that no conditions apply, as we will discuss below.

Name and source – Inspire licence without conditions	
http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/noConditionsApply	
Origins	<p>Inspire stands for ‘Infrastructure for Spatial Information in Europe’. The Inspire directive was initially adopted in 2007, and aimed to create a European Union spatial data infrastructure to support EU environmental policies. The spatial data infrastructure should thus facilitate the sharing of environmental spatial information among public sector organisations, facilitate public access to spatial information across Europe and assist in policymaking across boundaries.</p> <p>Inspire datasets should thus be made available in a manner that facilitates exchange and reuse, as a matter of principle.</p> <p>The ‘Inspire licence without conditions’ is the most extreme form of this philosophy, since it contains only the statement that no conditions apply for access and use. In that sense, it can be seen as a ‘CC0 for Inspire data’.</p>
Characteristics	<p>The licence grants permissions to reproduce and distribute works and create derivatives. It allows sublicencing.</p> <p>It has no obligations.</p> <p>It does not prohibit commercial use.</p> <p>It does not contain patent grants and does not cover trademarks or database rights.</p>
Impacts	<p>The ‘no conditions’ statement is of course very pragmatic to use, since it essentially communicates that no constraints apply. It leaves perfect freedom to reusers, with the sole limitation being that it is linked to Inspire data.</p>

4. Recommendations for data providers

Observations on key trends

The overview in this report is quite extensive and highlights a lot of facets and details with respect to licences applied in the data.europa.eu portal and the role of the Licencing Assistant. Nonetheless, a few key messages can be marked as noteworthy.

Firstly, it is interesting to observe that, despite the vast breadth of the datasets on the portal (more than 1.8 million datasets), only around 400 licences are commonly referenced. The existence of 400 different licences is of course still a very daunting prospect – very few people would have the time and inclination to familiarise themselves with all of them. But as the analysis in Section 3 showed, after filtering out data errors, in practice a fairly compact group of 14 licences is used by the majority of the datasets.

Secondly, when looking at those 14 licences, two clear observations stand out.

- **The only two licencing approaches commonly used today are a comprehensive waiver of rights (a ‘CC0-style’ approach), and an attribution requirement (a ‘CC BY 4.0-style’ approach).** All 14 of the common licences meet this criterion: either all rights are waived to the maximum possible extent in various ways or forms that are essentially equivalent from the perspective of the reuser, or an attribution of the source is requested.

The other limitations discussed in the section on characteristics of open licences (Section 2 of this report) in practice play no meaningful role: there is no significant usage of share-alike obligations (used in only 1 of the 14 licences, specifically the ODbL), copyleft obligations (not observed in the 14 licences) or prohibition of commercial exploitation (also used in only 1 licence, specifically the Norwegian Digital Licence). Thus, **there is a trend towards simplicity and openness**, which can be seen as very favourable.

- **The only commonly used international licences today are the CC licences, specifically CC0 and CC BY 4.0.** Looking at the 10 national/regional licences, all of these are national variants of these licences (with one being a literal reference to CC BY 4.0 with limited room for divergence). The principal differences are that:
 - where attribution is required, the approach is slightly different in the level of detail or in required elements to be mentioned;
 - language availability is more limited – 6 out of 10 offer English language versions of their licences, but apart from this, only national languages are supported.

National laws are referenced in 7 out of 10 licences. This means that, in practice, **the impact and appeal of these national and regional licences suffers from the lack of translations and international alignment.** They are intended to support and facilitate national reuse and, given their adoption rates (based on the available statistics), they might well achieve this goal, albeit at the cost of complicating cross-border reuse.



■ Publications Office
of the European Union