Licence Compatibility in Europe: A winding road to Creative Commons

A short exploration of legal issues, current trends and the practical reality for data providers and re-users in Europe

data.europa.eu
The official portal for European data

Introduction to licence compatibility

A primer on the role of licences for data providers and data re-users, and why licence compatibility matters

A central objective of European open data policy is ensuring that the potential value of data, including data made available via open data portals, can be optimally realised. Reducing or eliminating any constraints that impede open use of such data is a critical step in achieving this objective: 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'.

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. 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 denotes. While there is no official EU-level definition of the requirements for such a licence, common elements include 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.

The use of licences, however, can trigger a problem that is the focus of this research paper. When an aspiring re-user wants to combine multiple datasets into a single new product or service (e.g. to create a new enriched dataset, or to create a service or application that uses multiple datasets as its source), they will have to take into account all applicable licences. This is not unduly complex when all datasets use the same licence: the user will of course still have to evaluate that single licence and reflect on how its conditions can be met (e.g. how the source should be attributed), but this is usually a relatively simple exercise.

More challenging problems can however arise when the different datasets apply different licences. In that case, the aspiring re-user will have to evaluate them individually, and see whether and how it is possible to adhere to all of the different conditions cumulatively. This requires more significant legal expertise and resources, since the licences have to be compared with each other to examine their main characteristics and key features.

This question is referred to as 'licence compatibility': is it legally possible to combine two or more datasets that are covered by several licences, and to use, distribute or otherwise exploit the result?

Working definitions in this paper

This paper relies on a few important concepts that do not always have a clear or universally accepted meaning. To minimise issues of interpretation, the following working definitions are used.

Concept	Working definition
Dataset	A single logical group of data with a common structure, metadata and scope, made available as a downloadable file or via specific application programming interfaces.
Derivate work	An output (such as a new or combined dataset, or software) which was created by using the inputs of one or more datasets.
Licence	A document stating the terms under which a dataset is made available. While it is possible for a dataset to be made available without a licence, the implication is that the terms of use are unclear.
Attribution	A legal obligation to credit the source of a dataset. The manner and extent to which the source has to be credited is determined by the licence.
Distribution	The action of making a dataset available to third parties. Using a dataset for your own internal purposes (without making it available to third parties) thus does not imply any distribution.
Open data	A dataset that is made available for any lawful use with only minimal or no legal, technical or financial constraints.
Permissive licence	A licence that grants the right to access, reuse and redistribute a dataset, free of charge, without any limitation on use cases (commercial or non-commercial), and without limitation on the potential beneficiaries of the licence. Attribution may be included under a permissive licence.
Share-alike	A legal obligation to apply the same licence terms to a derivative work as were applied to the dataset(s) on which the derivate work was based. The licence to be used for the derivative work is determined by the licence.
Licence compatibility	The legal possibility to lawfully create a derivative work based on at least two datasets, in a manner that does not violate the licence under which those datasets were made available.

Problem statement and structure of this research paper

When a user wants to create a derivative work (as defined above, e.g. a new dataset or an application) on the basis of at least two datasets, they will in principle need to assess the licence compatibility of all of the datasets that they want to use. Conceptually, this exercise can have only two outcomes.

- The first possibility is that the licences are **compatible**, i.e. their conditions can be respected cumulatively. In this case, the re-user will still need to check how the conditions of the compatible licences can be met: compatibility does not mean that the conditions of the licences are identical, but only that it is possible to find a way to respect all of the conditions.

  - This task can be **relatively simple** in some instances, notably when only cumulative attribution is required. In this case, the re-user must determine which multiple attributions are required when releasing a new work, without needing to impose other limitations. This can be somewhat burdensome when the requirements are phrased differently – e.g. one licence requires the original administration to be identified; another also requires that the URL where the data is found is referenced; and a third imposes the inclusion of the date of the last update – but this is mainly a matter of administrative housekeeping.

  - The task can become **very complex** when multiple other limitations from the original licences must be 'stacked' on top of each other, e.g. by forbidding commercial use, limiting use to a specific sector or use case, forbidding further dissemination via sublicensing, forbidding changes in the data, or any other constraints that may be included in one of the original licences. In this case, the work is legally challenging, and can also be risky due to the need to make difficult judgments calls on whether a specific limitation in one licence is equivalent to a slightly different limitation in another licence. By way of example: if one dataset only allows use 'for private purposes', and another allows only 'non-commercial use', then can these be combined in a work that only allows use 'for private purposes', i.e. would that automatically exclude all commercial use, or does a re-user need to consider the possibility of private for-profit use? Opinions may reasonably vary.

  An inevitable side effect of compatibility assessment is also that, even if licences are found to be compatible, the strictest requirements must inevitably apply: if one dataset requires a specific attribution, one forbids commercial use and another forbids modification, then the only solution when combining them into a new product is to release a product that cumulatively contains the attribution, and forbids commercial use and modification. In that sense, even positive compatibility assessments (i.e. a finding that multiple licences are indeed compatible) result in increasingly larger limitations on the openness of datasets, since all constraints must be added up. This outcome is not beneficial from a public policy viewpoint, considering the current emphasis on enabling and facilitating reuse wherever possible.

- The second possibility of course is that the licences are **incompatible**, i.e. their conditions cannot be respected cumulatively. The most obvious example is the combination of two datasets with two different licences that each contain a strict share-alike condition, i.e. both licences allow only their own terms to be applied to a product that contains both datasets. In this case, such a product cannot be lawfully released.

Licence compatibility considerations can thus be a significant barrier to the reuse of multiple datasets that are made available under different licences, and in that sense the need for compatibility assessments is a barrier to achieving the EU's policy objectives with respect to open data. Some tools are available to mitigate the problem, such as the [Licensing Assistant](#) on data.europa.eu, but they too are ultimately not able to automatically determine how licences can be cumulatively adhered to; this remains a manual job that requires specific expertise and resources.

It is however not very clear whether this problem occurs in practice, or whether it is instead a largely theoretical issue. With that in mind, this research paper aims to examine the current licensing landscape in Europe, and to verify via stakeholder interviews whether licence compatibility problems occur, and how they are managed.

This legal research paper is **intended as a resource for data holders, data users and data policy makers**.

- **Data holders** can learn what licence compatibility challenges are, what licensing choices are being made in Europe, and what the impacts of these choices on data users are.
- **Data users** can learn how licence compatibility challenges can affect them, and which licences can cause difficulties in practice.
- **Policy makers** can learn which licensing choices minimise licence compatibility problems, thus contributing to an effective open data ecosystem.

The paper aims to answer three research questions, each tackling a different perspective of the topic:

1. **Empirically**, the paper will examine what the licence compatibility concerns are right now between data holders (if any). This will be done by examining the most common licences currently being used, based on an analysis of licences proposed by data providers via [data.europa.eu](#).

2. **Conceptually**, the paper will examine to what extent these compatibility concerns affect the possibility of combining datasets in practice, by talking to representatives of data providers about the problems they encounter (if any). In this way, it will explore to what extent the concerns affect data re-users.

3. **Operationally**, the paper will provide recommendations that data providers (to be understood as the entities that determine the licence terms) can implement to reduce compatibility concerns.

As a result, this paper is structured into three major sections.

- Firstly, a **short empirical assessment** of the state of play is provided, describing the main licences being offered today.
- Secondly, a **reality check** is provided, based on five interviews with data providers, to determine whether they face concerns or receive questions in relation to licence compatibility.
- Thirdly, **conclusions** are provided, containing guidelines on how to mitigate licence compatibility problems in practice.

A short empirical assessment of the state of play

What are the most common licences in Europe today?

The first question is of course how broad the licensing landscape in Europe is today. Licence compatibility problems can only occur when multiple different licences must be applied cumulatively, so it is important to understand how varied the landscape is in practice. If, by way of example, a single licence applies to 95 % of the datasets being offered today, then the risk of incompatibility is much lower than in a landscape where there is much more fragmentation.

Moreover, it is worth assessing to what extent licences actually do contain specific requirements to be adhered to. If 95 % of licences are permissive as defined above – i.e. they grant a right to access, reuse and redistribute a dataset, free of charge, without any limitation on use cases or potential beneficiaries, and optionally only an attribution requirement – again the licence compatibility challenges should be limited.

The most comprehensive resource available today is the data.europa.eu portal itself, which contains (at the time of drafting) **1 552 471 datasets** [for which the licences can be examined](.).

Focusing only on licences that are used more than 5 000 times on the portal – assuming of course that the metadata on this topic is accurate – the following overview table of most common licences can be provided.

Licence usage count	Licence name as indicated on the portal (i.e. the formal name used in the metadata on the portal. When a URL is shown, the URL is the formal name)	Short description of the licence
60.201	Data licence Germany – attribution – Version 2.0	German licence (created by GovData, the German open data portal). Permissive licence as defined above, with an attribution requirement (data provider, data source and licence name). Changes must be notified.
59.465	Creative Commons Attribution 4.0 International	International licence. Permissive licence as defined above, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
42.084	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/conditionsUnknown	Data error – no structured usage conditions.
35.785	https://creativecommons.org/licenses/by/4.0/deed.de	German translation of the CC-BY International licence mentioned above (which is identical in substance; only the language differs). Permissive licence as defined above, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
27.053	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). Permissive licence as defined above, with an attribution requirement (data provider, data source, date of latest update). References data protection, and affirms compatibility with CC-BY and

		the open government licence (OGL) (among others). Explicitly references applicable French legislation.
25.340	Open Government Licence 3.0	UK licence (created by the United Kingdom's National Archives, but used more broadly in the UK public sector). Permissive licence as defined above, with an attribution requirement (data provider, data source, licence). Explicitly states what is not covered by the licence (personal data, trademarks, identity documents, etc).
19.865	Creative Commons Attribution 4.0 International	Identical to the second licence above; it is referenced separately because datasets in this category include the version number '4.0' in the licence name, and the second licence above does not (i.e. the separate inclusion is a metadata parsing issue).
18.545	Creative Commons Attribuzione 4.0 Internazionale (CC BY 4.0)	Italian translation of the CC-BY International licence mentioned above (which is identical in substance; only the language differs). Permissive licence as defined above, with an attribution requirement (data provider, licence name, no warranties). Changes must be notified.
15.976	https://www.ine.es/aviso_legal	Licence of the Spanish Statistical Office (*Instituto Nacional de Estadística*). The URL is not correct, but the licence is available at https://www.ine.es/dyngs/AYU/en/index.htm?cid=125. Permissive licence as defined above, with an attribution requirement (data portal, data provider, date of latest update).
14.075	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/noConditionsApply	Inspire licence (geographical information), indicating that no conditions apply (a 'licence' in the broadest sense of the word).
9.176	Creative Commons CC0 1.0 Universal	International permissive licence, indicating that all rights are waived to the maximum possible extent (a 'licence' in the broadest sense of the word).
9.152	Licence not specified	Data error – no structured usage conditions.
7.084	http://www.gobiernodecanarias.org/istac/aviso_legal.html	Licence of the Canarian Institute of Statistics (*Instituto Canario de Estadística*) ([1]). Permissive licence as defined above, with an attribution requirement (data portal, data provider, date of latest update).
6.935	http://creativecommons.org/licenses/by/4.0/deed.nl	Dutch translation of the CC-BY International licence mentioned above (which is identical in substance; only the language differs). Permissive licence as defined above, with an attribution requirement (data portal, date of last update, references to metadata and licence). Explicitly references applicable Dutch law, and excludes some use cases (e.g. relinking statistical information to individual units).
6.345	http://data.vlaanderen.be/id/licentie/modellicentie-gratis-hergebruik/v1.0	Belgian (Flemish region) licence, created by the Flemish legislator. Permissive licence as defined above, 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.

([1]) The Canarian Institute of Statistics 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, on Statistics of the Autonomous Community of the Canary Islands.

6.019	Data licence Germany – Zero – Version 2.0	German licence (created by GovData, the German open data portal). Permissive licence as defined above, granting all rights unconditionally.
5.695	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). Permissive licence as defined above, with an attribution requirement (data provider, data source, date of latest update). References data protection, and affirms compatibility with CC-BY and the OGL (among others). Explicitly references applicable French legislation.
5.255	http://creativecommons.org/publicdomain/zero/1.0/deed.nl	Dutch translation of the permissive CC0 licence mentioned above (which is identical in substance; only the language differs), indicating that all rights are waived to the maximum possible extent (a 'licence' in the broadest sense of the word).
Total number of licences counted: 374 050		

Table 1: Overview of most common licences on data.europa.eu

As shown, the list above covers **374 050 out of 1 552 471 datasets** on the portal – **approximately 24 % of the total.** It is thus clear that there is a 'long-tail' limitation in the analysis of the data: 76 % of the datasets are not covered by this report. These are datasets that use a licence which is relatively uncommon, i.e. used less than 5 000 times on the portal.

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 Member States** (or in the case of the United Kingdom, former Member State)), **and Belgium (Flemish Region)** as a consistent user of a single regional licence.

18 licences are included in the table. However, as the table shows, there are **four data errors** in the table, at least for the purposes of this report.

- The third largest set indicates as a licence name 'http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/conditionsUnknown'. As that name clearly suggests, this is not actually a licence, but rather indicates Inspire datasets (geographical data) for which the terms of use are unknown.
- Similarly, the 12th largest set indicates as a licence name 'licence not specified', reflecting an absence of information (rather than any actual licence being identified).
- There are two separate mentions of the Creative Commons Attribution 4.0 International licence (second and seventh largest sets). This is a simple parsing issue: the first reference does not include the version number 4.0 in the metadata, and the second does (although both actually refer to the same version 4.0). Thus, **in reality, Creative Commons Attribution 4.0 International is the most common licence on the portal, with a combined 79 330 references.**
- There are also two separate references to the French open licence (fifth and 18th largest sets). 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 applies to 32 748 datasets, making it the second most popular national licence (after 'Data licence Germany – attribution – Version 2.0').

Thus, in reality **14 licences are covered.**

On the basis of these 14 licences, several very interesting observations can be made.

- **Firstly, 7 out of the 14 relate to actual licences used only in a single European national or subnational territory (national or regional)**: two from Germany, two from Spain (both statistical offices), and one each from Belgium (Flemish region), France and the United Kingdom.
- **Only 1 out of the 14 relates to an EU-level licence,** specifically for the Inspire datasets. As noted in the table, this is only a licence in the broadest sense, since it is actually a statement that no conditions apply. No actual EU-level licence exists that is used for more than 5 000 datasets on the portal.
- **The remaining 6 out of the 14 all relate to Creative Commons licences** (often abbreviated as 'CC licences'), either directly by referencing the CC-BY licence (i.e. with an attribution requirement) or the CC0 licence (a rights waiver statement) or indirectly by referencing a translation (in Dutch, German or Italian) of these CC-BY or CC0 licences, which are however substantively identical to the original CC-BY and CC0 licences.

Generally, based on an examination of these most popular licences on the data.europa.eu portal, **CC licences are the only cross-border licence templates in common use today**. Moreover, the **CC-BY 4.0 licence is by a wide margin the most popular licence on data.europa.eu**, accounting for 79 330 direct references as calculated above, and another 61 265 via indirect references (translations of the CC-BY 4.0 licence).

These CC licences are established and maintained by Creative Commons, an international non-profit organization establish 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 CC licence family, containing six different licences (including the CC-BY licence mentioned above) and one public domain dedication tool (the CC0 licence mentioned above). The licences were first made available to the public in 2002, but have been revised repeatedly since then.

Over time, the CC licences have clearly seen significant uptake in the EU open data sphere as well, as **140 595 out of the 374 050 examined most common licences (approximately 38 % of the sample) are CC-BY 4.0 licences.** With another 14 431 licences being CC0 declarations, the Creative Commons family is the only common non-national option in use today.

This status is also reflected in EU-level legislation and policy. The European Commission adopted a decision in 2019 that designated two CC licences as the default for Commission documents, with the explicit objective of facilitating the reuse of Commission documents.

A further substantive observation can be made that applies to all of the examined licences (both for CC licences and for non-CC licences): **the only two major licensing approaches today are a comprehensive waiver of rights (a CC0-style approach), and an attribution requirement (a CC-BY-style approach). All common licences on the portal are therefore permissive licences as defined at the beginning of this report.**

Other constraints, such as a **limitation to non-commercial use or a share-alike requirements, are not observed in the most commonly used licences at all**. A quick check of metadata on all licences on the portal (not just the most common ones) confirms this impression: only 1 197 datasets have a 'non-commercial' limitation flag, and only 2 693 a 'share-alike' flag – a negligeable segment of the more than 1.5 million datasets. This has implications for the licence compatibility problem, as will be examined directly in the next section.

Possible conflicts between licences

The analysis above would suggest that licence compatibility may not be a big problem in practice. If the only two major licensing models are the complete waiver ('do as you will') and the attribution model ('acknowledge the source'), then compatibility checks would fall into the relatively simple category as described above. An aspiring re-user who wants to combine multiple datasets would mainly need to check whether an attribution requirement exists in their data sources, and if so, find a correct way to observe them. The task should not be neglected, but it is not hugely complex either.

None the less, there are two issues that should not be overlooked. Firstly of course, there is the issue of sample size: the analysis above examines roughly a quarter of the datasets on the Portal, but requirements may vary across the other three quarters.

Secondly and more importantly, the devil is as always in the details. The table above provided only a very brief summary of legal requirements, but even that short summary shows various challenges. Even when focusing only on attribution itself, there is a broad variation in the scope and phrasing. Depending on the licence, attribution can cover (non-exhaustively):

- 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;
- the name of the licence;
- a link to the licence;
- the date of retrieval of the data;
- 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;
- mandatory inclusion of the original metadata; and
- an obligation to repeat the claim that no warranties are made.

It is possible to combine all of these attribution requirements in a single work, but significant legal homework is imposed on a data re-user – homework which is made especially burdensome when licences are only made available in a language that may not be familiar to the re-user.

Moreover, even if it is largely correct to say that only CC0 and CC-BY approaches are common, the summary table also shows that additional minor requirements are often included, such as the explicit references to national law, clauses on personal data, excluded use cases, moral rights, non-endorsement of use cases, and so forth. The very simple characterisation as a 'CC0 or CC-BY approach' is largely accurate, but hides significant complexity. A diligent re-user would, theoretically, need to identify and assess all of these points.

It seems likely however that such diligence is rarely practiced in reality, and that awareness of these legal minutiae may be limited. With that in mind, a reality check was also undertaken during the drafting of this report, by interviewing stakeholders and seeking their opinions on licence compatibility in reality.

A reality check: is licence compatibility a problem?

General approach: interviews with data providers

In the preparation of this legal report, the authors conducted interviews with data providers in five countries: Belgium, Denmark, Estonia, Germany and the United Kingdom. All interviewees were representatives of public administrations responsible for open data sharing or operators of open data portals in their countries.

The interviews were semi-structured, and built around three central questions.

1. *Do you think that licence compatibility is a problem in practice today, based on your own experiences or on feedback from data users or other stakeholders?*
    o *If so, what are the main problems that you're familiar with?*
2. *Did you (or do you) take into account licence compatibility when choosing the licences under which datasets are made available, assuming that you have an influence on this choice?*
3. *How do you think licence compatibility problems can be reduced or eliminated?*

The main findings and observations will be summarised in the following sections.

Summary of the individual interviews with data providers

Belgium

The representative from Belgium is responsible for supporting the federal open data portal platform data.gov.be, and for community building, also towards the various regional data portals in Belgium. There is a specific federal helpdesk to resolve data licence compatibility questions. Belgian federal law in principle requires public authorities to seek guidance from this helpdesk on their licence choices, but this rarely happens in practice.

The main compatibility problem is that various regions/communities in Belgium have elected to use slightly different unique licence templates than the federal ones (next to the CC licences). These unique templates all seem superficially fairly similar to existing CC licences, but vary in small details. This creates problems – e.g. when combining regional data sets that contain address information, since this is one case where the regional entities that issue these datasets apply different licences. Even within the Belgian borders (i.e. in a single country), this is challenging, and requires a lot of effort to do correctly.

Licence templates sometimes also overregulate – e.g. indicating that data cannot be used for unlawful or criminal purposes, which is already addressed by legislation and does not necessarily need to be addressed via licences. A historical concern was also whether CC licences would even be valid under Belgian law, but this concern has become smaller over time, as CC licences are now a commonly accepted practice across the EU now.

Moreover, enforcement practices for bespoke licences seem to be quite lax in many instances – unique licences can be quite detailed and granular, but the bodies using them have only limited ability and willingness to enforce / apply their terms or to help to address licence compatibility concerns.

As a result, when datasets with different licences need to be combined, either (1) parties accept the risk without formally checking compatibility (i.e. ignoring the problem); (2) an investigation into compatibility is done – which however only results in a specific opinion on this issue and thus a certain degree of risk acceptance; or (3) exceptionally the data is re-licenced under a new (compatible) licence on an ad hoc basis to resolve the issue. Finding solutions is a slow and sometimes costly process, Moreover, it does not scale – even within Belgium, compatibility checks are complex and time consuming. At the cross-border level, the effort is prohibitive, and the only realistic option is often to take a 'best efforts' approach and accept residual risks.

There is a growing awareness of this issue, and a growing tendency to move to CC licences, since this is the only currently scalable approach, even within Belgium. Certainly, at the federal level, efforts are being undertaken to promote CC licences as the main solution that works in practice. This is also important for cross-border collaborations with neighbouring Member States, where CC is the only recognisable option.

It was discussed whether this problem could also be solved by requiring authorities to justify when they choose something other than a limited range of standard licences. In the interviewee's opinion, however, this would not be a sufficient solution since the effort is often just not undertaken to draft a serious justification. Compatibility matrices that map the compatibility between licences can occasionally be useful, but certainly cannot solve the problem entirely. Ultimately, such a matrix is still just an opinion, and moreover it does not scale. The federal advice is generally to apply CC instead.

It is clear that the number of licences should be reduced, with a strong preference for CC, and perhaps guidelines should be provided on how they should be applied (e.g. on how attribution should be done; when share-alike is a useful option; or when commercial use should be managed) – but without modifying the licences or creating new variants.

Denmark

The representative from Denmark provides support to the new Danish national open data portal https://datavejviser.dk/ and assists public administrations in making their own licensing decisions. At the time of the interview, the portal contained 2 985 datasets, of which 2 508 (84 %) were available under the international CC-BY 4.0 licence. The only other significant licence shown on the Danish portal is a national licence designed for sharing energy data (the *Energinet licens*, which was developed as a standardised public sector licence around 2013).

This situation was partially the result of an explicit evaluation to determine the optimal licence choice for Denmark in order to satisfy its obligations under the open data directive, the outcome of which was to promote CC licences. This was however only decided in the last year; the high rate of adoption of CC-BY on the portal today is not only the result of this policy decision to favour CC, but also of the fact that the Danish statistical authority historically already used CC-BY licences for a large number of datasets. This inflates the number on the portal a bit (since the statistical authority publishes many datasets), but has also acted as a practical encouragement for other administrations to make the same default choice, even before the evaluation was done.

As a result, Denmark began to actively encourage the use of the CC licences (although no formal legal obligation to choose CC licences exists, outside of high-value datasets – CC is advised as a future proof option, but not mandatory). Denmark does not have another specific national licence template that is actively promoted. A short presentation and interactive webinar (a 'crash course' on licensing) is available for administrations to guide them in their licensing decisions, and open dialogue is often possible to discuss alternatives. The general outcome is normally that any data that should be openly available is best licenced under CC-BY. Other licences are really only considered when there are specific legal constraints that impede the choosing of CC-BY (e.g. because a third party owns the intellectual property rights). Even when there is specific legislation that limits the use cases, the general approach in Denmark is to reflect

these limitations in the metadata of the datasets, not in the licence. For example, for a dataset with company data that includes contact addresses, the prohibition to use these for direct marketing purposes is included in the metadata, not in a unique licence. In this way, user constraints can be communicated without modifying licences themselves. Licences only focus on intellectual property rights issues – CC licences do not mean 'do what you want', but 'do what you want within the limitations of the law' – and those limitations do not have to be repeated in the licences themselves.

Most of the questions raised by public administrations are actually not about copyright or licence compatibility, but on much more practical issues – who makes licensing decisions, what are the implications of a licence, or of choosing one licence rather than another, and what do other public authorities do? The issue of licence conflicts is not really considered in practice. The interviewee provides support in answering these questions via a standardised presentation, and the recommendation to opt for CC licences where possible is generally followed. While there are occasionally questions about the legal validity of CC licences under Danish law, the general consensus is that the legal soundness of CC licences under EU law has been sufficiently demonstrated through various legal cases over the past decades (notably in France and Germany; this validity was formally assessed and confirmed by the Danish General Attorney).

The ambition to facilitate internationalisation was another driver behind the choice for CC licences over any uniquely Danish template, as well as simple recognisability: prior to launching the new Danish portal, data re-users were consulted, and the feedback was that licences should be simple and intuitively recognisable – complex licences would often simply not even be read, due to legal uncertainty. Re-users felt the CC licences were simple, intuitive and recognisable, and that they eliminated most of the need for actual legal assessment. This is primarily true for CC-BY and CC0 licences, which are in line with the legal requirements for high value datasets; the CC-SA and CC-NC (share-alike and non-commercial licences respectively) are perceived as triggering new problems and questions without much benefit.

Estonia

The Estonian representatives work on open data policies, and support the Estonian national open data portal (https://avaandmed.eesti.ee/), including by providing an open data licensing guide on that portal. The guide explicitly recommends the usage of the CC licences; no national Estonian licence was ever created. The guide is not limited to CC – other open licences are referenced and even recommended as well (including the Open Data Commons Public Domain Dedication and License, the Open Data Commons Attribution License or Open Data Commons Open Database License), but a dedicated section notes that 'CC licences are international standard licences and recommended and suitable for use in Estonia'.

There is a policy preference for using CC0 in particular, and this is actually the most common choice for datasets on the Estonian portal. Administrations are allowed to choose other licences – there is no legal constraint on this point, only an official policy to preferably use explicitly referenced open licences. There is also a series of online instructions that explain the importance and role of licences and assist data providers in making an appropriate choices for open licences. Non-open licences are permitted, but not presented as a default option. Promoting only a limited number of open licences reduces complexity for data providers when making their choices.

The interviewees provide legal and policy support mainly to the data providers when they ask for it, and occasionally (but much more rarely) also to data re-users. Legal questions tend to relate to how licensing choices should be made, and how to deal with specific legal constraints (e.g. the responsibilities when the dataset contains personal data). This is also covered by the guide; the general position is that such constraints are not (and should not be) covered by the terms of the licences themselves. Generally, the questions should be whether the data should be open. If yes, then a CC licence should usually be the best choice. If more restrictions are really necessary, then quite likely open data is not

the right model: if data has been found to be 'safe' to be open and available on the open data portal, then a flexible and international licence (such as one of the CC licences) is normally the best choice. The portal should ideally be a 'safe space' for open data without particular constraints.

There is little experience with licence compatibility problems in Estonia; questions on this topic virtually never come up. The limited focus on a smaller set of internationally recognised open licences is probably a part of the reason, but equally the interviewees suspect that data re-users simply do not check this and just assume that logically, data on an open data portal should be openly available for any reuse, without legal verification.

Explicit conflicts or questions relating to licence compatibility are thus avoided by creating a 'safe space' environment where compatibility problems are much less likely to occur, because of a systematic preference for permissive licences such as CC0. The 'safe space' approach goes in both directions: it should comfort the data re-users, but also the data providers, who should be confident that their responsibility does not extend to any and all reuse of their data. It should not be their role to avoid all potential problems that could occur due to the behaviour of the re-users.

Germany

The representative for Germany works on the German federal open data portal https://www.govdata.de/ as a product manager. At the time of the interview, the portal contained 86 238 datasets, and the principal licences in use were the German Data Licence with attribution v2.0 (*Datenlizenz Deutschland Namensnennung 2.0*, with 61 295 usages, i.e. approximately 71 % of the datasets), the German Data Licence Zero v2.0 (*Datenlizenz Deutschland Zero 2.0*, with 7 222 usages (approximately 8 %)) and the International CC-BY 4.0 Licence, with 4 152 usages (approximately 5 %). Thus, the German landscape is clearly dominated by national licence variants.

The interviewee explains that the German open data licensing landscape is regulated, in the sense that public administrations must choose from a closed list of 30 permitted licences. Four of these (the two aforementioned German licences, CC-BY and CC0) are actively promoted, but preferences can vary significantly between the different German states ([2]).

The closed list does not imply a strong licensing constraint, but is rather a metadata constraint – the list largely consists of explicitly named licences such as those mentioned above, but also includes multiple open categories of licences such as other open licences (*andere offene Lizenz*), other closed licences (*andere geschlossene Lizenz*), and other commercial licences (*andere kommerzielle Lizenz*).

The interviewee affirms that licence compatibility questions do come up frequently, even between seemingly very comparable licences (such as the German attribution licence and CC-BY, which are both permissive open data licences with an attribution requirement). Opinions seem to vary a lot; in practice, the two German licences seem to be principally chosen because they were locally designed and therefore are commonly considered to be inherently suited to a German legal and policy context. The CC licences are created from a US-based perspective, and do not have this same perception. For some German lawyers, there is still the concern that the CC licences may not be valid under German law, although there is no case-law that supports this concern. None the less, for some stakeholders the conviction remains that the German licences are a safe choice and international licences are not.

There is also a disconnect between the data providers and the data user communities. Data providers have a tendency to pick the most common choice (which is currently the German Data Licence with attribution), without much consideration of the impact that this has on the data users (e.g. what the impact is of having to mention the source in every use case, when combining multiple datasets; this perspective is not on the radar of most public administrations).

([2]) Germany is a federal country, that has granted the competence to its 16 constituent states to decide on the licences that should be used by their respective public administrations.

Moreover, once data is made available, there is normally no significant verification by the data providers of how, or whether, their licence terms are complied with. This means that user communities are not necessarily very focused on remaining compliant with licensing constraints.

In general, there is limited attention paid to the practical implications of licensing choices, and this issue can occur even without combining multiple different licences. A mobility app that combines 40 different datasets that all use CC-BY has no divergence in the licence terms as such, but it can be difficult or impossible to correctly display the 40 mandated attribution notices. There is insufficient awareness of these issues; guidance on this topic would be useful.

The interviewee has a strong preference for using CC licences only (no national variant licences) to improve EU wide usability, and then in particular to favour CC0 licences, which also eliminate the attribution problem. This should significantly facilitate usage; and better benchmarking and comparisons between Member States could be useful as well to show the effectiveness of opting only for international licences such as CC licences.

United Kingdom

The representatives from the United Kingdom noted that they were initiating an exercise at present to review the UK's own licensing policies, and that they could not presently comment extensively on licence compatibility issues.

At the current time, however, the United Kingdom favours the Open Government Licence 3.0, as developed by the National Archives. This is an open licence with a limited attribution obligation for re-users. Use of the OGL is not mandatory under UK law, but its use is recommended by the National Archives for any public sector bodies owning their own copyright and database rights. The UK open data portal https://www.data.gov.uk/ contained 56 460 datasets at the time of the interview, and 26 270 (46 %) of these used the OGL.

Extensive guidance on the meaning and use of the OGL is made available under the UK government licensing framework, which also identifies exceptions and limitations of the OGL, contains branding guidelines and provides examples of how to attribute correctly. A small range of alternative licences is presented as well, including a non-commercial government licence, a charged licence and a developer licence.

No immediate feedback could be provided on licence compatibility issues, given the ongoing evaluation of the United Kingdom's national open data policies.

Key findings, trends and lessons learned across the examined countries

While there are significant differences between the interviewees, several horizontal observations can be made. Firstly, the national interviews confirm two observations with respect to the data.europa.eu portal: there is an increasing trend of relying on CC licences as a way to simplify reuse, both for data providers and for aspiring data re-users; and where national licences exist, they follow the CC0 or CC-BY logic – i.e. they focus on comprehensive rights grants to the users, or only contain an attribution requirement. Other requirements – such as a limitation to non-commercial use and share-alike obligations – play a marginal role at best, since they create new problems without a commensurate benefit.

Secondly, the interviews confirm that licence compatibility is a real challenge in practice, for which re-users currently do not have a very satisfactory answer. The main resolution strategy focuses on risk acceptance – i.e. not attempting to resolve possible licence conflicts at all, but rather tacitly accepting the problem, and counting on the inaction or the good will of the data providers. That approach is typically successful, since data providers have limited interest in following up and limited capacity to follow up on individual reuse cases, or to enforce their chosen licence terms.

Looking at the preference of the interviewees, they unanimously favoured an increased use of CC-BY and (especially) CC0 licences, which were seen as the main feasible solutions to foster international reuse of open data (with the exception of the UK interviewee, which could not yet express an opinion on this topic since evaluations were still ongoing). This position was even observed in Member States (as well as in the United Kingdom) with their own national or regional licences that were tailored to their own legal systems – even in these countries, the increased simplicity and practicality that results from systematic use of CC licences was seen to outweigh the legal risk of licences that were not optimally aligned with European legal frameworks – especially because, after more than a decade of increasing usage in open data sets, no CC licence had been found to be rejected as invalid by a European court.

Reducing the problem: what can be done to resolve licence conflicts?

Lessons learned on licence compatibility – avoiding or reducing conflicts

Based on the interviews, there are some clear lessons that seem relatively universally applicable in any open data context. They can be summarized as follows.

- **Understand that licensing must be made easy.** Data providers should be actively guided towards a very small number of licence models from which they can choose. Offering a breadth of licence choices has an excessive cost in terms of complexity. Moreover, the text of the licences themselves should also be kept simple, since the analysis of complex and specific licences may be a burden for re-users, and thus act as a disincentive to use that data. The consensus is that neither most data providers nor most data re-users are willing to consider or capable of considering details of complex licence texts, so a simpler landscape makes it more likely that licences will have any meaning in practice.
- **Avoid, or at least limit the use of, unique national (or regional) licences.** Such licences were initially presented as being more suited for the European policy context, especially since the alternatives of the CC licences were based on US legal traditions. In practice, local licences are seen to create disproportionate burdens for re-users: combining datasets from different countries or regions becomes prohibitively difficult. Moreover, the conceptual legal risk behind international licences (i.e. the risk that they would be declared legally invalid by an EU court) does not seem to have manifested itself in reality.
- **Default to CC licences, specifically CC-BY and CC0.** The interviewees unanimously noted that CC-BY and CC0 were the optimal choices for almost any open data use case. CC0 was the choice most favoured by the interviewees, since it eliminates most licence compatibility problems; but the statistical reality presently seems to be that most public administrations wish to receive attribution. The driver and benefit of this preference is not clear.
- **Avoid share-alike licences, and licences that are limited to non-commercial use or to use within a specific sector or use case.** The present report doesn't find support for *all* CC licences without distinction, but rather specifically for CC-BY and CC0. CC-SA and CC-NC (or other variants of CC licences that contain share-alike and non-commercial provisions) do not find the same favour with the interviewees (nor indeed are they commonly found on European data portals), as these licences create new difficulties in their interpretation and in their compatibility. Indeed, several interviewees noted that a key challenge in convincing administrations to use CC-BY and CC0 related to the idea that public sector data should not be used for private gain. Once it became

clear that such private use could benefit the public too, the transition to CC-BY or CC0 became easier. The same applies for any non-CC licences that exclude commercial use or require a share-alike approach: they are not favoured.

- **Be transparent, modest and realistic about attribution requirements, and show examples.** A recurring challenge – and the principal reason why interviewees favour CC0 over CC-BY – is that attribution requirements are often insufficiently clear and therefore hard to interpret, difficult to apply cumulatively when combining multiple datasets, and unsuitable for some use cases where interfaces do not easily allow a re-user to give credit to multiple (sometimes dozens) of data providers. When using CC-BY or other licences with an attribution requirement, Member States should issue guidance on how this requirement can be observed, and data providers should be explicit in stating what exactly would satisfy their requirements. Moreover, data providers should be challenged on what benefit they expect when imposing attribution requirements and be encouraged to reflect on the implications of an attribution duty on re-users.

- **Based on the above, evaluate the feasibility of modifying the licences applied to existing datasets.** As this report shows, there is a significant number of datasets that are made available under local licences that are largely unknown in other countries than those of the data provider. These create occasional licence compatibility challenges. It is worth revisiting these datasets and examining how the recommendations above can be implemented – i.e. by changing the licence on open data portals, to make the datasets available under (preferably) CC0 or CC-BY licences instead.

- **Improve data literacy and data licensing skills, e.g. by providing e-learning material that steers data providers towards CC-BY and CC0 licences where possible.** The interviewees affirm that many data providers need a primer on what licensing entails, what the best choices are and how they should proceed. Compact guidance that steers data providers appropriately (i.e. towards CC-BY and CC0 licences according to this research paper) can be highly effective in driving changes.

- **Engage in a dialogue. Ask for justifications and reasons when a different choice is made.** Building on the data literacy/data licensing recommendation above, it is important to leave margin for interaction, questions and simple dialogue. A local (national or regional) contact point should be available to data providers where they can raise questions on concerns that they might have, from simple governance points (who chooses the licence in our organisation?), to policy points (can we change our minds later?), and legal risk management (where does our responsibility end when re-users misbehave?).

- **Create a safe environment for data providers and data re-users.** The ideal open data portal is not just a portal that provides transparency to data providers and data re-users on what their choices, risks and obligations are, but rather a data portal that ensures that these questions do not have to be at the forefront of their minds all of the time. Regarding data providers, this implies that they are guided towards accessible licence choices and are informed of where their responsibilities stop. Regarding data re-users, they should be able to trust that the data they find can be reused without requiring in-depth legal assessment of unique obligations contained in the licences. For this reason too, simple and well-known licences such as CC-BY and CC0 are generally preferable.

- **Respect the fact that licences cannot and should not resolve all legal questions.** One common criticism of CC licences is that they do not address all possible legal issues, including unlawful reuse, e-government legislation, unfair commercial practices, data protection and consumer protection. This is correct, but no licence template could or should address all legal challenges that could possibly emerge, especially where those challenges are already governed by binding law. Where needed, guidance on such challenges can be published separately; it is not useful or necessary to create specific licences that identify nationally applicable legal constraints that likely will not be understandable in international use cases.

In conclusion – broader policy recommendations for the future

The central finding of this legal report is that **licence compatibility challenges are indeed a problem today**, and that they can be **most effectively mitigated by systematically opting for CC-BY and CC0** licences. The analysis of statistical data from the data.europa.eu portal also showed that the shift towards CC licences is already happening today.

It is however also important to recognise that a simple shift to CC-BY and CC0 licences is not a panacea that would resolve licence conflicts comprehensively. It is at any rate worth **developing guidance at the national level, or preferably at the EU level, on how to optimally deal with attribution obligations**, since this is a problem for which no clear solution is available today.

Finally, European countries should reflect on **how they can support data providers with their licensing decisions**, in a way that facilitates decision-making for the data providers and is conducive to meeting the concerns of data re-users – i.e. in a manner that is simple and pragmatic, e.g. via licensing crash courses, webinars and help desks. This approach appears to be effective in practice in guiding a shift towards licensing choices that reduce the risk of compatibility issues and thus encourage reuse of open data.

Bibliography – sources and references

- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast) (OJ L 172, 26.6.2019, p. 56) (Open Data Directive), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024.
- List of licences on the data.europa.eu portal, data.europa.eu/api/hub/search/search?filter=dataset (see under 'licences').
- Bernelin, M., 'The compatibility of open/free licences: a legal imbroglio', *International Journal of Law and Information Technology*, Vol. 28, No 2, Summer 2020, pp. 93–111, https://doi.org/10.1093/ijlit/eaaa010.
- Khayyat, M. and Bannister, F., 'Open data licensing: More than meets the eye', *Information Polity*, Vol. 20, No 4, 2015, pp. 231–252, https://content.iospress.com/articles/information-polity/ip357.
- data.europa.eu open data licensing e-learning module, https://data.europa.eu/elearning/en/module4/#/id/co-01.
- Creative Commons licensing overview, https://creativecommons.org/licenses/.
- European Commission, Directorate-General for Communications Networks, Content and Technology, *Study to support an impact assessment for the review of the database directive – Final report*, Publications Office of the European Union, Luxembourg, 2022, https://ec.europa.eu/newsroom/dae/redirection/document/83514.
- OpenAire, 'How do I license my research data?', OpenAire website, https://www.openaire.eu/how-do-i-license-my-research-data.
- Joint Research Centre, Gentile, S., Georgieva, I., Iglesias, M. et al., *Reuse policy – A study on available reuse implementing instruments and licensing considerations*, Publications Office of the European Union, Luxembourg, 2019, https://op.europa.eu/en/publication-detail/-/publication/971c48e2-51e3-11e9-a8ed-01aa75ed71a1/language-en.