Data and licensing management guidelines

Research report and guidelines prepared within the project

STIRData: Specifications and Tools for Interoperable and Reusable data

Co-financed by the Connecting Europe Facility Programme of the European Union, under GA n.
INEA/CEF/ICT/A2019/2063078

Author: Jakub Míšek, Ph.D., Radim Charvát, Ph.D., Matěj Myška, Ph.D.

Masaryk University, Faculty of Law

June 2023

This research report and these guidelines were prepared within the "STIRData: Specifications and Tools for Interoperable and Reusable data" project, co-financed by the Connecting Europe Facility Programme of the European Union, under GA n. INEA/CEF/ICT/A2019/2063078.

The availability of resources has been verified as of 30 June 2023, unless a different date is specified for a particular resource.

1. Introduction and methodology 4
Part I: Legal analysis 8
2. International and EU law 8
2.1 International treaties 8
2.2 Law of the European Union 9
2.3 Future and anticipated legislative actions 15
2.4 Chapter summary 17
3. Protection of data & intellectual property rights 19
3.1 The protection of plain data 20
3.2 Copyright protection 21
3.2.1 Legislative overview 21
3.2.2 Copyright and open data 25
3.3 *Sui generis* database rights 26
3.4 Chapter summary 29
4. Personal data protection 31
4.1 Privacy and personal data protection: An overview of the regulation 31
4.1.1 Introduction to the topic and an overview of the relevant legal instruments 31
4.1.2 An overview of the basic concepts 33
4.2 Open data as personal data processing 37
4.3 The open-data provider as data controller 38
4.4 The open-data reuser as a personal-data controller 43
4.5 Chapter summary 48
5. Terms of use of open data 50
5.1 Contract 50
5.2 Open Licenses 52
5.3 Providing information about the dataset 54
5.4 Chapter summary 55
Part II: Guidelines and a licensing scheme 57
List of selected resources 62

1. Introduction and methodology

Open data is a global phenomenon that is intended to enable the effective use of public sector information (PSI).[1] When we are discussing PSI and its legal regulation, we first need to distinguish between two basic layers – access to PSI, and its reuse. Public administration authorities and local governments produce large amounts of information and data in the course of their statutory tasks. Whether it is, for example, the files on the basis of which a law is applied authoritatively, the environmental data used to decide whether or not to allow construction, or, if we look a level higher, the data used to determine policies for the future development of the state or parts of it, it is always data that public administration works with and which it uses as a foundation to produce any of its outputs. The ability to see the data that has led to such outputs, or even the outputs themselves, is then a prerequisite for civil society to effectively evaluate the performance of public administration. Furthermore, the right to access such data (or PSI in general) is a necessary part of the freedom of expression, a basic civil right that allows for transparency and public scrutiny of how public sector bodies fulfill their duties. Without access to specific important data, it would not be possible to hold a public discussion regarding the actions of our elected officials and public sector bodies.[2]

In addition to these factors, there is high potential value in public sector data, as best characterized by a statement by Rufus Pollock, founder of the Open Knowledge Foundation, who argued that "The best thing to do with your data will be thought of by someone else." This aspect of public sector data relates to the possibility of its subsequent and repeated use. The economic value of public sector data lies, for example, in creating financial savings by linking data that would otherwise be separate due to their different origins,[3] or in developing applications that offer better services to citizens who are happy to pay for them,[4] thereby creating a market that generates economic benefits as indicated by profitability, tax collection, employment and other factors. In one study, the McKinsey Global Institute analyzed seven areas (education, transport, consumer products, energy, oil and gas, healthcare, and consumer finance) and quantified the potential value inherent in the effective publication and reuse of data in these areas at over USD 3,000 billion per

[1] Parts of this analysis draw upon *inter alia* the results of previous research of the main author, which was assembled in a doctoral thesis he defended in 2019 (MÍŠEK, Jakub. *Právní aspekty otevřených dat* [online]. Brno, 2019. Online in Czech: https://is.muni.cz/th/sqe7a/. Doctoral thesis. Masaryk University, Faculty of Law.).

[2] A good example of such a situation was the very difficult access to health data relating to the COVID-19 pandemic in the Czech Republic, which led to further diminishing of public trust.

[3] E.g. Jetzek Thorhildur, Michel Avital and Niels Bjørn-Andersen, 'Generating Value from Open Government Data', *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design* (2013) 16.

[4] E.g. Melissa Lee, Esteve Almirall and Jonathan Wareham, 'Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements' (2016) 59 Communications of the ACM 82, 89.

year.[5] The growing potential of public sector data has also been recognized by the I Union, which since the beginning of the new millennium has used legislation to promote the usability of public sector data in its Member States.[6] In addition to the economic benefits, we can also talk about the "societal value" of such data, for example in terms of greater citizen engagement in public affairs in the form of policy planning, in the areas of economic growth, education, conservation, security, among others. [7]

Access to PSI and its reuse are two interrelated phenomena; the latter cannot naturally occur without the former. However, different legal regimes apply to the two. PSI access is primarily a public authority activity. This is the case even though this activity is often initiated by a person who claims a fundamental right to information and who therefore has the right to demand positive performance from the state in the form of providing this information. The provision or non-provision of PSI is then a direct exercise of public authority in the field of information rights, with all the consequences that this entails.

In contrast, the reuse of PSI primarily concerns the recipients of the data, i.e., third parties who further handle the data. In this case, the law regulates how they can use the data, the limits to use of the data, and what obligations they have to fulfill. The basic principle that applies in this case is the principle of legal license, according to which anyone can do that which is not prohibited by law. However, the fact that the applicant has the right to access the data does not in itself mean that he or she can freely use all the data thus obtained. On the contrary, the applicant is still limited by other legal institutions such as personal data protection, intellectual property rights, and others. Furthermore, a large part of the regulation of PSI reuse is legislation that ensures that data providers provide data in a way that makes it as easy as possible to use. Making the data available in appropriate formats and under clearly defined legal conditions is a prerequisite for its subsequent reuse.

The concept of public sector data reuse has recently been closely linked to the concept of open data as a technologically efficient way of providing and reusing PSI. Of course, one of the possible

[5] MCKINSEY GLOBAL INSTITUTE. Open data: Unlocking innovation and performance with liquid information | McKinsey & Company [online]. 2013 [last accessed 30. 6. 2023]. Online: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information.

[6] See Directive 2003/98/EC, the subsequent amendment Directive 2013/37/EU, and then the most recent version (recast), Directive EU No. 2019/1024.

[7] See, e.g., Erna Ruijer and Evelijn Martinius, 'Researching the Democratic Impact of Open Government Data: A Systematic Literature Review' (2017) 22 Information Polity: The International Journal of Government & Democracy in the Information Age 233; Igbal Safarov, Albert Meijer and Stephan Grimmelikhuijsen, 'Utilization of Open Government Data: A Systematic Literature Review of Types, Conditions, Effects and Users' (2017) 22 Information Polity: The International Journal of Government & Democracy in the Information Age 1.

uses of such data is in public oversight. However, this is not an automatic consequence of the provision and use of PSI. In recent years, there has been an inappropriate conflation of the terms "open data" and "open government" in the sense of transparent and auditable public administration.[8] However, open data in itself has no particular value in the pursuit of open government. Governance can be transparent without the efficiencies brought by new technologies. Conversely, public authorities may indeed provide a lot of open data, but it will not contribute in any way to greater transparency in the exercise of public authority (e.g., public transport timetables).[9] It always depends a great deal on the specific nature of the data provided, i.e., what it tells us. This is also why, in the context of political rights, which can be used to oversee public administration, access to PSI mainly functions on the basis of requests. As Peixoto has argued, even if data is by its nature capable of providing better oversight of public administration, other factors such as free media and a mature political culture are necessary for the actual exercise of such control.[10]

It is necessary to stress that, at least from the legal perspective, open data is understood to be an extremely efficient way to provide PSI for the easiest subsequent reuse. When we discuss open data and its legal regulation, we are mostly concerned with answering the question "how" PSI is provided. The question of "what" PSI is provided is subject to specific regimes of access to PSI. However, when deciding on the questions of what data a public sector body can publish and how it should be done, both of these layers must be analyzed and answered.

This analysis focuses on the legal issues of open data publication and its reuse, with a focus on the datasets of public registers of companies and other legal entities, as this is the main research goal of the whole STIRData project. Nevertheless, the conclusions and recommendations arising from this legal analysis are applicable in broader scope of any PSI and open data publication and reuse.

The aim of this analysis is first to provide some guidance for open-data providers on what the relevant legal questions during the publication process are. Second, this document should help open-data providers find answers, or at least point them to the sources of the answers on these relevant questions in their respective jurisdictions. Thirdly, for the sake of completeness, where relevant, this document aims to provide at least a basic overview of relevant legal questions regarding the reuse of open data. Therefore, this document is divided into two main parts: I) Legal

[8] Some authors, such as Evgeny Morozov, subsequently criticized such merging of these originally technical and social concepts, because this blurring of the boundaries between them gives rise to the illusion that many problems can be solved "at the click of a button." See Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (Public Affairs 2013) Chapter 3.
[9] Harlan Yu and David G Robinson, 'The New Ambiguity of Open Government' (2011) 59 UCLA Law Review Discourse 178.
[10] Tiago Peixoto, 'The Uncertain Relationship between Open Data and Accountability: A Response to Yu and Robinson's the New Ambiguity of Open Government' (2012) UCLA Law Review Discourse 200.

analysis and II) Guidelines and a licensing scheme. The first part provides theoretical background and argumentation for the second part, which is more practically oriented. It offers a checklist of steps that must be considered when publishing open data in order to ensure that such data can be published legally and in the most reusable way.

The first part is divided into four chapters (Chapters 2 to 5). Chapter 2 offers a brief overview of the existing and anticipated international and European Union legal instruments that are relevant to regulation of the access and reuse of PSI. The chapter also contextualizes the position of datasets of public registers of companies and other legal entities. Chapter 3 focuses on the issues of intellectual property protection, which in certain situations can pose an obstacle to the publication and reuse of PSI and open data in particular. The chapter, from the perspective of international and European Union law, briefly explains basic concepts of copyright and database protection. It also provides a recommendation on licensing schemes and the use of public licenses. Chapter 4 focuses on the questions of personal data protection, because this issue is very deeply connected with the publication and reuse of datasets of public registers of companies and other legal entities, as they contain a great deal of personal data regarding ownership of companies. Finally, Chapter 5 offers recommendations on how to properly prepare terms of use and which licenses to use in which cases.

This analysis is prepared from the perspective of international and European Union law. That is because a preparation of detailed analysis that would incorporate specifics of Member States is outside the scope of the STIRData project. The focus on the international and European Union law represents the main limitation of this analysis. However, at the same time, it ensures its flexibility and applicability regardless of the specific national jurisdiction, because the international and European Union legal instruments offer a common, harmonized framework that is applicable throughout the EU.

Part I: Legal analysis

2. International and EU law

2.1 International treaties

At the international level, treaties have focused on access to PSI rather than its reuse. Furthermore, access to PSI was historically incorporated into freedom of expression, rather than expressed as an independent right. The right to information first became relevant in the 1948 Universal Declaration of Human Rights, Article 19, which declares the right to freedom of expression, including the right to seek, receive and impart information and ideas by any means.[11] The next treaty, probably the most important one, was the European Convention on Human Rights (1950), Art. 10 para. 1 of which enshrines the freedom of expression as follows:

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.

Again, we see that access to information is inherently part of freedom of expression. This legal construct was reaffirmed by the European Court of Human Rights in *Magyar Helsinki Bizottság v. Hungary*.[12]

The Convention allows for limitation of freedom of expression, and accordingly, right to access the information. It states that

> [t]he exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[13]

[11] Art. 19 reads as follows: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."
[12] Decision of the Grand Chamber of European Court of Human Rights from 8. 11. 2016, No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*).
[13] Art. 10 para. 2 of the European Convention on Human Rights.

We also find this same formulation in the International Covenant on Civil and Political Rights (1966).[14]

When looking at the historical development of a right to information, other documents from the Council of Europe are also interesting, even if they are only of a consultative nature.[15] They contain basic principles that still comprise key elements of the right to information. They are as follows: i) Every person has the right to receive information held by public authorities; ii) The applicant for information is not obliged to prove the reasons for his/her request; iii) Access to information is to be exercised on the basis of the principle of equality; iv) Information is to be provided within a reasonable time; v) The state authority must state the reasons for withholding information; and vi) Rejection of a request must be reviewable.

More recently, the Council of Europe prepared a Convention on Access to Official Documents (CETS No. 205; Tromsø, 18 June 2009), which aimed to create a "genuine starting point for an effective right of access to official documents in the European region."[16] Convention No. 205 guarantees the right of any person to have access to official documents (PSI) held by public authorities on request. "Public authorities" are defined quite broadly in the Convention and include, in addition to state and local authorities, legislative bodies, the judiciary, and natural and legal persons when they exercise public powers. However, this is only the minimum necessary set of entities - ratifying States may decide on a broader definition of "an obliged entity." The Convention regulates possible restrictions on the provision of information, and generally addresses the proceeding of making a request, including its processing. In terms of reuse of information, Convention 205 does not offer specific regulation.[17] So far, fourteen Council of Europe countries have ratified Convention 205, including Finland, Estonia, Lithuania, Hungary, Montenegro and Ukraine.

2.2 Law of the European Union

The right of access to information, or the obligation of public authorities to provide certain information, has not received much attention in the context of European law. The fundamental provision of European law dealing with the right of access to information is the Charter of Fundamental Rights and Freedoms of the European Union.[18] It includes the right to information as part of the right to freedom of expression under Article 11, but it also includes, in Article 42, a

[14] Art. 19 of the International Covenant on Civil and Political Rights.
[15] For example, see the Declaration of the Committee of Ministers of the Council of Europe on Freedom of Expression of Information of 29 April 1982.
[16] See the Explanatory Report to the Convention [online, cit. 25. 7. 2021]. http://www.worldlii.org/int/other/COETSER/2009/2.html.
[17] See Mireille Van Eechoud and Katleen Janssen, 'Rights of Access to Public Sector Information' (2013) 6 Masaryk University Journal of Law and Technology 471.
[18] Document No. 2010/C83/02.

specific right of access to documents of the Union's institutions, bodies, and other entities, irrespective of the medium in which the documents are held. At the level of secondary legislation, this right has been expressed in general terms in relation to the institutions of the European Union by Regulation (EC) No. 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council, and Commission documents.

Although European legislation is not very extensive in the field of legal regulation of access to information,[19] the situation is quite different in the field of legal regulation of the re-use of PSI. In fact, the European Union (or formerly the European Community) has been the main driving force in Europe in the field of introducing legislation enabling the reuse of PSI since the 1980s.[20] In 1989 the Commission issued its "Guidelines for improving the synergy between the public and the private sectors in the information market."[21] Article 1 stated that public organizations should make their information available to the private sector in a reusable form through electronic information services. Exceptions could be made in cases where access to information was precluded on grounds of legitimate public interest. However, binding legislation at the level of secondary law only came with Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the reuse of public sector information, popularly known as the "PSI Directive." The primary objective of PSI Directive was to create a workable legal environment for the transparent and non-discriminatory use of public sector information. It should be stressed that the PSI Directive did not impose any obligation to provide information. It merely stipulated that, where information is provided, it should be done in such a way as to facilitate its reuse by the private sector to the greatest possible degree. This approach was not changed by the directive's amendment in the form of Directive 2013/37/EU, nor has it been since.

Directive 2003/98/EC, as amended by Directive 2013/37/EC, set out at a general level the basic rules and principles concerning the reusability of published public sector information. The Directive was applied to all documents[22] made available by public sector bodies, except where specified in the Directive. These included documents that did not fall within the scope of the public tasks of the public sector body concerned, documents encumbered by third-party intellectual property rights, documents with restricted access for reasons of personal data protection, and/or

[19] It must be noted that there is a problematic and yet unresolved issue of whether the European Union even has the authority or competence to oblige Member States to ensure general access to public sector information. See Eechoud and Janssen (n 17) 478–480.

[20] For more context, see Herbert Burkert, 'Public Sector Information: Towards a More Comprehensive Approach in Information Law' (1992) Journal of Law and Information Science 47, 49.

[21] Online: https://op.europa.eu/en/publication-detail/-/publication/7c37bbee-4363-4ec7-91ff-b6848142ec97/language-en.

[22] The term document is broadly defined in Art. 2 para. 3 as "any content whatever its medium (written on paper or stored in electronic form or as a sound, visual or audiovisual recording)" or any part of such content.

documents held by cultural institutions other than libraries, museums, and archives.[23] Thus, situations could arise where national law allowed for the *provision* of a document (i.e., access to it), but since the document was covered by an exception to the application of the directive, the State concerned was not obliged to ensure the *reusability* of such information.[24] The general principle of the directive was to ensure that information would be provided as openly as possible (both technically and legally) so that it is as easily reusable as possible.

It is important to note that the directive does not differentiate between a situation when the data is provided voluntarily at the discretion of the data provider and a situation when there is an existing legal duty to provide the data (regardless of whether it arises from the EU or national law). In both cases, voluntary and mandatory data provision, as well as the principles of the directive and the duties arising from it, will all apply, because it is assumed that even voluntary data publication is done in accordance with the law of the Member State. Thus, if it falls within the scope of the directive, the directive will apply.[25] However, data providers should be aware of their national legal obligations and should verify whether they must or at least can provide the data in question.

The PSI Directive was replaced by Directive (EU) 2019/1024 of 20 June 2019 on open data and the re-use of public sector information (known as the "OD Directive"), which preserves the same grounding principles as the previous legal instrument. The OD Directive, which is a recast (i.e., a new, rewritten version), of the PSI directive, came into force on 17 July 2021.

Art. 1 para. 2 of the OD Directive sets down a list of exceptions that are outside the scope of the directive. In the context of this analysis, the most important of these are:

- c) documents for which third parties hold intellectual property rights;
- d) documents, such as sensitive data, which are excluded from access by virtue of the access regimes in the Member State, including on grounds of:
    - the protection of national security (namely, State security), defence, or public security;
    - statistical confidentiality;

[23] The exemption from the exception, i.e., the return of libraries, museums, and archives to the scope of application of the Directive, was a change introduced by Directive 2013/37/EU. Its importance is seen particularly in the context of facilitating the digitalization and dissemination of the content of library collections.

[24] One example we can mention would be the copyrighted work of a third party which the public sector body provides for access, but cannot license for further reuse.

[25] For instance, in the Czech Republic relatively many databases and information sources have a legal requirement to be accessible; however, there also exists a possibility of voluntary data publication, because of the general discretion to do so given by Section 5 para. 5 of the Freedom of Information Act (act No. 106/1999 Sb.).

    - commercial confidentiality (including business, professional, or company secrets);
- h) documents, access to which is excluded or restricted by virtue of the access regimes on grounds of protection of personal data, and parts of documents accessible by virtue of those regimes which contain personal data the reuse of which has been defined by law as being incompatible with the law concerning the protection of individuals with regard to the processing of personal data or as undermining the protection of privacy and the integrity of the individual, in particular in accordance with Union or national law regarding the protection of personal data.

It is important to realize that point h) does not exempt personal data processing from the application of the OD Directive outright. If European or national law allows for or requires publication of a given dataset containing personal data, the OD Directive will generally apply, unless there is another kind of exception on the level of national law.

The general principle of the directive is formulated in Art. 3 para. 1 as follows: "Member States shall ensure that documents to which this Directive applies in accordance with Article 1 shall be reusable for commercial or non-commercial purposes."[26] Art. 5 para. 1 sets out basic requirements for the quality of provided documents. They must be made available in "any pre-existing format or language and, where possible and appropriate, by electronic means, in formats that are open,[27] machine-readable,[28] accessible, findable and reusable, together with their metadata." Furthermore, both the format and metadata should, where possible, be compliant with "formal open standards."[29] a formal open standard is defined as a "standard which has been laid down in written form, detailing specifications for the requirements on how to ensure software interoperability."[30] It is an important, albeit sometimes overlooked, tool for ensuring the interoperability of content provided by different providers. When the same type of dataset is provided by multiple data providers in their specific fields, it can easily lead to the proliferation of data standards, which in turn would lower the possibility of the effective connection and reuse of such data. Legislators realized the necessity for interoperability; however, at the same time it is not practically possible to legislate specific technical standards in necessary detail at the European level. Therefore, they introduced the concept of formal open standards. A formal open standard is a non-legislated document of technical norms that describes in a great detail how (in

[26] As in the previous directive, the term "document" means any content, whatever its medium or any of its parts.
[27] An open format is defined in Art. 2 para. 14 as "a file format that is platform-independent and made available to the public without any restriction that impedes the re-use of documents."
[28] A machine-readable format is defined in Art. 2 para. 13 as "a file format structured so that software applications can easily identify, recognise and extract specific data, including individual statements of fact, and their internal structure."
[29] Art. 5 para. 1 OD Directive.
[30] Art. 2 para. 15 OD Directive.

what formats and structures) any data is to be provided; the OD Directive does not provide for any official authority to issue formal open standards. The advantage of a formal open standard is its flexibility. Since it may become necessary to specify formal open standards in more detail or update them over the course of their use and lifetime in light of future technical developments, the informal way in which they are issued is an appropriate feature. At the same time, a formal open standard is a purely technical document, whose primary audience is the people who directly prepare the data for publication (usually the IT department).

The OD Directive also covers the question of costs for providing information and data. Art. 6 stipulates a general rule that the reuse of documents must be free of charge. The main exception to this rule is that data providers can demand "recovery of the marginal costs incurred for the reproduction, provision and dissemination of documents as well anonymisation of personal data and measures taken to protect commercially confidential information," which includes costs for particularly extensive searches for requested information.[31] Art. 6 para. 2 enumerates specific exceptions from the general rule of feeless access to and reuse of PSI; however, none of those are relevant for the scope of this report.

Although the OD Directive was built on the PSI Directive, it also incorporated several new important concepts. It broadened the scope of its application to public undertakings[32] and research data,[33] it introduced rules for providing dynamic data,[34] and it in fact excluded application of a *sui generis* right to database protection,[35] to name a few. In the context of this analysis, however, probably the most important change was the introduction of so-called "High Value Datasets" (hereafter HVDs). In the view of European policymakers, certain types of data were so important Union-wide for their socioeconomic potential that they should be mandatorily provided by each Member State.[36] Therefore, Annex I of the directive was created based on Art. 13, which a list of general categories of important data. As of now, these include statistics, earth observation and environment, meteorological, geospatial, companies and company ownership and mobility.[37] Specific datasets, which will be mandatorily provided EU-wide, are listed in an implementing act based on Art. 14 of the OD Directive.

[31] Recital 36 OD Directive.
[32] Art. 1 para. 1 letter b) OD Directive.
[33] Art. 10 OD Directive
[34] Art. 5 para. 5 and 6 OD Directive.
[35] Art. 1 para. 6 OD Directive. For more about this topic, see Chapter 3.3 of this report.
[36] Recital 68 of the OD Directive states: "A Union-wide list of datasets with a particular potential to generate socioeconomic benefits together with harmonised re-use conditions constitutes an important enabler of cross-border data applications and services."
[37] The list will most likely be enlarged in the future.

The Commission issued Implementing Regulation No. 2023/138 on 21 December 2022.[38] The Regulation specifies not only which specific data from a given category member states are obliged to provide, but also sets out (albeit in very general terms) the basic technical requirements for the interoperability of these data. Member states have to fulfil obligations arising from the regulation by June 2024.

Thematic Area 5 of the Annex, which is relevant to the scope of this research report, describes in point 5.1 mandatorily disclosed data on companies and company ownership. Specifically, the following data are required to be published:

- Name of the company (full version; alternative names when applicable);
- Company status (such as when it is closed, struck off the register, wound up, dissolved (as well as the date of these events), economically active or inactive as defined in national law);
- Registration date;
- Registered office address;
- Legal form;
- Registration number;
- Member State where the company is registered;
- Activity/activities that are the object of the company, such as the NACE code.

Furthermore, the member states must ensure publication of accounting documents, which include: i) Financial statements (incl. the list of participating interests, subsidiary undertakings and associated undertakings, their registered office address and proportion of capital held), audit reports; ii) non-financial statements, management reports and other statements or reports, and iii) annual financial reports. Unfortunately, the European law maker did not take the chance to include also at least the most important data regarding company ownership. The reason behind this decision is personal data protection and unwillingness of several member states to do so. Such data would necessarily include personal data of natural persons that are involved in company ownership. The relation of open data publication and personal data protection will be discussed further in this report. However, at this moment it can be said, that even though the reason of personal data protection is understandable, it should be noted that it had negative effect on the possibility of ensuring and strengthening transparency across the EU.

Point 5.2 of Annex sets out the basic requirements for the data interoperability. The datasets about companies must be made available for re-use without undue delay after the latest update, under

[38] Online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2023.019.01.0043.01.ENG

conditions and licence that will not make further use of the data impossible or difficult, in an open and machine-readable format, through APIs and bulk download and at individual company level. The datasets must also include documentation describing at least the data structure and semantics. Finally, the datasets must use Union or internationally recognised and publicly documented controlled vocabularies and taxonomies, where available, such as the Core Business Vocabulary. These requirements provide only a very basic standardisation, which of course fails to cover the necessary level of detail to ensure full interoperability across the EU. In view of this, it is necessary to create non-legal standards, in the sense of formal open standards, which specify sufficient detail to achieve the desired interoperability.

Apart from the general legal framework of PSI reuse, there are area specific legal instruments, with priority application, due to the general legal doctrine of *lex specialis*.[39] Examples of such cases include the INSPIRE Directive[40] in the case of spatial information, or Directive No. 2010/40/EU[41] and its implementing acts,[42] which regulate transportation and traffic data. However, these areas lay beyond the scope of this report, and they are not discussed in any further detail.

2.3 Future and anticipated legislative actions

Currently, we find ourselves in the midst of a major wave of ongoing European legal reform of the regulation of data. This wave already includes several existing legal documents, like regulation No. 2016/679, the General Data Protection Regulation (GDPR),[43] and the aforementioned OD Directive. However, as publicly available documents demonstrate, we can expect a lot more in this area. In February 2020, the Commission published "A European Strategy for Data," a policy

[39] This principle states that if there are multiple laws that govern the same factual situation, the one that governs a more specific subject matter will be applied.
[40] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).
[41] Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the Framework for the Deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport.
[42] Commission Delegated Regulation (EU) 2015/962 of 18 December 2014 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide real-time traffic information services; Commission Delegated Regulation (EU) 2017/1926 of 31 May 2017 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services; Commission Delegated Regulation (EU) No. 886/2013 of 15 May 2013 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to data and procedures for the provision, where possible, of road-safety-related minimum universal traffic information free of charge to users; and Commission Delegated Regulation (EU) No. 885/2013 of 15 May 2013 supplementing ITS Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of information services for safe and secure parking places for trucks and commercial vehicles.
[43] And related legislation, such as Directive No. 2016/670, Regulation No. 2018/1725, and a proposal for e-Privacy regulation.

document that foreshadowed future legislative efforts concerning the regulation of data.[44] The aim of the strategy is to create a single digital space that will facilitate the efficient sharing and use of large amounts of data from both public and private spheres. Infrastructures for the data-based economy should be promoted – large repositories that enable efficient big data analytics and machine learning. In return, organizations providing data would gain access to the data of others. At the same time, conflicting interests (privacy, trade secrets, etc.) need to be protected. The Commission identified several issues that hindered the potential of highly efficient data handling, like the low availability of data, imbalances in market power, and low data interoperability and quality, as well as low skills and data literacy on the side of data reusers and the general public.

The first proposal following "A European Strategy for Data" was proposal for a regulation on European data governance (the Data Governance Act, "DGA") presented by the Commission in November 2020. The final version of the act was enacted in May 2022 as Regulation 2022/868 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).[45] It consists of nine chapters comprising 38 articles. The regulation is designed to facilitate data sharing (personal and non-personal data) in the EU by making public administration data more accessible. Furthermore, it aims to strengthen trust in data sharing intermediaries, a new type of data controllers, whose services are expected to be used by third parties in different data spaces. The regulation does not create any new duties to provide specific data, or to process that in a specific way. In some ways, Regulation 2022/868 is a meta-regulation for creating a trusted and secure space for sharing and managing different data without providing detailed rules for specific kinds of data. The regulation itself "regulates" three main areas: i) the reuse of PSI, which lies outside of the scope of the OD Directive; ii) securing rules for creation of safe and trustworthy data intermediation services; and iii) creating a safe environment for data altruism. In the context of this report, the first of these is the most important; Chapter II of the regulation addresses this area.

Chapter II focuses on public sector data that does not fall within the scope of the OD Directive, because such data are subject to the rights of others, such as trade secrets, statistical confidentiality, third-party intellectual property rights, or the publication of which would constitute an infringement of data protection law. If approved, the DGA will contribute to the creation of an internal market for data by facilitating the emergence of new services by creating a set of harmonized provisions. The Commission believes this will make it easier for ?providers of

[44] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, A European Strategy for Data. Brussels, 19. 2. 2020, COM(2020) 66 final. Online: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066.
[45] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

these new services to operate across borders. At the same time, like the OD Directive, it does not impose any direct obligation to provide specific data. It is left to Member States to decide what data will be made available this way. The DGA, built on the FAIR data principles,[46] only sets a minimum standard of quality and principles for access to the data that falls under the purview of the Act.

As mentioned previously, Chapter II of the DGA is built on the same principles as the OD Directive, and together they create a complex codified environment, setting an EU-wide minimal common denominator for the legal regulation of PSI access and reuse. Art. 4 of the regulation prohibits exclusive arrangements between data providers and data reusers, with the possibility for exceptions in the cases of necessity "for the provision of a service or the supply of a product in the general interest that would otherwise not be possible." Art. 5 sets out basic rules of how the information should be provided. Art. 6 deals with the question of fees, and it is build on the same framework as is the case of the OD Directive. Art. 7 and 8 then envisage a system of regulatory bodies and the creation of a national single information point that will help with accessing the information.[47] The DGA is a general regulation, which presents an identical set of rules for any data regardless of the specific area or data space. We can expect more proposals governing area-specific data in the coming months.

2.4 Chapter summary

This report focuses on legal instruments at the international and European level, because they illustrate the basic outlines and context of the relevant regulation shared throughout the EU and its Member States. Based on this presumption, it is possible to conclude this chapter by stating that on the international level, there is effectively no single shared legal framework covering access to PSI. The right of access to PSI bound up with the freedom of speech, but the specifics of this legal concept are different in every country, with very limited consolidation due to the case law of the European Court of Human Rights. The Council of Europe's Convention on Access to Official Documents (CETS No. 205) might help in this regard in the future, but as of this moment, its relevance is very limited due to the small number of contracting parties. Furthermore, on the level of national law, there may be *lex specialis* legal instruments covering specific areas of information. Therefore, the question of whether the specific information can be provided must be resolved on a case-by-case basis in accordance with national law.

In the second relevant area, the reuse of PSI (including open data), the situation is much better on the European law level. The EU's PSI Directive set out a sound initial standard of legal

[46] For more, see, e.g., https://www.go-fair.org/fair-principles/.
[47] A more detailed analysis of the DGA will be available in the final version of this report, when more detailed versions of the regulation appear.

requirements for the quality of information and data publication; the OD Directive has made this standard even higher. For this reason, it is possible to expect the same minimal requirements for data quality in all countries in the EU and EEA. It is too soon to evaluate the quality of implementation.[48] Furthermore, we can expect there to be some differences in the way the implementation is done. The OD Directive sets a minimal standard, but the Member States can still choose to implement a higher, more open, one. Finally, commission Implementing Regulation No. 2023/138 regarding list of high value datasets was published on 21 December 2022 and will come into effect in June 2024. This regulation specifies which specific data from categories set in Annex of the directive 2019/1024 must member states make publicly available online in the open data format. Furthermore, the regulation sets out the basic technical requirements for the interoperability of these data. These are, however, very general. Thus, it will be necessary to specify the detailed requirements through technical standards.

[48] Art. 18 of the OD Directive anticipates the Commission carrying out an evaluation of the directive no sooner than July 2025.

3. Protection of data & intellectual property rights

Intellectual property rights is one of several obstacles that need to be dealt with during the publication and reuse of PSI. If an institution provides any content that is protected by any kind of intellectual property rights, that provider must license it properly (if possible) to ensure the effective reusability of the content. On the other hand, as will be discussed below, in a lot of situations the provided content will be not protected by any intellectual property rights at all. In such cases the data provider will not need to license the content in the strict meaning of the word "license," in the sense of a contract that allows a third party to use content protected by intellectual property rights. Furthermore, the provider cannot license the content (in the strict meaning of the word) because there is no protected content to be licensed. However, the data provider must still declare and set the terms of use.[49]

Relevant international normative documents consisting of multilateral treaties regarding copyright protection include:

- The Berne Convention for the Protection of Literary and Artistic Works;
- The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) (WTO);
- The WIPO Copyright Treaty.

Relevant European Union documents regarding copyright and database protection include:

- Directive 2001/29/EC on the harmonization of certain aspects of copyright and related rights in the information society;
- Directive 2006/116/EC on the term of protection of copyright and certain related rights (codified version);
- Directive 96/9/EC on the legal protection of databases;
- Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

All European Union member states are parties to international treaties mentioned above and are bound by the European Union directives. This creates a harmonised legal framework for the whole European Union and European Economic Area of the pertinent aspects of copyright. The following analysis builds on this harmonised framework. However, it is important to note that there may be differences among the specific regulatory frameworks of individual Member States.

[49] For more detail, see Chapter 5.

3.1 The protection of plain data

Before addressing issues of IP rights protection and how these issues interact with PSI and open data, it is necessary to first address the issue of the legal regime of so-called "plain data."[50] Plain data are data that are not protected by any other legal regimes regulating information and its use; these include as intellectual property law, protection of trade secrets, personal data protection, handling of classified information, etc. A particularly pressing and interesting question is whether there is such a thing as "ownership" of plain data, i.e., whether plain data is capable of being the subject of property rights.

There is currently no regulation at the international or the European level to address this issue. In legal doctrine, however, the debate about data ownership is quite lively.[51] The possibility of creating a specific "Data Producer's Right," which would at least partially resolve the issue of an absolute right to plain data, has been raised by the European Commission in its communication on "Building a European Data Economy."[52] The idea behind this proposal was to increase legal certainty and encourage the reuse of data, and motivate markets to create new data. However, this proposal was met with several critical voices. Montagnani[53] argued that, first, data is a very unstable object of protection due to the speed of its creation and consumption. Second, the proposal did not sufficiently define the "rightful producer" of data. Finally, there is great risk in overlap between the new right and currently existing protection frameworks, such as copyright and *sui generis* database rights. Nevertheless, the European policymakers' efforts to address these problems can be seen in the draft of the European Data Strategy[54] and, in particular, the forthcoming Data Act.[55]

---

[50] There is a theoretical question concerning different possibilities of data and information legal regulation. Even though the concepts of "data" and "information" hold different meanings, legislatures quite often use these terms interchangeably. Furthermore, the regulation of information (e.g., what a person can or cannot do with specific information) often primarily affects "data" that is in a context that actually begins to constitutes "information." However, although this issue is important, it lies beyond the scope of this research report. For this reason, at this point we only refer to some select relevant sources that deal with the topic. On the theoretical concept of "information," see Pieter Adriaans, 'Information' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University 2013); Luciano Floridi, *Information: A Very Short Introduction* (Oxford University Press 2010); Michael Keeble Buckland, 'Information as a Thing' (1991) 42 Journal of the American Society for Information Science and Technology 351.

[51] See, e.g., Lee A Bygrave, 'Information Concepts in Law: Generic Dreams and Definitional Daylight' (2015) 35 Oxford Journal of Legal Studies 91; Martin Fadler and Christine Legner, 'Who Owns Data in the Enterprise? Rethinking Data Ownership in Times of Big Data and Analytics' (2020) Proceedings of the European Conference on Information Systems (ECIS) 1; Maria Lillà Montagnani and Antonia von Appen, 'IP and Data (Ownership) in the New European Strategy on Data' (2021) 43 European Intellectual Property Review 156.

[52] European Commission, Communication on Building a European Data Economy SWD (2017) p. 13.

[53] Lillà Montagnani and von Appen (n 51) 161.

[54] Online: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

[55] Online: https://digital-strategy.ec.europa.eu/en/policies/data-act.

In the context of EU law, the most critical thing to mention is the "Regulation on a framework for the free flow of non-personal data in the European Union."[56] While this Regulation governs certain obligations that Member States have with respect to plain data, it does not address the issue of data ownership at all. Of course, this is because the definition of ownership and the objects that are eligible for ownership is a matter of private law, which is mainly regulated at the level of Member States. To the best of our knowledge, and within the limits of this report, we are not aware of a case of a jurisdiction granting plain data status as something protected by property law.[57] We cannot, of course, rule out the possibility that such arrangements exist somewhere.[58] For this reason, it is essential to investigate the situation in a particular country before starting the process of providing open data. However, we consider the presence of such an arrangement to be highly unlikely.

3.2 Copyright protection

3.2.1 Legislative overview

In the context of the legal regulation of public sector information and open data, we may encounter cases where the content provided, or parts thereof enjoy copyright protection. Copyright is an absolute right that protects the results of the author's creative activity (works of authorship) in such a way that it excludes anyone else[59] from any profiting from and influencing the copyrighted work. A work of authorship is thus understood as the personal expression of a natural person, which corresponds to a dual concept of protection that consists of a combination of moral[60] and commercial rights. It is worth mentioning that moral rights have not been codified at all by any European Union directives, and thus the scope of their protection depends on legal order of each Member State.

Art. 2 section 1 of the Berne Convention, a cornerstone of international copyright protection, defines protected copyrighted work (or "authorial work") as follows:

[56] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.
[57] For a good overview, see Andreas Boerding and others, 'Data Ownership - A Property Rights Approach from a European Perspective' (2018) 11 Journal of Civil Law Studies 323, 334–345.
[58] Boerding claims that generally speaking, legal frameworks in European countries could support the concept of data ownership. ibid., 369.
[59] It should be noted that there are exceptions and limitations to copyright. The reason for their existence is that copyright is not intended to act as a tool for the monopolization of ideas. The exceptions and limitations are therefore intended to ensure the further dissemination of ideas and the growth of creativity. See, e.g., Christophe Geiger, 'Promoting Creativity through Copyright Limitations: Reflections on the Concept of Exclusivity in Copyright Law' (2009) 12 Vanderbilt Journal of Entertainment and Technology Law 515.
[60] Art. 6 of the Berne Convention.

> The expression "literary and artistic works" shall include every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression, such as books, pamphlets and other writings; lectures, addresses, sermons and other works of the same nature; dramatic or dramatico-musical works; choreographic works and entertainments in dumb show; musical compositions with or without words; cinematographic works to which are assimilated works expressed by a process analogous to cinematography; works of drawing, painting, architecture, sculpture, engraving and lithography; photographic works to which are assimilated works expressed by a process analogous to photography; works of applied art; illustrations, maps, plans, sketches and three-dimensional works relative to geography, topography, architecture or science.

In the context of access to PSI and open data, literary works expressed in speech or writing, photographic works, visual works, architectural works, and cartographic works are especially relevant.

EU copyright law does not provide for a clear-cut general definition of copyrighted work.[61] However, the CJEU has laid down basic preconditions that an intangible creation must fulfill to qualify as a copyrighted work.[62] The basic standard is that the work must be original, i.e., ''the author's own intellectual creation''.[63] Consequently, a work must not be copied and must be a result of the intellectual creative activity of the human author.[64] The work must thus reflect personal choices of its creator. These are not present[65] when the expression is dictated by rules,[66] technical function,[67] and technical considerations,[68] or by the information content itself[69], which prevent the creator from expressing her "creative abilities in the production of the work by making free and creative choices" [70] – i.e., leaving no room for creative freedom. In the context of

[61] P Bernt Hugenholtz and João Pedro Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 International Review of Intellectual Property and Competition Law 1190, 1193.
[62] For an extensive debate what is protected as „work" in the EU see Caterina Sganga, 'The Notion of "Work" in EU Copyright Law after Levola Hengelo: One Answer Given, Three Question Marks Ahead' (2019) 41 European Intellectual Property Review 415.
[63] As ruled in the CJEU decision of 16. 7. 2009, No. C-05/08 (Infopaq International). This requirement was adopted as the basic standard in the subsequent topical cases Levola Hengelo (CJEU decision of 13. 11. 2018, No. C-310/17), Funke Medien (CJEU decision of 19. 7. 2019, No. C-469/17), Cofemel (CJEU decision of 12. 9. 2019, No. C-683/17) and Brompton Bicycle (CJEU decision of 11. 6. 2020, No. C-833/18).
[64] Hugenholtz and Quintais (n 61) 1196.
[65] These constraints on creativity were identified by ibid 1198.
[66] E.g., a football match where the game rules are applicable (joined Cases C-403/08 and C-429/08 Premier League, para. 98).
[67] For example, the functionality of a graphical user interface as presented in the CJEU decision of 22. 12. 2010, No. C-393/09 (Bezpečnostní softwarová asociace), para. 49-50.
[68] CJEU decision of 11. 6. 2020, No. C-833/18 (SI and Brompton Bicycle Ltd), para. 26.
[69] CJEU decision of 29. 7. 2019, No. C-469/17 (Funke Medien), para. 24.
[70] As required according tothe CJEU decisions of 19. 7. 2019, No. C-469/17 (Funke Medien), para. 19 and of 7. 3. 2013, No. C-145/10 (Painer), paras 87–88.

PSI, it is noteworthy that the fact that "mere intellectual effort and skill" were required for the creation of the intangible result has no influence on its copyrightability.[71] Artistic merit or aesthetic quality are also not required.[72] Moreover, economic investment as such is no guarantee of protection by copyright law.[73] Finally, as noted by Hugenholtz and Quintais, the work must be expressed "in a manner which makes it identifiable with sufficient precision and objectivity"[74] – ideas that have not been materialized in a form or a shape are not "works."[75]

In the context of access to information, analyses and expert opinions are frequently cited as relevant examples of works, but works may also include other copyrighted works, as in the case of works held by libraries and museums.[76] In the context of open data, copyrighted works may appear as part of a provided dataset, but the dataset itself may also be a work. An example of the former is a database of entries to a literary competition organized by a local governmental agency, or a database of expert opinions on various issues commissioned by a public authority.[77] One practical example would be a database of tourist destinations, such as the one run by CzechTourism, a public sector entity,[78] which includes photos and descriptions of interesting places and destinations. An example of the second variant could be map documentation of one of these destinations, which is a cartographic work.

In addition to the above examples, a database itself may be protected as a copyrightable work. This follows from Article 3 of Directive 96/9/EC which states, "In accordance with this Directive, databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright. No other criteria shall be applied to determine their eligibility for that protection." A database is defined as "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."[79] It is important to note that copyright protects the structure of the database as such, i.e., the selection and arrangement of the content. The content itself is not protected in this way.[80]

[71] CJEU decision of 19. 7. 2019, No. C-469/17 (Funke Medien), para. 23.
[72] Hugenholtz and Quintais (n 61) 1197.
[73] Ibid.
[74] CJEU decision of 13. 11. 2018, No. C-310/17 (Levola Hengelo), para. 40.
[75] Hugenholtz and Quintais (n 61) 1199.
[76] The goal of ensuring reusability of content held by, e.g., public libraries was one of the main legislative aims of 2013 PSI directive amendment.
[77] In this variant, the difference between the work itself and the metadata that describes it must be taken into account. Only the work itself would be copyrighted, but the dataset would certainly include metadata such as the identity of the author, date of publication, etc., which are not protected.
[78] In Czech online: https://www.kudyznudy.cz/.
[79] See Art. 1 (2) Directive 96/9/EC.
[80] See Art. 3 (2) Directive 96/9/EC.

The author of the database is the person who organized the content. The conditions for granting copyright protection to a database are therefore that: a) it is the author's own intellectual creation; and b) the selection of the elements and their arrangement must be the result of creative activity. As Hugenholtz points out,[81] the first condition was clearly addressed by the Court of Justice of the European Union in Case C-604/10 (Football Dataco), in which it held that the "criterion of originality is satisfied when, through the selection or arrangement of the data which it contains, its author expresses his creative ability in an original manner by making free and creative choices… and thus stamps his 'personal touch'."[82] This criterion is not met "when the setting up of the database is dictated by technical considerations, rules or constraints which leave no room for creative freedom."[83] The second condition (that the selection of elements is a creative activity) is well demonstrated by Hugenholtz's example, according to which a list of the author's favorite restaurants in Amsterdam could enjoy protection, while a list of the most expensive restaurants probably would not.[84] A database that meets the above conditions for granting copyright protection is referred to as an "original database."

Some jurisdictions may include an exception for governmental works. Generally, this exception ensures that certain content, which would otherwise fulfill the criteria of copyrighted work, is not protected by copyright because of its importance for the public interest. For example, the Czech copyright act addresses such an exception as follows:

> Protection under copyright law does not apply to an official work, which includes legal regulations, decisions, measures of a general nature, public documents, publicly accessible registers and collections of the documents therein, or official drafts of official works and/or other preparatory official documentation, including official translations of such works, parliamentary and senate publications, commemorative books by a municipality (municipal records), state symbols and/or symbols of a unit of local self-government*, and other such works for which there is a public interest in exclusion from protection.*[85]

The possible existence of such exception is quite important, because if the provided content (which would otherwise fulfill the criteria of copyrighted work) falls within its scope, it can be provided and further reused without any obstacles and without a need for any licenses.

---

[81] P Bernt Hugenholtz, 'Directive 96/9/EC' in Thomas Dreier and P Bernt Hugenholtz (eds), *Concise European copyright law* (Second edition, Kluwer Law International 2016) 392 <https://media.wolterskluwer.com/pdfs/SampleChaptersPDF/6651.pdf>.
[82] Para. 38 of CJEU case No. C-604/10 (Football Dataco and Others).
[83] Ibidem, para. 39.
[84] Hugenholtz (n 81) 393.
[85] Sec. 3 letter a) Act. No. 121/2000 Sb., copyright act. Translation by the authors, emphasis by the authors

3.2.2 Copyright and open data

When it comes to the context of the publication and reuse of PSI that is in some way restricted by third-party copyright (even though these situations will not in fact be very common), there are two areas which must be taken into account. First, the distinction between access and reuse is absolutely crucial. In the event of a conflict with copyright, there may be a situation where, based on certain copyright limitations or exceptions, information can be disclosed even though it is copyrighted. For example, some countries have enacted an exception for the use of a work for an official purpose. The disclosure of information undoubtedly qualifies as fulfilling the "official purpose" criterion, as it is a decision about the right to information (and thus, by implication, the right to freedom of expression). If the disclosure of a copyrighted work is proportionate and passes the three-step test as defined, for example, by the Berne Convention, the obliged entity may disclose the work to the applicant upon request. However, this does not mean that the applicant can freely continue to use it without further permission, as the statutory license no longer entitles him or her to do so. Thus, the applicant may, at most, make use of other legal grounds (statutory licenses and copyright limitations), but may not redistribute or exploit the work in any other way without further permission. For this reason, relying solely on exceptions for the further use of PSI is inappropriate, and if any of the content provided is copyrighted it is necessary to license it (see more later in Chapter 5 of this research report).

The second area is connected to copyrighted databases. When providing open data, it is possible to encounter copyrighted original databases. We think that this is more likely to be the case when the obliged entity creates the database and provides data on a discretionary basis, because in the case of mandatory provision, the selection of elements in the database is determined by law and therefore lacks room for creative freedom. However, even for these databases, protection can result due to the possibility of an original arrangement of the elements prescribed by law. This is even more true as the database structure becomes more complex, such as when it interconnects multiple information resources. With regard to the scope of STIRData project, we can assume that a majority of such databases will be protected by a copyright, (or more specifically, their structure will be protected). However, that does not necessarily mean that this protection will also extend to the open-data export of the content of such databases. The export may be in a different structure than in the source database, and in such case the copyright protection of the source database would not be infringed. The same is true for situations when the data reuser accesses the source database via API (application programming interface) without copying the structure of the original database. Again, if the units of data are later stored in structure that differs from the original structure, no copyright protection would be infringed. However, when the copyright protection conferred because of the way the *content* of the source database were selected, a

different outcome may result. In any case if there is copyrighted content, the open-data provider should publish it only under a proper open license.[86]

3.3 *Sui generis* database rights

The *sui generis* database right, or the right of the maker of the database, has its roots in European Directive 96/9/EC. As can be seen from Recitals 10 to 12 of the directive, the main purpose of the *sui generis* database right is to protect the investment made in its creation.[87] The above-mentioned definition of a database[88] applies not only to the case of an original copyrighted database, but also to databases protected by *sui generis* right. This also means that a given database may be protected by both means, either of them, or none of them.[89] In other words, protection of a database via copyright is entirely independent of protection via *sui generis* rights.

The "maker" of the database is defined as the legal or natural person who takes the initiative and the risk of investing.[90] This is the first fundamental difference from copyright, because if a database provided by a third party achieves the conditions for the creation of a *sui generis* right, this right is created directly for the customer and not for the external provider (e.g., an IT company that provides database solutions for a public-sector body). As Hugenholtz has pointed out, in order to be granted the status of a maker of the database, it is essential that the entity in question both invested in the creation of the database and instigated its creation.[91]

The right itself is defined in Art. 7 of Directive No. 96/9/EC as the right to "to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database." Even though the main purpose of *sui generis* right is to protect the investment in the database,[92] it is clear that this right has effect on the content of the database, regardless of whether the content itself is protected in any other way. This right is absolute; no one can without permission interfere with the content by extracting or re-utilizing a substantial part of the database. Furthermore, the protection applies not only to the primary

[86] For more on this topic, see Chapter 5.
[87] See also Hugenholtz (n 81) 402.
[88] Art. 1 (2) of Directive 96/9/EC reads as follows: a "database shall mean a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."
[89] See CJEU decision of 15. 1. 2015, No. C-30/14 (*Ryanair Ltd v. PR Aviation BV*).
[90] Rec. 41 of Directive 96/9/EC.
[91] Hugenholtz (n 81) 403.
[92] See CJEU decision of 3. 6. 2021, No. C-762/19 ('CV-Online Latvia' SIA), in which the court stated that extraction and reutilization of the content within the meaning of the provision of Art. 7 is when there is a "risk to the possibility of redeeming… investment through the normal operation of the database in question."

database, but also to copies and exports of the database.[93] However, the maker may grant another person the right to exercise this right with a license.

The term "extraction" is practically identical with the concept of copying of the database content. A broad interpretation would also lead to the conclusion that the application of the concept would also include, for example, the creation of temporary copies for the purpose of displaying the database on a computer screen, since this activity is not covered (unlike copyright) by any of the exceptions to the *sui generis* database right, with the result that such activity must be subject to the consent of the maker of the database.[94]

The term "reutilizing" means any activity that makes the contents of the database available to the public. In the Innoweb decision, the Court of Justice of European Union (hereafter the CJEU) held that the operator of a "metasearch engine," a service that redirects users' queries to other search engines and offers the user results from specialized databases through this process, reutilizes the content of those databases and their substantial parts.[95]

For example, in the context of open data, database extraction is the downloading of a copy of a database file to the data storage of the data reuser. Database re-utilization is then the creation of an application that interacts with the database and also allows third parties to access its contents, e.g., via API.

It follows from the above that the *sui generis* database right of the database maker indirectly protects the content of the database. However, it does not protect individual data, but only becomes applicable when a substantial part of the database is affected. Unlike copyright, the *sui generis* right does not have a personality component, it can be waived, and it is transferable.[96]

Directive 96/9/EC states in Art. 7 (1) that the *sui generis* database right protects a database "which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents." A "quantitative" investment is one that consists of quantifiable resources, such as time and money;[97] in contrast, a "qualitative" investment consists of non-quantifiable effort, such as mental effort or energy expended. Examples of qualitative input, according to Hugenholtz, would be a web designer's skills in creating an online database, or a lexicographer's knowledge in creating keywords.[98] The assessment of when an investment is "substantial" must be made on a case-by-case basis; it is not

[93] See para. 52 of CJEU decision of 9. 11. 2004 No. C-203/02 (The British Horseracing Board).
[94] See Hugenholtz (n 81) 406.
[95] CJEU decision of 19. 12. 2013 No. C-202/12 (Innoweb).
[96] See Art. 7 para. 3 Directive No. 96/9/EC.
[97] See Recitals 7, 39, and 40 of Directive 96/9/EC and para. 44 and 46 of the CJEU decision of 9. 11. 2004, No. C-338/02 (*Fixtures Marketing Ltd v. Svenska Spel AB*).
[98] Hugenholtz (n 81) 404.

entirely clear when an investment is sufficiently substantial, i.e., what the threshold for a substantial investment is.[99] Recital 19 of the directive gives some slight direction on this matter:

> …as a rule, the compilation of several recordings of musical performances on a CD does not come within the scope of this Directive, both because, as a compilation, it does not meet the conditions for copyright protection and because it does not represent a substantial enough investment to be eligible under the sui generis right…

Recital 7 also hints at this, stating that "...the making of databases requires the investment of considerable human, technical and financial resources."

The case law of the CJEU has not helped in this matter either.

Substantial investment must be made in the obtaining, verification, or presentation of the contents of the database, taking into account the total input to these three components. Obtaining consists of locating existing independent elements and placing them in the emerging database. However, the costs for obtaining the content do not include the costs necessary to create the elements.[100] The content verification criterion consists of the resources expended by the database builder to verify the veracity and accuracy of the data already present in the database. The investment in the presentation of the database content consists of an investment in resources in "the resources used for the purpose of giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in that database and the organisation of their individual accessibility."[101] Hugenholtz mentioned the digitalization of analog files, the creation of a thesaurus, or user interface design as examples of such activity.[102] In our opinion, another such activity would be the creation of an API (application programming interface) that could be used to access the contents of the database in an automated way. One practical example of a substantial investment regarding the presentation of content in the context of open data would be the cost of 2,000,000 CZK (€85,100) incurred by the Ministry of Finance in modifying the Czech Republic's ARES register[103] to allow open data from it to be published.[104]

---

[99] Ibid.
[100] See Hugenholtz (n 81) 405. Also see para. 42 of the CJEU decision of 9. 11. 2004 No. C-203/02 (The British Horseracing Board) and para. 53 of the CJEU decision of 9. 11. 2004, No. C-338/02 (*Fixtures Marketing Ltd v. Svenska Spel AB*).
[101] Para. 43 CJEU decision of 9. 11. 2004, No. C-338/02 (*Fixtures Marketing Ltd v. Svenska Spel AB*).
[102] Hugenholtz (n 81) 405.
[103] ARES is the Access to Registers of Economic Subjects, which contains data from registers of legal and physical persons, company ownership, and other data.
[104] Slížek D. Dodám vám systém na otevření dat ARES za 1 Kč, nabízí ministerstvu Michal Bláha [I will give you the ARES open-data system for 1 Kč, Michal Bláha offers the Ministry of Finance]. Lupa.cz [online], 2017 [last accessed 30. 6. 2023]. Online: https://www.lupa.cz/aktuality/dodam-vam-system-na-otevreni-dat-ares-za-1-kc-nabizi-ministerstvu-michal-blaha/.

The OD Directive brought a major change in the context of a *sui generis* database right and its application during PSI and open data publication and its reuse. As mentioned earlier, Art. 1 (6) OD Directive states: "The right for the maker of a database provided for in Article 7(1) of Directive 96/9/EC shall not be exercised by public sector bodies in order to prevent the reuse of documents or to restrict reuse beyond the limits set by this Directive." For a time, it was disputed whether public sector bodies could even acquire *sui generis* database rights. The purpose of the *sui generis* right is to protect the investment in the database; public sector bodies generally do not need that, since they only handle public money, and they only create databases to fulfill legal duties. As a result, there is not a risk to any "investment" in the typical sense.[105] The provision of Art. 1 (6) pragmatically reflects practice in the Member States, which allow for such protection to exist. However, at the same time it effectively cancels any possibility to enforce *sui generis* database rights in a way that would prevent any kind of future reuse of the data. Therefore, the data providers must license or waive[106] the *sui generis* right when they publish the protected database.

At first sight, it may seem that this entire chapter, which has dealt in detail with the *sui generis* database right, is superfluous because of the impossibility of enforcing this right. And that it would suffice to simply state that EU law shows that this right cannot constitute an obstacle to further use of PSI. However, the provision of Art. 1 (6) OD Directive does not preclude the existence of such a right. Therefore, if the database of a public-sector body (or another data provider falling within the scope of the OD Directive) fulfills the requirements of substantial investment in the obtaining, verification, or presentation of its contents, the *sui generis* database right will objectively exist, and it will protect such a database. It is necessary to properly license such databases to give reusers legal certainty. Without a correct license, reuse of the protected database in the context of open data applications and further activities would be illegal, regardless of the provision of Art. 1 (6) OD Directive.

3.4 Chapter summary

When providing open data, it is possible to encounter several legal obstacles created by intellectual property rights related to the content. These obstacles need to be overcome by proper licensing; without it, unlimited reuse of the provided data is not possible. First of all, it should be noted that most of the content provided as part of open data from public registers will be in the form of "plain data." Data is "plain" if it is the type of data that is not subject to any specific form of legal protection, such as intellectual property rights, trade secrets, personal data, etc. The legal

[105] Hugenholtz noted that on the basis of this argument, the Dutch Council of State (*Raad van State*) refused to recognize a *sui generis* database right of the City of Hamburg. See Hugenholtz (n 81) 405.
[106] A waiver of the *sui generis* database right should be applicable in the most jurisdiction of EU and EEA Member States. It is not excluded in the directive 96/9/EC and since the database right stems directly from this directive, it is safe to assume that national legislators did not include this kind of exclusion.

assessment of plain data may vary from jurisdiction to jurisdiction, as there is no European or international harmonization of which specific types of data are "plain." Within the scope of the limitations of this research report, we can conclude that plain data is not protected by property law, and therefore not subject to absolute rights of ownership. In view of this, there is no legal barrier present that would *per se* limit their provision and reuse.

In the area of intellectual property rights, the chapter identified three rights that may be relevant in the provision and reuse of PSI and open data. The first is copyright protection of the *content* of the database in question. This will be a relatively atypical and uncommon type of case, because it requires that the content provided meets the standard for a copyrighted work. However, if such a situation does happen to arise, it is essential to license the content properly.

The second is the copyright protection of the *database* itself. In this case, the copyright specifically protects its structure and the way its content is selected. In the case of public databases, the structure of the database is more relevant for protection, where the intellectual work of its creators is protected, rather than on the selection of its content. This is because the content will often be based directly on the requirements of legislation, and the choice of content will therefore not offer any scope for intellectual activity itself. Copyright protection of a database does not protect the content as such. It is therefore theoretically possible to allow the sharing and reuse of content without the need for licensing, as long as this does not also copy the structure of the source database. Nevertheless, in order to ensure legal certainty, we recommend that the database be properly licensed if this right exists.

The third is the *sui generis* right of the database maker. Although this protection is mainly concerned with the investment made in the source database, it also indirectly affects the content of the database. If content is provided from a database protected by a *sui generis* database right, it must be licensed; otherwise, the recipient will be restricted in the further use of the data provided. This requirement therefore follows directly from Art. 1(6) of the OD Directive, which states that any *sui generis* database right must not constitute an obstacle to the further use of the data provided.

All three types of protection can be combined. At the same time, it is quite common that the dataset provided is not protected by any of these rights. Assessing the level of copyright protection (if any) should be conducted on an *ad hoc* basis directly by the open-data provider before providing the open data. The details of how to license the content are discussed in Chapter 5 of this research report.

4. Personal data protection

4.1 Privacy and personal data protection: An overview of the regulation

4.1.1 Introduction to the topic and an overview of the relevant legal instruments

The right to privacy and personal data protection has traditionally conflicted with the right of free access to information and its reuse. This situation stems from the very basic principles underlying the rights in question. While the right of access to information respects the essential character of information and presupposes its dissemination in accordance with the principle of publicity, the protection of privacy and, in particular, the protection of personal data requires the opposite. This is particularly true for personal data protection, whose systems are based on the prevention of damage, the control of data, and the effort to ensure that such data is not used in violation of the rights of data subjects.[107] Both PSI and the Open Data Directives note this opposition, and therefore they include provisions stating that the directive is without prejudice to Union and national law on the protection of personal data.[108] This does not mean that personal data cannot be provided in open-data quality. However, when dealing with publication and reuse of PSI, it is still necessary to take into account the legal framework of personal data protection as a whole.[109]

One basis for the right to data protection is the right to informational self-determination.[110] This guarantees every person the possibility to determine how information about his or her person will be treated.[111] This is not an absolute right and there are exceptions to it, for example in the form of the processing of information by state authorities. In general, however, the right to informational self-determination gives each person the possibility to determine how he or she will present him or herself in public, what information he or she wants to be known about him or her and, ultimately, what information he or she wants to receive.

Data protection legislation pursues two objectives. The first is to ensure that the right to protection of personal data of natural persons (data subjects) is not infringed. Data protection

---

[107] See, e.g., Raphaël Gellert, 'Understanding Data Protection as Risk Regulation' (2015) 18 Journal of Internet Law 3.

[108] See Art. 1, section 4 of the OD Directive.

[109] For a general overview, see, e.g., Cristina Dos Santos, 'On Privacy and Personal Data Protection as Regards Re-Use of Public Sector Information (PSI)' (2013) 6 Masaryk University Journal of Law and Technology 337; Frederik Zuiderveen Borgesius, Jonathan Gray and Mireille van Eechoud, 'Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework' (2015) 30 Berkeley Technology Law Journal 2073.

[110] See, e.g., Theo Hooghiemstra, 'Informational Self-Determination, Digital Health and New Features of Data Protection' (2019) 5 European Data Protection Law Review 160; Orla Lynskey, 'Deconstructing Data Protection: The "Added-Value" of a Right to Data Protection in the Eu Legal Order' (2014) 63 International & Comparative Law Quarterly 569.

[111] The right was *inter alia* asserted by the decision of German Federal Constitutional Court from 15. December 1983 No. BvR 209/83, BVerfGE65, which concerned a case of the publication of an unsatisfactorily anonymized dataset of public census records.

legislation regulates the behavior of obliged entities in relation to the personal data they process. If someone decides to handle personal data, in the vast majority of cases, processing will also take place. As a consequence, the data controller must comply with the obligations arising from the legislation. From the point of view of data protection law, it does not fundamentally matter much who the personal data belongs to.

Data protection law regulates how personal data is treated. It is a preventive instrument, based on the premise that if the data controller complies with the obligations imposed on him, the risk of damage, harm, or misuse of personal data caused by the processing will be reduced. As a result, other fundamental rights that could be affected by the mishandling of personal data are indirectly protected. The first of these rights is that of privacy, as the risk of disclosure of sensitive information about a person will be limited if personal data is handled properly. However, other examples of rights indirectly protected in this way include the right to property (the mishandling of personal data can lead, for example, to identity theft and the misuse of payment cards) or the prohibition of discrimination (discriminatory practices can occur on the basis of the mishandling of personal data). Data protection law is essentially like an umbrella protecting other rights that could be violated by the processing of personal data.

The second objective of the legislation is generally to facilitate processing of personal data that is lawful and that respects the protection of the rights of data subjects.[112] Data protection legislation is pragmatic, because it is based on the assumption that the processing of personal data happens and is generally appropriate for society to do it. A popular saying claims that personal data is the new oil because it enables economic exploitation as well as progress. Data protection legislation is therefore not intended to prohibit the processing of personal data altogether. Processing is generally possible – as long as the data controller approaches it responsibly so as to minimize the risks that the processing may pose to the data subject. The two objectives mentioned above are dynamically complementary. The object of any processing of personal data is to strike a balance between them.

Legal instruments regarding personal data protection on international level relevant for this analysis are:

- The European Convention for the Protection of Human Rights and Fundamental Freedoms
    - Art. 8 protects the right to a private and family life

[112] See, e.g., Christopher Hood, Henry Rothstein and Robert Baldwin, *The Government of Risk: Understanding Risk Regulation Regimes* (Oxford University Press 2001); Raphaël Gellert, 'Data Protection: A Risk Regulation? Between the Risk Management of Everything and the Precautionary Alternative' (2015) 5 International Data Privacy Law 3.

- The Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108)

Legal instruments regarding personal data protection on European Union level relevant for this analysis are:

- The Charter of Fundamental Rights of the European Union (Document No. 2010/C 83/02)
    - Art. 8 reads as follows:

      Protection of personal data

      1. Everyone has the right to the protection of personal data concerning him or her.

      2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

      3. Compliance with these rules shall be subject to control by an independent authority.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, hereafter the GDPR)
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)

4.1.2 An overview of the basic concepts

This section presents basic concepts that are necessary for the following legal analysis.

The term "**personal data**" is defined in Art. 4 para. 1 of the GDPR as

> Any information relating to an identified or identifiable natural person ("**data subject"**); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Personal data is essentially any information that can directly or indirectly lead to the identification of a natural person. Personal data protection legislation does not apply to legal persons. "Directly

identifying" personal data is information or records which, directly from their own context, are clearly capable of identifying a person. This includes, for example, a name, a permanent residence, an identity card number, or a telephone number. "Indirectly identifying" data, on the other hand, includes data points that cannot on their own lead to the identification of a person, but if combined with other data in the right context, can be used to identify a person. The definition of personal data is extremely broad due to this definition, as it can cover an extremely large amount of information. This includes, for example, IP addresses that identify users' personal devices, which can be used in criminal investigations. It should be stressed that indirectly identifying personal data is indeed personal data, even if a specific data controller cannot use it to identify a specific person at the time. The CJEU set down in its *Breyer* decision (C-582/14) the limits of scope of what is personal data that it is no longer personal data if "the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and manpower, so that the risk of identification appears in reality to be insignificant."[113] This is essential for the notion of anonymization.

Anonymous data is data that has undergone a process of irreversible anonymization and can no longer be used to identify a specific person. Therefore, such data can no longer considered personal data.[114] However, it is more appropriate to think of use of anonymization techniques as a scale between identification and absolute anonymity, because the more anonymous the data, the less informative it is, and vice versa.[115] Based on the decision in the *Breyer* case, we can consider data to be anonymous when it would clearly be disproportionately difficult or costly to reidentify a person with it given the state of the art. Furthermore, in the context of open data, there is always a risk of reidentification, because of the free access to the data and the possibility of its reuse and the ability to connect it with other datasets. Moreover, evolving technology may also allow reidentification in datasets in future instances, where it had previously not been possible.[116]

Pseudonymous data can be found where direct identifiers are replaced by indirect identifiers.[117] This is, for example, the substitution of a number for a name and surname. However, as the above shows, it is still personal data, albeit more secure. Generally, we can say that the application of anonymization techniques to a dataset creates pseudonymous data, and once it is practically impossible to reidentify the original data subjects, we can then consider it anonymous data. It is

[113] See para. 46 of the CJEU decision of 19. 10. 2016 No. C-582/14 (*Breyer*).
[114] See Recital 26 of the GDPR.
[115] Similarly, see Paul M. Schwartz and Daniel J. Solove, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 New York University Law Review 1814.
[116] For more information on anonymization and its legal consequences and shortcomings, see e.g., Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701.
[117] Art. 4 para. 5 of the GDPR.

useful, especially in the context of open-data publication and reuse, to consider anonymity and pseudonymity as a scale, rather than as a number of discrete states. The following figure shows this relationship:

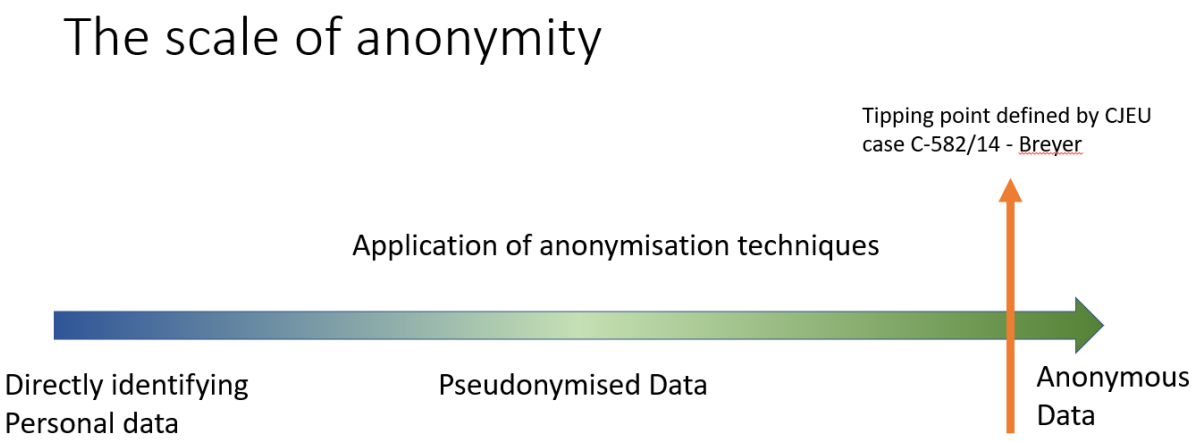

Figure 1: The scale of pseudonymity and anonymity

Special categories of personal data are defined and enumerated in Article 9 of the GDPR. It includes

> personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

The processing of this type of data is generally prohibited, unless one of the specific exemptions listed in Art. 9 para. 2 applies. This is because these categories of personal data pose an increased risk to data subjects and their rights, due to the possibly discriminatory nature of this data. Generally, these special categories of personal data cannot be published as PSI (not to mention in open-data quality) because it would constitute too great an interference with the rights of data subjects.[118]

A "data controller" is defined in Art. 4 para. 7 of the GDPR as a "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data." The purpose of the processing is an absolutely essential element in the overall context of the legal framework regarding the protection of personal data. The purpose that the controller has identified at the beginning of the processing is then measured against how the processing can lawfully take place, how long the data can be kept, who can have

[118] See the CJEU decision of 22. 6. 2021 No. C-439/19 (*Latvijas Republikas Saeima*).

access to it, and so on.[119] The data controller often processes data for a variety of purposes and is thus accountable for multiple processing processes.

The "data processor" is defined in Art. 4 para. 8 of the GDPR as "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." The main difference between the data controller and the data processor is that data processor does not set the purpose of processing, but has to follow the purpose set by the controller.

Personal data processing is defined in Art. 4 para. 2 of the GDPR as

> Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

As with personal data, the definition of processing is extremely broad. In general, it is any activity that a controller or processor carries out that uses personal data during the data's life cycle. It is useful to think of processing of personal data procedurally – i.e., over time. Thus, a single processing process, which is defined by a previously identified purpose, may comprise several sub-activities of handling personal data.

Article 5 of the GDPR enumerates **basic principles** relating to the processing of personal data that are applicable in every instance of personal data processing. In the context of this general overview, the most important are the principles of lawfulness, fairness, and transparency,[120] the principle of purpose limitation,[121] the principle of storage limitation,[122] and the principle of accountability.[123]

An important part of principle of lawfulness is the requirement to state a specific **legal ground** for data processing, without which the data controller cannot even start with the processing. These legal grounds are listed in Art. 6 para. 1 of the GDPR. For the context of this analysis, the only legal grounds that are relevant for publication and the reuse of PSI (and open data) are the fulfilling of a legal obligation to which the controller is subject,[124] the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller,[125] and a

[119] Art. 5 para. 1 of the GDPR.
[120] Art. 5 para. 1 letter a), GDPR.
[121] Art. 5 para. 1 letter b), GDPR.
[122] Art. 5 para. 1 letter e), GDPR.
[123] Art. 5 para. 2.
[124] Art. 6 para. 1 letter c), GDPR.
[125] Art. 6 para. 1 letter e), GDPR.

legitimate interest pursued by the controller or by a third party.[126] Other legal grounds are not applicable in the context of the publication or reuse of PSI.

4.2 Open data as personal data processing

The OD Directive, similar to its predecessor, sets out in Art. 1 para. 2, point h) that the directive does not apply to

> documents, access to which is excluded or restricted by virtue of the access regimes on grounds of protection of personal data, and parts of documents accessible by virtue of those regimes which contain personal data the re-use of which has been defined by law as being incompatible with the law concerning the protection of individuals with regard to the processing of personal data or as undermining the protection of privacy and the integrity of the individual, in particular in accordance with Union or national law regarding the protection of personal data.

The specific implementation of this legislation is nevertheless up to individual EU Member States, some of which, such as Germany or Belgium, require complete anonymization of personal data, while others allow further use of personal data disclosed in this way according to certain legal authorization (for example, France, Slovenia, and the Czech Republic).[127]

A more theoretical question is whether personal data can be ever published as open data at all. This question relates to the fact that traditionally the term "open data" has meant data provided without any legal restrictions. At the same time, the processing of personal data in the European context will always present some kind of restriction arising from the relevant legal framework, the GDPR in particular, because data subjects cannot waive their rights, even if they want to. Therefore, it may seem that personal data cannot ever be published as open data (or in open-data quality).

In our opinion, it is necessary to understand the concept of open data in the broader context of European law. Should the stricter interpretation (that is, that "open data" must be without any restriction) be true, not even databases containing licensed copyrighted content could be considered open data, because there will be residual duties to attribute the work; after all, in many jurisdictions, the author cannot waive her or his rights. Therefore, the principle that open data must be provided without any legal restriction needs to be interpreted in a way that implies that data providers cannot impose new legal restrictions, and must do their best to minimize restrictions arising from the statutory law.

[126] Art. 6 para. 1 letter f), GDPR.
[127] Santos (n 109) 338.

Regardless of this theoretical question, it is necessary to note that there might be practical legal consequences stemming from the approach of European and national lawmakers and their decision about how to publish a dataset with personal data. For example, describing the legal characteristics of a published dataset vis-à-vis the concept of open data, or stating expressly that that the dataset is purposed for further reuse will help reusers to bear the burden of proof in regard to the legal grounds of any new data processing.[128]

In the context of open data, we distinguish between two types of data controllers. The first are data providers who publish PSI containing personal data on the basis of a legal obligation. The purpose of such personal data processing is therefore the relevant legal provision itself, or the direct consequence of that obligation. It is also theoretically conceivable for data providers to process data for a purpose of their own choosing (for example, voluntary publishing information to the public). However, in the case of such processing, it would be quite difficult to ensure its lawfulness, as will be discussed in detail later in this section. The second type of controllers are open-data reusers. These are usually private entities, app developers, or individuals who want to use the data to inform themselves about a given issue. The purposes of processing published personal data can vary widely, and it is up to these new data controllers to determine these purposes in accordance with the GDPR and comply with all the obligations imposed on them by data protection legislation.

4.3 The open-data provider as data controller

Anytime an open-data provider, (be it a public-sector body or a public undertaking) provides personal data in an open-data format, it clearly puts itself in the position of a data controller and therefore it has to fulfill the obligations arising from the GDPR. This means that several variations of the situation have to be taken into account. The first question is when a data provider can publish personal data, if at all. Only after these cases have been identified can one consider whether it is possible to provide it in a qualitatively better way, i.e., as open data.

The first fundamental question is whether an open-data provider can decide, at its discretion, to publish personal data. This issue directly follows from a question addressed in Chapter 2 of this report, whether the data provider has the legal standing to provide the data in general. If the data provider were to provide personal data on a discretionary basis as open data, it would first have to determine the appropriate purpose of such processing of personal data, and to justify such processing by one of the legal grounds in accordance with Art. 6 of the GDPR. Determining the purpose for such processing is difficult, but not impossible. This could be, for example, to ensure law and order in the case of the publication of a list of dog owners by a municipality, or for

[128] This will be more discussed further in Part 4.4 of this report.

transparency, where lists would be provided with information on the recipients of welfare benefits and aid that the municipality has decided to pay. However, for most of these examples, the GDPR does not offer legal basis for such processing of personal data. Since these would be voluntary disclosures, not directly based on law, or which would have to be made in the course of the performance of a public authority's tasks, the legal grounds of Art. 6 para. 1 points c) and e) cannot be applied. Nor is it possible to rely on "legitimate interest" as listed in Art. 6 para. 1 point f), because this legal basis cannot be used for data processing carried out by public authorities in the performance of their tasks. Theoretically, the data provider could acquire consent from the data subjects and later rely on the legal grounds listed in Art. 6 para. 1 point a), but this option would be not realistic for practical and technical reasons in most of situations. Therefore, we can conclude that it is not possible to publish personal data as open data voluntarily based on the discretion of data provider, and thus this possibility remains outside of scope of the OD Directive.

One way to overcome this problem is anonymization. The data provider, most likely the public-sector body that owns the relevant database, might be able[129] to conduct anonymization of the data and publish the results as statistical or aggregated data. Once the data has been anonymized, (i.e., it would clearly be disproportionately difficult or costly to reidentify given the state of the art)[130] the data provider can publish it online. However, the anonymization process must be done very diligently, due to the risk of reidentification[131] and possible further high risk to the rights and interests of data subjects.

The situation is different when the data provider has a legal obligation to provide personal data as PSI (or directly as open data).[132] In this situation, the OD Directive will indeed apply. However, even in these cases, the data provider is still in the position of a personal data controller. The purpose of such processing (provision of data) stems from the law under which the processing is carried out. In the case of the above example, the purpose is, for example, transparency and ensuring the possibility to contact the persons responsible for the management of a commercial company. The legal grounds for the processing of personal data in these cases are processing resulting from a legal obligation or the performance of a task carried out in the public interest or

[129] In cases where it is allowed by the law.
[130] See para. 46 of the CJEU decision of 19. 10. 2016 No. C-582/14 (*Breyer*).
[131] For more on the risks of reidentification, see Ohm (n 116). For more on the legal issues of anonymization, see WP29 Opinion No. 5/2015, on Anonymisation Techniques, online [last accessed 30. 6. 2023]. Online: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
[132] One practical example from the Czech Republic of such data is information from the registers of persons established by Act 304/2013 Sb., on public registers of legal and natural persons. Section 7 para. 1 of the Act stipulates that data from the six registers defined in the Act (the Association Register, the Foundation Register, the Register of Institutions, the Register of Unit-Owners' Associations, the Commercial Register, and the Register of Public-Benefit Corporations) must be provided by the Ministry of Finance upon request to the extent specified in the act., compulsorily

in the exercise of official authority within the meaning of Art. 6 para. 1 points c) and e) GDPR. It is therefore necessary to have this legal obligation – imposed on the data provider – to have the possibility to publish personal data. It should be stressed that the processing in question must also always be strictly limited to the purpose and wording of the enabling legislation.

When the data provider publishes personal data, the technical means by which such publication takes place has an enormous effect in terms of ensuring the protection of the rights of data subjects.[133] In terms of potential interference with the right to privacy, it makes a difference whether the information – personal data – is provided in the form of a freely downloadable complete database file (with open data being only a better version of this) or whether the individual data entries are only accessible individually through a technical measure, such as by filling in a form on a website. Of course, in the latter case it would be theoretically possible to save the registry gradually using an automatic script. But this activity first of all requires more effort, and for a lay user of the Internet quite difficult. Second, the reuser in such a case would have considerably more difficulty in fulfilling the conditions imposed on them by the GDPR as a data controller.[134] The particular method of publication is also important because it can create legitimate expectations among data subjects about how their data is treated (how much it is protected), which is one of the essential aspects of assessing the legitimacy of the data's possible subsequent use.[135]

Providing personal data in the form of open data is potentially the riskiest way to disclose information, due to the easy technical possibility to use and misuse the data. This fact was highlighted, for example, in the report of the UN Special Rapporteur on the right to privacy.[136] The WP29[137] addressed the issue of the relationship between personal data and PSI (hence open data)

[133] Technical barriers to publication were addressed by WP29 in Opinion No. 6/2013, ("on open data and public sector information ('PSI') reuse"), online, p. 10-11 [last accessed 30. 6. 2023]. Online: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf.

[134] For more on this topic, see Section 4.4 of this report.

[135] As will be discussed in more detail in the following section, the main legal grounds on which a data controller may rely if that person wishes to reuse data from public registers is the legitimate interest of the controller in the sense of Art. 6 para. 1 letter f) of the GDPR. One of the criteria which intervene in the assessment of the legitimacy of the intended processing is the degree of expectation of the data subject that such processing may take place. For more on this, see, e.g., WP29 Opinion No. 6/2014, on the notion of legitimate interests of the data controller under Art. 7 of Directive 95/46/EC [last accessed 30. 6. 2023]. Online: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[136] Report on Big Data and Open Data from 17 October 2018, No. A/73/438, p. 11 [last accessed 30. 6. 2023]. Online: https://undocs.org/A/73/438.

[137] WP29 was a working group created on the basis of Art. 29 of Directive 95/46/EC and was later transformed into the European Data Protection Board when the GDPR came into force.

in two opinions;[138] while they did not expressly rule out the possibility of providing personal data in this way, they strongly warned of the risks that such processing of personal data may entail. In Opinion No. 3/2013, which addressed the principle of purpose limitation of processing, WP29 stated:

> Once personal data are publicly available for reuse, it will be increasingly difficult, if not impossible, to have any form of control on the nature of potential further use, be it for historical, statistical, scientific or other purposes. This is especially the case if the data are available in digital, searchable and machine readable format and have been published on the internet, hence, the selection of the information that will or will not be made publicly available becomes all the more important.[139]

In both opinions, the WP29 referred to the need to carry out a careful impact assessment of the processing of personal data,[140] i.e., what effects the disclosure of personal data may have on data subjects. When assessing the impact, it is useful to consider, for example, whether and what form of anonymization or pseudonymization has taken place vis-à-vis the data, because this may act as a technical solution to increase the level of protection of personal data.[141]

The solution to the question of how the publication of information – personal data – should be technically (qualitatively) carried out is surprisingly poorly addressed, both legislatively and doctrinally. If (as is typically the case) legislation has not specified the method of disclosure itself, i.e., does not determine a specific norm to be followed, it is necessary to follow general norms. On the EU level of regulation, these general norms are Art. 5 of the OD Directive and the relevant provisions of the GDPR. A key aspect in assessing the qualitative manner in which PSI/personal data is to be provided is therefore the purpose of such processing, i.e., an assessment of why the information should be provided. As mentioned above, the purpose of the processing is derived from the legal provision that the obligation to disclose the information is based on. Decisions about the technical details of the disclosure method must be based on the nature of the personal data provided, the risk that its misuse may pose to data subjects, and other factors.[142] In doing so,

[138] WP29 Opinion No. 6/2013, on open data and public sector information ('PSI') reuse and WP29 Opinion No. 3/2013, on purpose limitation, [last accessed 30. 6. 2023]. Online: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.
[139] WP29 Opinion No. 3/2013, on purpose limitation, p. 49-50.
[140] WP29 Opinion No. 6/2013, on open data and public sector information ('PSI') reuse p. 20 and WP29 Opinion No. 3/2013, on purpose limitation, p. 50.
[141] Borgesius, Gray and van Eechoud (n 109) 2114–2121.
[142] See WP29 Opinion No. 3/2013, on purpose limitation, p. 10-12.

the principle of 42minimization of interference states that the processing should be such that it is the least intrusive on the rights of the data subject in fulfilling the purpose.[143]

The publication of PSI/personal data in the form of open data requires that the publication be consistent with the purpose for which the information was provided. This purpose must be discernible from the statutory formulation of the obligation to provide certain PSI/personal data, or from the formulation that outlines the public-administration activity in which such provision is to take place. Therefore, if the legislator explicitly states that the specific PSI/personal data is to be provided as open data, the data provider must comply with this obligation.[144] If the law is not as precise, e.g., it merely states that certain PSI/personal data must be published online, the data provider should evaluate the character of provided data, the purpose of publishing such data that is enshrined in the law, and if possible follow the requirements of Art. 5 of the OD Directive (or more precisely, its national implementation).[145]

To summarize this section, the data provider can only provide PSI/personal data when it has a legal duty to do so. Optimally, the legislation will state precisely that the data should be provided in open-data quality. But even without this specification, the level of data quality as set out by Art. 5 of the OD Directive is almost at the level of open data. Still, it is important to be stress that there are limits on legislatures as well about what data can be legally made public for further reuse. These limits have been established by the case law of the CJEU. In *Volker und Markus Schecke GbR and Eifert v. Hessen*,[146] the court stated that a law requiring overly detailed publication of data regarding recipients of agricultural subsidies constituted a breach of the right to privacy and the right to data protection. Later, CJEU decided three cases, that are very important in regards of the national legislator's possibilities to set a legal duty for data publication. The first one is Case C-439/19 – *Latvijas Republikas Saeima (Points de pénalité)*. In it the court categorically denied the possibility of publishing PSI concerning "special categories" of personal data in the sense of Art. 9 of the GDPR for further reuse.[147] The second one is Case C-184/20 – *Vyriausioji*

[143] A good example from the Czech Republic is access to data from the cadaster. Let us assume that the purpose of this provision is to enable contact with the owner of the property and verify who the owner of the specific property is, but not to enable easy verification of how many properties someone holds. The technical setup of the described system should thus allow a specific property to be traced via the form, but prevent a search based on the property owner's identifier.

[144] For example, this could have been the situation of the implementing regulation No. 2023/138. Would the European law maker decided that high value dataset of "Companies and company ownership" also contains data on company ownership including personal data, it would be a legal duty for the data publishers to do so.

[145] That is, to provide such data "by electronic means, in formats that are open, machine-readable, accessible, findable and re-usable, together with their metadata. Both the format and the metadata shall, where possible, comply with formal open standards."

[146] CJEU, decision of 9. 11. 2010, joined cases C-92/09 and C-93/09 *Volker und Markus Schecke GbR and Eifert v. Land Hessen*.

[147] CJEU, decision of 22. 6. 2021, C-439/19 – *Latvijas Republikas Saeima (Points de pénalité)*.

tarnybinės etikos komisija.[148] The CJEU ruled that while the online publication of data from declarations of private interests is legitimate in relation to the fight against corruption, it does not meet the requirements of the balancing test, in particular taking into account the general principles of data protection such as the principle of minimisation. Finally, the third one is the joint cases C-37/20 a C-601/20 – *Luxembourg Business Registers*.[149] In this decision the CJEU ruled that he obligation to disclose information on the beneficial owners of companies within the meaning of Article 30 of the amended Directive 2015/849 is in breach of the Charter of Fundamental Rights of the EU, given that, although this regulation is an appropriate tool in the fight against corruption, it is not strictly necessary (as there is a less invasive solution) and at the same time it does not offer sufficient guarantees for the effective protection of fundamental rights and therefore disproportionately interferes with them. These decisions set limits for legislatures on what they can legislatively require with respect to publishing PSI/personal data.

4.4 The open-data reuser as a personal-data controller

This part addresses legal issues of the data reuser's role as a personal-data controller, albeit very briefly because this topic is not within the primary focus of this report. Once the PSI/personal data is published in the form of open data, anyone can take it and process it. However, PSI does not cease to be personal data simply because it has been published. It cannot be accessed and then immediately used for any purpose whatsoever, because it is still fully subject to the legal framework of personal data protection. Exceptions to this general rule are situations where the GDPR does not apply, as Article 2 para. 2 sets out. Within this paragraph, point c) is relevant for open data; it concerns the processing of personal data carried out by a natural person in the course of a purely personal or household activity.[150] Like all exceptions to the scope of GDPR, processing data solely for personal use should be interpreted as strictly as possible. The narrow interpretation of this exception was also later upheld by the CJEU in the *Lindqvist* case,[151] where the court held that the publication of personal data on a website did not constitute processing for personal or domestic activities, and in the *Ryneš* case, where it held that the use of a CCTV system to monitor the surroundings of a house did not fall within this exception, despite the fact that only the owner of the property had access to the footage.[152] As a practical example of the use of open data while maintaining the exception of processing personal data solely for personal use, it is possible to conceive of a case where the interested person downloaded the provided information

[148] CJEU, decision of 7. 10. 2022, C-184/20 – *Vyriausioji tarnybinės etikos komisija.*
[149] CJEU, decision of 22. 11. 2022, C-37/20 a C-601/20 – *Luxembourg Business Registers*.
[150] Art. 2 para. 2 c), GDPR.
[151] CJEU decision of 6. 11. 2003, C-101/01 – *Lindqvist*.
[152] CJEU decision of 11. 12. 2014, C-212/13 - *Ryneš*.

and analyzed it to inform themselves or find out interesting facts, search for connections, etc. However, they would not be allowed to disseminate this information in any way. This is not a typical use of open data. Given this, the applicability of the exception for open data is very limited.

A common way of dealing with open data is to create various applications and other services that work with the data, which are then offered to end users. If the dataset used contains personal data, the creator of the application is in the position of a data controller, as they determine the purpose of the processing. The purpose may vary significantly depending on the nature of the specific application. Mere "use of data in an application" is insufficient to serve as a purpose, as the purpose must be sufficiently specific.[153]

Another issue regarding the purpose of data processing relevant from the open-data perspective concerns possible difficulties with the interpretation of Art. 6 para. 4. Some authors have interpreted this paragraph to mean that it applies even in cases of personal-data transfer by a new controller.[154] This would mean that the (new) user of open data (the creator of the new application) would be bound by the (original) purpose (or compatibility with it) for which the data was originally published when determining the purpose of the new processing. However, in our view, this interpretation is not correct, because the purpose-limitation principle applies to processing carried out by one and the same controller who established the original purpose for collecting the personal data.[155] In the context of the app creator, the collection of personal data occurs at the moment of download from the data provider's server. We therefore assert that Art. 6 para. 4 of the GDPR should be read as a provision permitting certain types of processing that would otherwise be prohibited by the purpose-limitation principle, rather than as a provision further restricting the processing of personal data by third parties.

Once the purpose of the processing of personal data has been determined, the creator of the application working with personal data must provide a legal justification that allows such processing of personal data. Of course, in theory, the data subject's consent is an option,[156] but it will be extremely technically difficult for the data controller to secure it. In practice, therefore, this legal basis is not applicable. From the range of legal grounds listed in Art. 6 para. 1, the only

[153] See, e.g. WP29 Opinion No. 3/2013, on purpose limitation.

[154] See, e.g., Michal Nulíček and others, GDPR v otázkách a odpovědích. *Bulletin-advokacie.cz* [online]. Published 3. 11. 2017 [last accessed 30. 6. 2023]. Online: http://www.bulletin-advokacie.cz/gdpr-v-otazkach-a-odpovedich. It should be added that the same team of authors did not state this opinion in their following publication – Michal Nulíček et al., *GDPR - obecné nařízení o ochraně osobních údajů* (Wolters Kluwer 2017) 140–144.

[155] The same view has been elegantly articulated by Nonnemann, who tried to argue the opposite position in his interpretation, but arrived at the absurd result of the *de facto* impossibility of assessing the first two conditions introduced by Art. 6 para. 4 of the GDPR. See (in Czech) František Nonnemann, 'Zpracování Veřejně Dostupných Osobních Údajů a GDPR' (2018) 26 Právní rozhledy 167, 169.

[156] See Art. 6 para. 1 letter a) of the GDPR.

practical option available to developers of open-data applications with personal data is the legitimate interest of the controller or of a third party set out in point f). Legitimate interest may be the exercise of any right or activity that is generally allowed by law. However, the data controller may rely on this legal justification only if the protection of the rights and interests of the data subject do not take precedence over the declared legitimate interests. Therefore, it can be argued that there is a small institutional (intra-systemic) proportionality test present, which every data controller must assess before starting to process data. It is then the case that the lower the risk of interference with the privacy or other rights and interests of the data subject (whether due to the nature of the personal data or due to their technical security or other aspects), the more likely the data controller may be to be granted these legal grounds.[157] It is important to stress that it is not only the fundamental rights and freedoms of the data subject that are at stake, but any interests in general, as is clear from the wording of Art. 6 para. 1, point f) of the GDPR when it states: "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data…" On the other hand, it is clear from the text of the GDPR that European policymakers did indeed intend to include the possibility of applying this legal basis to a wide range of various situations, including direct marketing.[158]

Recital 47 of the GDPR mentions that the data subject should have a reasonable expectation that the intended processing of personal data may take place, and that the controller or third parties should take these expectations into account when assessing the possibility of applying the "legitimate interest" standard. For this reason, it is very helpful that in cases of the publication of PSI with personal data, the legislation formulates the intention of further reuse of the data. In view of the intra-system proportionality test, processing data for the purpose of the controller's legitimate interests of *de jure* constitutes effective protection against misuse of the published personal data. The test of balancing the legitimate interest of the controller against the rights and interests of the data subjects is fundamentally tilted in favor of the data subjects. However, the legitimate interest can nevertheless serve as a valid legal basis for further processing, especially in situations when the new data controller creates a new added value to the used content. For example, a mere republication of public database would not survive the test of legitimate interest.

[157] For more detailed analysis regarding the legal grounds of legitimate interest see, e.g., WP29 Opinion No. 6/2014, on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC [last accessed 30. 6. 2023]. Online: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.
[158] See Recital 47 of the GDPR.

However, if the new data controller can find new (still legal) uses for the data, which generally add value to what is already public, the balancing test will more likely end up in their favor.

The final aspect addressed in this chapter is the "information duty" of the data controller. In the context of open data, the GDPR's Article 14 sets out what information the controller must provide when personal data is not directly collected from the data subject (which is the case when working with open data). Paragraph 1 of Article 14 stipulates that the controller is obliged to notify the data subject about information concerning the identity of the controller, the purpose of the processing, the categories of personal data, and other information necessary to ensure fair and transparent processing of personal data. According to Recital 61 of the GDPR, this notification must be provided at the time of the collection of the personal data by the controller. Information duty is an essential part of the data-protection legal framework.[159] Its practical and functional performance is one of the biggest obstacles for effective reuse of PSI with personal data. Due to the large amount of personal data of different data subjects that are processed in the course of such activities by the controller, there is no technically efficient way to comply with this obligation. We are therefore concerned that the current method, whereby the creator of an application using open data with personal data notifies those affected by publishing this information about the processing of personal data on its website or in the application documentation, is inefficient due to the lack of outreach in getting this notification to data subjects.

GDPR offers two exceptions which might be helpful in this situation. The first one is a provision in Article 11 which reads as follows:

> If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.

However, data subjects often are indeed fully identified or identifiable if data from the relevant registers is published in open-data quality. Therefore, this provision will help only in situations when a data provider publishes effectively pseudonymized data that nevertheless still do not meet the level of security required for their full anonymization. On the other hand, a number of information sources, such as the commercial register, contain contact details, although often only

[159] See, e.g., Paul de Hert, Serger Gutwirth. Privacy, Data Protection and Law Enforcement: Opacity of the Individual and Transparency of the Power. In: Erik Claes, Antony Duff and Serge Gutwirth (eds), *Privacy and the Criminal Law* (Intersentia 2006) 77–78. See also Lynskey (n 110) 595; Gabriela Zanfir. Forgetting About Consent: Why The Focus Should Be On "Suitable Safeguards" in Data Protection Law. In: Serge Gutwirth, Ronald Leenes and Paul De Hert (eds), *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges* (Springer 2014) 245.

a physical address. Nevertheless, in such cases it would be possible to provide this notification, at least in theory.

The second exception can be found in Art. 14 para. 5, point b), which reads:

> Paragraphs 1 to 4 shall not apply where and insofar as the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 89(1) or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available [emphasis by the authors].

We can see purposes generally connected with the freedom of speech are expressly mentioned in this exception. And although other purposes are not, the exception is generally applicable to them as well. However, like every exception in the GDPR, it must be interpreted as narrowly as possible. Therefore, the question of whether there is a "disproportionate effort" must be answered with great detail and scrutiny. For example, if the dataset contains email addresses of data subjects, it would be hard to argue that sending notification information about the prepared processing constitutes a disproportionate effort. Thus, application of the second exception is quite questionable.

In our opinion, the best solution would be to enact a law that would create a register of all applications and services that process data taken from public resources, including open data. This register could be run, e.g., by the respective national data protection agency, or the institution that runs national data portal. If the law expressly states that the register can also contain information regarding existing applications, then it is possible to expect data subjects to know about it.[160] Even on the practical level, the main purpose of the information duty in the GDPR is to be able to know about ongoing processing of personal data. The existence of a central database of open data services would fulfill this purpose and, at the same time, would allow data subjects to practically check which applications use their data, and to counteract this if necessary.

[160] This is because of the legal principle that ignorance of the law is not an excuse to violate it.

4.5 Chapter summary

When it comes to intersection of PSI and personal data, several things must be taken into consideration. Firstly, the application of the OD Directive is limited to cases when personal data is lawfully provided to the public. A *sine qua non* condition of this is that there is a legal duty for the data provider to make the data public. However, in some cases, when national law conflicts with international instruments, namely European Convention on Human Rights as well as EU law, represented by the EU Charter of Fundamental Rights and the GDPR, or with the national constitution, not even a legal duty to make the data public would be sufficient – such duty would be unconstitutional and thus invalid in the long run.[161]

Second, when personal data is present in the process of publication of PSI (or more specifically, open data), all parties involved must fulfill any duties arising from personal-data protection legislation, specifically the GDPR, because they will occupy positions as data controllers. The data provider can publish a dataset that contains personal data only if there is a legal duty to do so. If there were no such duty, the data provider as a data controller would not have applicable legal grounds (Art. 6 para. 1 of the GDPR) for data processing. This insight is especially true when it comes to publication of personal data with open-data quality, because the risk of misuse or abuse of such data is higher than in the case of mere online accessibility. The data controller (i.e., the public-sector body providing the data) must also fulfill other duties arising from the GDPR, such as making sure that the data is provided in a way that will not endanger the source database and its integrity and that the data is correct and precise. However, these duties are manageable and achievable.

Third, recipients of personal data that is published in open-data quality are in a much more difficult position. They are in the position of data controllers, because they determine the purpose and the means of data processing. These must be lawful and sufficiently specific. Furthermore, the reusers must pass the balancing test, because legitimate interest (Art. 6, para. 1, point f)) is the only legal grounding that is applicable for this kind of personal data processing. The information duty represents another significant problem. It is very complicated, and almost impossible, to meaningfully inform data subjects about ongoing data processing in an application that uses personal data. The controller cannot simply send every data subject in a database a message, and a notification published on the webpage of the controller is not sufficient, because it does not fulfill the main purpose of information duty. Data subjects must have a chance to know about the ongoing processing so they can oppose it if necessary. A mere message on a webpage would be not enough because the data subject would not know where to look for that webpage in the first

[161] This might include a situation when the publication of data was not proportionate to the risk.

place. Thus, we suggest the implementation of an alternative solution. A central, publicly run database of applications and services that use open data datasets with personal data would solve this problem.

5. Terms of use of open data

When publishing PSI in the form of open data, it is necessary to add information about the conditions under which the data from the distribution can be used to each dataset, in the form of a metadata record. So-called "open terms of use", which ensure maximum reusability of content and minimum restrictions beyond the law, are a prerequisite for open data. Thus, the terms of use may include, for example, a license to allow the sharing of content protected by intellectual property rights, information about the presence of personal data, or conversely, information about the legal freedoms concerning the content provided. A technically appropriate way to specify the terms of use of open data may be the placement of a hyperlink to the document in the metadata record, where the terms themselves are expressed in a form understandable to the recipient of the data. Typical variants of the terms of use have included a link to standardized Creative Commons licenses, a link to a webpage with text informing readers about the open nature of the data provided, or a link to the provider's own terms of use displayed on its website. The main problem with this third group is that this method is not standardized in any way and is not machine-readable, which makes it difficult to work with the data afterwards. However, from a legal point of view, it is quite interesting. These types of terms of use are generally used in cases where the available PSI is not protected by any absolute IP rights. There are mainly two types of provisions in the text of the actual terms of use that are of legal interest. Statements of the first type are intended to set out various obligations for future recipients and reusers of the data. Statements of the second type are intended to exclude or limit liability for future damages that may arise in connection with the use of the data.

5.1 Contract

In a situation where no IP rights are involved during the publication of data, the data providers cannot rely on any license (*in stricto sensu*), because there is no content that could be licensed. However, the data can still be provided under some legally binding conditions. For the legal effect of these conditions, i.e., binding the recipient of the data to a legal obligation to behave or to tolerate something, there must be a contract concluded between the data provider and the data recipient. In the context of open data, it is appropriate to refer to such contractual arrangements as, for example, simple open-data contracts, as they do not involve the licensing of any content. It must be a contract, because a data provider cannot bind a data recipient to any performance merely by a unilateral declaration. Thus, in providing open data, it would be necessary to interpret the publication of the data and the publication of the conditions as an offer, with the beginning of the use of the data in line with these conditions as an implicit acceptance of the contract. The object of the performance is then the provision of the data on the part of the data provider, and the fulfillment of the conditions set out in the corresponding document on the other.

This part of this research report is necessarily limited by its scope. The question of the legality and conditions of a valid contract is subject to national law and is not harmonized in any way. Therefore, contract conditions must be analyzed in each jurisdiction separately.[162] From the case law of the CJEU, it is clear that a contract governing the provision of data is generally possible.[163] Generally, we can presume that the following condition should be met regardless of jurisdiction, as it stems from general legal principle of necessity of legal certainty. In order for the terms of use of open data to be understood as an offer to enter into a contract, which could then occur through the use of the data itself, it is necessary that data providers make it absolutely clear that what is at stake is an offer to enter into a contract. This can be achieved, for example, by formulating the text of the terms of use in such a way that it is clear at a first glance that the download and use of the data constitutes a contract with the provider. If the data provider does not comply with this requirement in such a way that makes it clear that the terms of use are also an offer to enter into a contract, the contract cannot be concluded, and any failure to comply with the provider's requests has no legal consequences. The other way to ensure that a contract is definitely concluded is to make the provision of data conditional on prior registration, which would remove the problem of non-addressability of the offer. However, this method contradicts the principle of open data, according to which access to data should be as simple as possible; for this reason, this report does not recommend this approach.

In any case, the contract must be within the limits set by the OD Directive. Art. 8 (1) stipulates:

> The re-use of documents shall not be subject to conditions, unless such conditions are objective, proportionate, non-discriminatory and justified on grounds of a public interest objective. When re-use is subject to conditions, those conditions shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition.

This provision sets a limit to any contract governing the reuse of provided open data. It sets a minimum standard of legal compatibility, as analogous to other provisions of the OD Directive that set a minimum standard of technical interoperability with the requirements of "formats that are open, machine-readable, accessible, findable and re-usable."[164]

In our opinion, the maximum appropriate level of restriction that can be accepted in order to still be open data is the obligation to give an attribution to the originator of the data. Anything above

[162] For example, Czech law generally demands identification of contracting parties. Thus, it is quite complicated to argue that a general contract can be closed in similar way as Creative Commons licenses, which have specific regulation in Czech law that expressly allows this licensing practice (but is applicable only when IP-protected content is present).
[163] See CJEU decision of 15. 1. 2015, No. C-30/14 (*Ryanair Ltd v. PR Aviation BV*).
[164] See Art. 5 (1) of the OD Directive.

that level is already contrary to the principles of open data and to the abovementioned provision of the OD Directive.

5.2 Open Licenses

As the findings of Chapter 3 of this report demonstrate, in most cases, there is no need to consider licensing in any way when providing open data because of the legal freedom of plain data and the fact that the data provider can license *in stricto sensu* only content protected by IP rights. Furthermore, the easiest way for PSI publication would be a case when we could apply statutory exceptions or limitations. Unfortunately, neither Directive 2000/29/EC nor Directive 96/9/EC offer any such exceptions and limitations that would be applicable for the reuse of open data.[165] One possible solution, at least for some reuse applications, can be found in Directive 2019/790 (the DSM Directive).[166] Articles 3 and 4 of the DSM Directive introduced exceptions for text and data mining, which the Member States have been mandated to introduce into their legal frameworks. Unfortunately, neither these exceptions are sufficient for the effective and complex reuse of PSI.[167] Therefore, licensing is the only viable option for securing the possibility of reusing PSI protected with IP rights.

There are three situations in which content licensing is possible and necessary for open data. A necessary prerequisite for all three is that the data provider is entitled to grant a license (or sub-license). If this were not the case, and a license were nevertheless granted, this would be acting contrary to the *nemo plus juris* legal principle.[168] A license may be granted in the following situations: i) the copyrighted work is part of the distribution of the dataset provided; ii) the structure of the database is copyrighted as an original database; iii) the content of the database to be provided is protected by a special right of the acquirer. In addition, these three options may be combined.

One suitable tool for licensing content protected by intellectual property rights may be what are known as "public licenses." The aim of public licenses is to ensure that the licensed content is open, i.e., that anyone can freely use the licensed work for any purpose, subject at most to attribution and openness. The public nature of the license is determined by its regulated contracting process. This takes the form of a non-addressed public offer to enter into a license agreement, which occurs implicitly at the beginning of the use of the work, subject to the terms

[165] For more on this argumentation in the context of database rights, see Jakub Míšek, 'Open Data, Open API and Database Rights' (2019) Jusletter IT 1.
[166] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.
[167] For more context, see, e.g., Jakub Míšek, 'Exception for Text and Data Mining for the Purposes of Scientific Research in the Context of Libraries and Repositories' (2020) The Grey Journal.
[168] This principle states that no one can transfer more rights than they have themselves.

and conditions set out by the licensor through publicly available license terms. This implies that no real interaction between the licensor and the licensee is necessary, and that there is no prior knowledge or limitation on the range of persons who may enter into the contract. It also means that if the licensee breaches the terms of the license, the license is lost and further use of the work is then unlawful. We can define a public license with a list of the following minimum defining characteristics:

- non-addressability and irrevocability;
- automatic implied acquisition of the license by use of the work;
- non-exclusivity;
- territorial, temporal, quantitative, and material unlimitedness;
- the right to share;
- royalty-free distribution;
- minimization of the licensor's liability and warranties;
- the condition of attribution.

There are a number of public licenses that data providers can use to license open data.[169] Good examples of a public license that is recommended by the general open-data community are the Creative Commons licenses, specifically the CC BY 4.0 and the CC0 1.0 waiver. Therefore, in this report we focus on these licenses as well. Creative Commons licenses are made up of three layers: i) the legal text; ii) a human-readable version summarizing the rights and obligations under the license (the "deed"); and iii) a machine-readable layer that allows automated evaluation of which license with which terms is being used.[170] One of the reasons why Creative Commons licenses are recommended as standardized licenses for open data publishing is precisely because their machine-readability enables the automated evaluation of licenses within open data catalogues and thus makes it easier for data recipients to work with them. The second advantage is their international familiarity and availability. Together, these two reasons significantly contribute to legal interoperability among different jurisdictions and different data providers. Open-data providers may choose to use other licenses (including their own). However, this will result in a loss of machine readability, and, especially if the license is written in a relatively uncommon language – such as Czech or Hungarian, for example – to deterioration of the ability to use the data

[169] See, e.g., Primavera De Filippi and Lionel Maurel, 'The Paradoxes of Open Data and How to Get Rid of It? Analysing the Interplay between Open Data and Sui-Generis Rights on Databases' (2014) International Journal of Law and Information Technology <http://ijlit.oxfordjournals.org/content/early/2014/10/16/ijlit.eau008>.
[170] See https://creativecommons.org/.

abroad. Therefore, we cannot recommend this approach and instead recommend that data providers use standardized public licenses.

The CC BY 4.0 (Attribution 4.0 International) license is the most open license in the Creative Commons family of licenses. This license is usable for both copyrighted works and *sui generis* database rights, and allow the recipient of the content to share, modify, and use the protected content for any purpose. The only conditions for use of the work under the CC BY 4.0 license is to give attribution. This means that that the origin of the work must be indicated, i.e., its title (or other clear identification) and the author, and furthermore there must be a reference to the text of the license.[171]

The CC0 1.0 is not a contract, but a unilateral waiver of absolute rights to the protected content.[172] In view of this, it cannot be applied as waiver in jurisdictions where the waiving of copyright is prohibited by law. If the CC0 license is wrongly used by the licensor to license the copyright work, Article 3 of the CC0 license will apply (the so-called "Public License Fallback"), according to which in cases where national law does not allow the waiver of an absolute right, CC0 is to be interpreted as granting the most open license (e.g., CC BY). However, the CC0 license can be used to waive the *sui generis* database right; in such a case, the *sui generis* right is irrevocably terminated. Because CC0 is a unilateral act, it can be used also as an information tool to communicate to data recipients that the content is not protected by intellectual property rights in any way. However, this approach may create some level of legal uncertainty, and therefore we cannot recommend it without reservations.

5.3 Providing information about the dataset

One important duty of an open-data provider is to maximize the legal certainty of the data recipient so the provided open data can be reused as easily as possible. A good way of fulfilling this duty is to prepare transparent and understandable "terms of use" for the provided dataset. The main purpose of the terms of use is to provide information about the legal status of published open data. However, in case legally protected content is present in the dataset, the terms of use should deal with this problem as well. From a technical point of view, a good way to communicate the terms of use to the data recipient is to add them to the metadata record of the provided open-data release.

The terms of use may include licenses that allow for the use of the content of the dataset. As the previous chapters of this report have shown, during the publication of open data, it is possible to

[171] See online: https://creativecommons.org/licenses/by/4.0/.
[172] See online: https://creativecommons.org/share-your-work/public-domain/cc0/.

encounter, albeit exceptionally, copyrighted works as part of the open data content, the copyrighted database (its structure or way of selection of the content), and *sui generis* database rights of the database maker. In cases where there is no content present in the dataset protected by IP law, it is nevertheless also advisable to inform the recipient of the data of this fact in order to reassure him of his legal certainty. In licensing, it is absolutely essential to ensure that open-data providers apply licenses only where possible and necessary, and that the licensed content is clearly identified. Granting a single "global" license for the entire dataset release is inadequate and confusing because it does not identify the exact content to which the license applies. This reduces the legal certainty of the recipient of the data. If in a particular case it is not possible to identify which content is specifically covered by the license, it would probably result in voiding the license due to ambiguity. One way to include adequately licenses in the metadata structure describing the distribution of the dataset is to split the "terms of use" section into three entries according to the intellectual property rights that may be involved in the publication of open data.

In addition to these intellectual property rights, personal data may also be present in the dataset. Therefore, the terms of use should also inform the recipient about the presence or absence of personal data, thus adding a fourth entry in the metadata record in the "terms of use" section. Each of these four entries would make it possible to clearly identify whether or not the listed protected content is present, or under which license its reuse would be allowed.

A possible proper construction of the terms of use is described by the licensing scheme in the Part II of this research report (see Figure 2).

5.4 Chapter summary

An essential part of providing open data is specifying the terms of use. The primary purpose of the terms of use is to inform the recipient of the data about the legal status of the dataset provided. However, the nature of the terms of use will vary depending on the type of content present in the dataset provided. If the dataset includes content protected by copyright (be it the content itself or a copyrighted database) or *sui generis* database rights, it is necessary to license such content to properly allow the reuse of such data. In this case, the terms of use will be constituted by a license agreement. One appropriate solution is to use public licenses such as Creative Commons CC BY 4.0 or a CC0 waiver. The great advantage of this solution is its wide international recognition and applicability.

If the open data provided does not contain any content protected by intellectual property rights, it is still possible to make the provision and reuse of the data subject to a data disclosure contract. Under this contract, the recipient commits to complying with the conditions set by the data provider. This solution raises three main problems. The first is primarily legal. It relates to the fact

jurisdictions have widely varying conditions that need to be fulfilled with respect to the formulation of the contract and the (indefinite) number and range of recipients to be validly concluded and therefore legally enforceable. A possible solution to this problem would be to introduce prior registration before the provision of data. However, this runs contrary to the principles of open data as such. The second problem is that the terms of setting up such a contract are limited by the requirements of the OD Directive and, in particular, must not restrict in any way the further use of the data provided beyond the scope of the OD Directive. The third problem relates to legal interoperability and its technical implementation. The custom contract and the custom terms of use will by definition not be standardized. They will also probably not be machine-readable and automatically processable. This makes the automatic reuse of the data provided much more complicated. In view of these problems, we cannot recommend the provision of open data on the basis of custom *ad hoc* contracts.

The main purpose of the terms of use is to inform the recipient of the data about the possible legal obstacles that may be associated with the provided content and, where appropriate, to provide solutions to them (licenses for content protected by intellectual property rights). The terms of use must therefore include a specification of what specific IP rights are associated with the data provided and how these rights are licensed. If the data provided contains personal data, it is also necessary to inform the recipients of the data of this fact so that they know that by processing it, they will be in the position of data controllers. Finally, if the data provided is truly just plain data without any accompanying legal protection, it is still advisable to inform the recipients of the data of this fact to reassure them of their legal certainty. Guidelines for the creation of the terms of use and a flowchart for this are contained in Part II of this report.

Part II: Guidelines and a licensing scheme

The following guidelines summarize the legal analysis presented in Part I of this report in a form of a checklist. To make further use of the data at European level a success, the reuser must have a high level of legal certainty that there will not be any legal obstacles for further use of the data. This issue, however, must be properly addressed by the data providers. The following checklist should serve as a tool that helps data providers to be sure that they have not neglected any important issue during the process of data publication.

Step 1: Access to the data

- Check regulation of access of information
    - What laws govern access and publication of information?
    - How has the Open Data Directive been implemented into the national legal order?
    - Is there any specific act that covers the publication of datasets of public registers of companies and other legal entities?
        - If yes, this act will generally be applicable.
        - If no, a general freedom of access to information or similar act will be applicable.
- Does the public-sector body have a legal duty to publish the data?
    - Is the data part of HVD in accordance with OD Directive?
    - Is there a national regulation that stipulates a duty to publish the data?
    - How precisely is the duty stipulated? (For example, is it general, like "Data from the register," or is it a more specific enumeration of data categories?)
- Does the public-sector body have discretion to publish data?
    - Can the public-sector body generally publish data without a specific legal obligation to do so?
        - This will apply in cases when there is not a duty to publish specific datasets, or in cases where there is such a duty, but the public-sector body would like to publish more than legislation requires.

Step 2: Intellectual property rights protection

- For a proper licensing scheme, it is necessary to identify protected content and make sure that the data provider can license it.
- Each dataset can be subject to several forms of intellectual property rights.

- Is plain data in your jurisdiction protected by any kind of intellectual property or property rights?
    - If no (which will probably be the majority of cases), this will not constitute an obstacle for the publication and reuse of data.
    - If yes, it is important to identify whether the data provider is the owner of the data (or whether they exercise the rights to the content).
        - If yes, the content can be provided and licensed under a public license (preferably CC0 or CC BY 4.0).
        - If no, the content cannot be provided.
- Does the dataset contain copyrighted content?

  For example, there are copyrighted literary or artistic works in the dataset, or the dataset is a map.
    - If no (which will probably be the majority of cases), this will not constitute an obstacle for the publication and reuse of data.
    - If yes, it is important to identify whether the data provider can exercise the rights to the content.
        - If yes, the content can be provided and licensed under a public license (preferably CC BY 4.0, or CC0 if your jurisdiction allows the provider to waive copyright).
        - If no, the content cannot be provided.
- Is the database itself protected by copyright?

  In other words, does the content fulfill the requirements that a) it is the author's own intellectual creation, and b) the selection of the elements and their arrangement are necessarily the result of creative activity?
    - If no (which will probably be the majority of cases), this will not constitute an obstacle for the publication and reuse of data.
    - If yes, it is important to identify whether the data provider can exercise the rights to the content (which will probably be most of the cases).

      In some cases, it is possible to provide the dataset without licensing the copyrighted database. It will depend on whether the export of the database directly reproduces its structure and the selection of elements. If not, then the exported dataset will not infringe the copyright.
        - If yes, the database can be provided and licensed under public license (preferably CC BY 4.0, or CC0 if your jurisdiction allows the provider to waive the copyright).
        - If no, the content cannot be provided.

- Is the database protected by a *sui generis* right?

  In other words, has there been a qualitatively and/or quantitatively substantial investment in the obtaining, verification, or presentation of the contents?

  This step is essential because a *sui generis* database right directly affects the provided content.

  - If no, this will not constitute an obstacle for the publication and reuse of data.
  - If yes, it is important to identify whether the data provider is the creator of the database (which will be the majority of the cases) or whether the data provider can exercise the rights to the content.
    - If yes, the database can be provided, but must be licensed for further reuse, preferably under public license (CC BY 4.0, or CC0 if the applicable jurisdiction allows creators to waive their *sui generis* database right).
    - If no, the content cannot be provided.

Step 3: Personal data protection

- If the dataset contains personal data, it is essential that this data is provided only in accordance with personal data protection legislation (primarily the GDPR) and that data reusers are properly informed about the presence of personal data, so they can effectively comply with their duties as data controllers.
- Does the dataset contain personal data?
  - If no, this will not constitute an obstacle for the publication and reuse of data.
  - If yes, is there a legal duty to provide such data as open data?
    - If yes, the dataset can be provided.
    - If no, the content cannot be provided.
- The data provider can lower the risk of misuse of the provided personal data by conditioning their provision with a conclusion of a contract, which would bind the reuser to use the data in a certain foreseeable way, e.g., the data provider can limit purposes of reuse. However, it must be noted that because such conditions have been added, it would not be possible to describe any information provided in this way as “open” data.
- If the personal data is not allowed to be provided, the data provider still can conduct a thorough anonymization process, which would render personal data non-personal. It is important to note that the anonymization has to be prepared thoroughly and diligently, because there is a higher risk of de-anonymization once the dataset is publicly and freely available.

Step 4: Preparation of Terms of Use

- Once the data provider identifies legal obstacles that could prevent the publication or reuse of the data, it is essential to communicate the results of their findings to the data reuser.
- The terms of use should contain information about any content that is subjected to any kind of information protection, whether it is copyright or personal data protection. It also must contain any solution which addresses such protection, such as licenses.
- Terms of use should be easily accessible and preferably part of the metadata of the dataset.
- For the sake of legal certainty, it is advisable to precisely identify what intellectual property rights are present in the specific instances of a particular dataset and how precisely these rights are licensed.
    - For example, the data provider can decide that copyrighted content is licensed with CC BY 4.0 public license and at the same time, that the *sui generis* database right is waived according to CC0.
- If the dataset contains personal data, the terms of use must contain this information. It is not possible to license personal data. The reuser will have to fulfill all the duties arising from the GDPR and national laws of data protection.
- If the dataset does not contain any protected content, the terms of use should also state this information for the sake of legal certainty.
- The data provider can provide the data under other conditions and after a conclusion of a contract. However, making additional obligations for data reuser contingent on a contract would lead to a situation in which the provided content would no longer be truly "open" data.
    - Furthermore, should the data provider decide to do so, firstly the provider must determine what the national law allows for and what conditions are for the legally valid conclusion of a contract online.

Flowchart of the licensing (terms of use) process

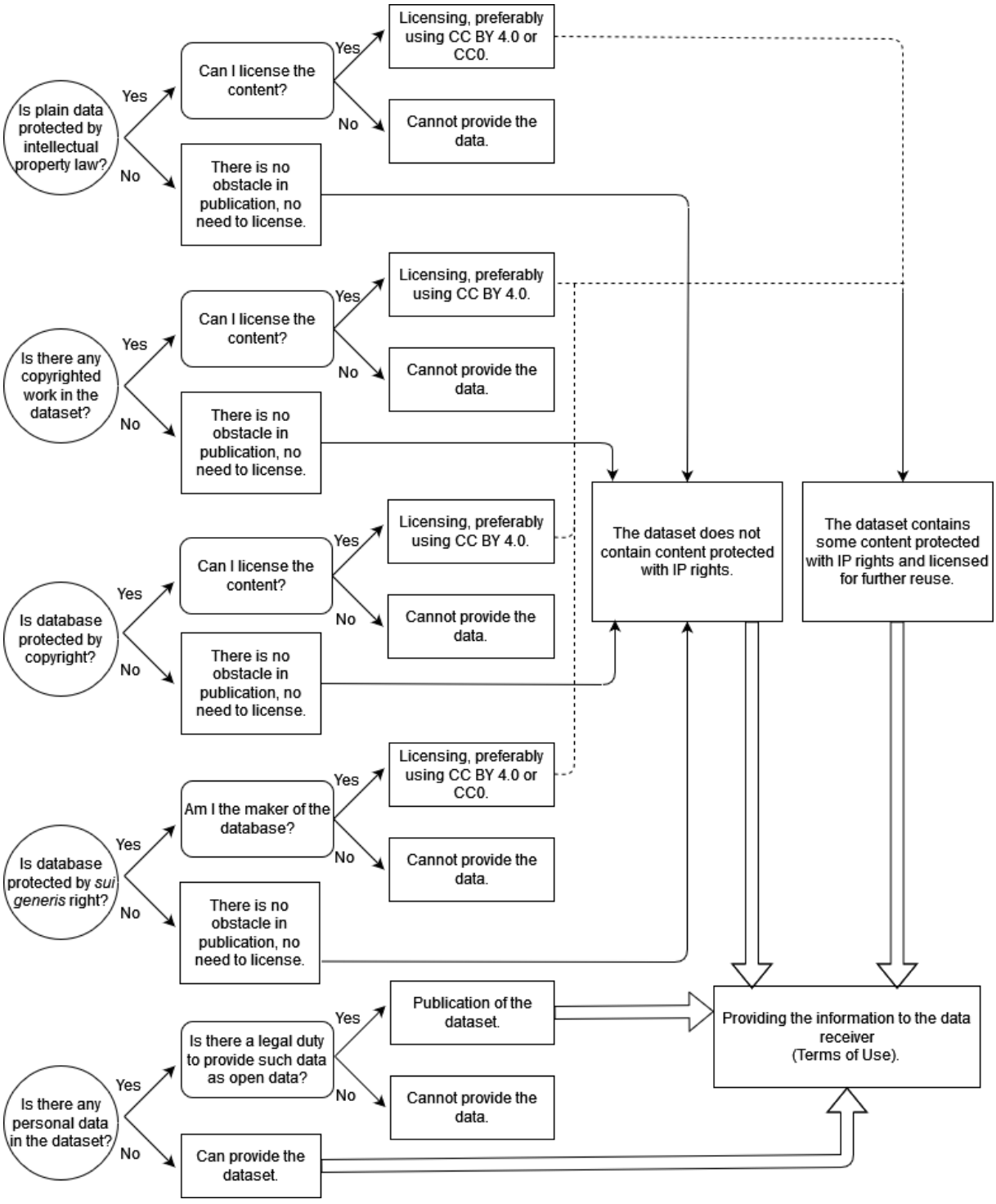

Figure 2: Terms of Use Flowchart

List of selected resources

Adriaans P, 'Information' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University 2013) <https://plato.stanford.edu/archives/fall2013/entries/information/> accessed 30 June 2019

Boerding A and others, 'Data Ownership - A Property Rights Approach from a European Perspective' (2018) 11 Journal of Civil Law Studies 323

Borgesius FZ, Gray J and van Eechoud M, 'Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework' (2015) 30 Berkeley Technology Law Journal 2073

Buckland MK, 'Information as a Thing' (1991) 42 Journal of the American Society for Information Science and Technology 351

Burkert H, 'Public Sector Information: Towards a More Comprehensive Approach in Information Law' (1992) Journal of Law and Information Science 47

Bygrave LA, 'Information Concepts in Law: Generic Dreams and Definitional Daylight' (2015) 35 Oxford Journal of Legal Studies 91

Claes E, Duff A and Gutwirth S (eds), *Privacy and the Criminal Law* (Intersentia 2006)

Eechoud MV and Janssen K, 'Rights of Access to Public Sector Information' (2013) 6 Masaryk University Journal of Law and Technology 471

Fadler M and Legner C, 'Who Owns Data in the Enterprise? Rethinking Data Ownership in Times of Big Data and Analytics' (2020) Proceedings of the European Conference on Information Systems (ECIS) 1

Filippi PD and Maurel L, 'The Paradoxes of Open Data and How to Get Rid of It? Analysing the Interplay between Open Data and Sui-Generis Rights on Databases' (2014) International Journal of Law and Information Technology <http://ijlit.oxfordjournals.org/content/early/2014/10/16/ijlit.eau008> accessed 4 December 2014

Floridi L, *Information: A Very Short Introduction* (Oxford University Press 2010)

Geiger C, 'Promoting Creativity through Copyright Limitations: Reflections on the Concept of Exclusivity in Copyright Law' (2009) 12 Vanderbilt Journal of Entertainment and Technology Law 515

Gellert R, 'Data Protection: A Risk Regulation? Between the Risk Management of Everything and the Precautionary Alternative' (2015) 5 International Data Privacy Law 3

——, ‘Understanding Data Protection as Risk Regulation’ (2015) 18 Journal of Internet Law 3

Gutwirth S, Leenes R and De Hert P (eds), *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges* (Springer 2014)

Hood C, Rothstein H and Baldwin R, *The Government of Risk: Understanding Risk Regulation Regimes* (Oxford University Press 2001)

Hooghiemstra T, ‘Informational Self-Determination, Digital Health and New Features of Data Protection’ (2019) 5 European Data Protection Law Review 160

Hugenholtz PB, ‘Directive 96/9/EC’ in Thomas Dreier and P Bernt Hugenholtz (eds), *Concise European copyright law* (Second edition, Kluwer Law International 2016) <https://media.wolterskluwer.com/pdfs/SampleChaptersPDF/6651.pdf>

Hugenholtz PB and Quintais JP, ‘Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?’ (2021) 52 International Review of Intellectual Property and Competition Law 1190

Lee M, Almirall E and Wareham J, ‘Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements’ (2016) 59 Communications of the ACM 82

Lillà Montagnani M and von Appen A, ‘IP and Data (Ownership) in the New European Strategy on Data’ (2021) 43 European Intellectual Property Review 156

Lynskey O, ‘Deconstructing Data Protection: The “Added-Value” of a Right to Data Protection in the Eu Legal Order’ (2014) 63 International & Comparative Law Quarterly 569

Míšek J, ‘Open Data, Open Api and Database Rights’ (2019) Jusletter IT 1

——, ‘Exception for Text and Data Mining for the Purposes of Scientific Research in the Context of Libraries and Repositories’ (2020) The Grey Journal <https://is.muni.cz/auth/publication/1608958/cs/Exception-for-Text-and-Data-Mining-for-the-Purposes-of-Scientific-Research-in-the-Context-of-Libraries-and-Repositories/Misek> accessed 28 November 2022

Morozov E, *To Save Everything, Click Here: The Folly of Technological Solutionism* (PublicAffairs 2013)

Nonnemann F, ‘Zpracování Veřejně Dostupných Osobních Údajů a GDPR’ (2018) 26 Právní rozhledy 167

Nulíček M and others, *GDPR - obecné nařízení o ochraně osobních údajů* (Wolters Kluwer 2017)

Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701

Peixoto T, 'The Uncertain Relationship between Open Data and Accountability: A Response to Yu and Robinson's the New Ambiguity of Open Government' (2012) UCLA Law Review Discourse 200

Ruijer E and Martinius E, 'Researching the Democratic Impact of Open Government Data: A Systematic Literature Review' (2017) 22 Information Polity: The International Journal of Government & Democracy in the Information Age 233

Safarov I, Meijer A and Grimmelikhuijsen S, 'Utilization of Open Government Data: A Systematic Literature Review of Types, Conditions, Effects and Users' (2017) 22 Information Polity: The International Journal of Government & Democracy in the Information Age 1

Santos CD, 'On Privacy and Personal Data Protection as Regards Re-Use of Public Sector Information (PSI)' (2013) 6 Masaryk University Journal of Law and Technology 337

Schwartz PM and Solove DJ, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 New York University Law Review 1814

Sganga C, 'The Notion of "Work" in EU Copyright Law after Levola Hengelo: One Answer Given, Three Question Marks Ahead' (2019) 41 European Intellectual Property Review 415

Thorhildur J, Avital M and Bjørn-Andersen N, 'Generating Value from Open Government Data', *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design* (2013)

Yu H and Robinson DG, 'The New Ambiguity of Open Government' (2011) 59 UCLA Law Review Discourse 178

Authors:

JUDr. MgA. Jakub Míšek, Ph.D.

JUDr. Radim Charvát, Ph.D., LL.M.

doc. JUDr. Matěj Myška, Ph.D.

Masaryk University

Žerotínovo nám. 617/9, 601 77 Brno, Czechia

Brno 2023