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

Briefing for national legislative implementation in the interest of supporting Open Access to research data and its re-use

This policymaking for Open Science was a collaborative effort between the Digital Curation Centre (DCC), EBLIDA, IFLA, LIBER, and SPARC Europe with SPARC Europe as co-ordinator. We hope this guidance will prove useful to academic libraries in Europe when implementing this into national legislation.

1. In a nutshell: new rules for the sharing of public and publicly funded data

On 16 July 2019, the Directive on Open Data and Public Sector Information ("the Open Data Directive" or "the Directive" – Directive (EU) 2019/1024) entered into force.[1] The harmonisation of national rules and practices on the re-use of public information are to contribute to the smooth functioning of the internal market and the proper development of the information society in the European Union (EU).

Member States have to implement the Directive into their national laws by 16 July 2021.[2] Further opportunities for research institutions and libraries are highlighted in boxes in this document to indicate where Member States have some leeway when implementing the Directive into their own national legislation to increase the benefits of openness for the research community.

The Open Data Directive provides a common legal framework for public sector information ("PSI"—not to be confused with publicly accessible data[3]) in the EU, founded on the principles of transparency and fair competition, and brings that framework up to speed with the evolution of digital technologies.

The Directive updates the existing rules governing the re-use of **PSI held by Member States' public sector bodies** at national, regional and local levels, such as ministries, state agencies as well as publicly funded libraries, research organisations and other cultural heritage bodies. In addition, it also determines the re-use of documents held by **public undertakings**, i.e. organisations funded mostly by or under the control of public authorities such as meteorological, water, energy, transport, and postal services.[4]

---

[1] The Open Data Directive supplants the rules introduced by the 2003 Public Sector Information Directive ("the PSI Directive" (Directive 2003/98/EC)) which had been amended by Directive 2013/37/EU to stay abreast of technological developments and to add the content held by museums, libraries (including university libraries) and archives to the PSI Directive's scope of application. A proposal for a revision of the Directive was adopted by the European Commission on 25 April 2018. On 22 January 2019, negotiators from the European Parliament, the Council of the EU and the Commission reached an agreement on the revision proposed by the Commission. When adopted in June 2019, the Directive was renamed as the "Open Data and Public Sector Information Directive." See also European Commission, "From the Public Sector Information (PSI) Directive to the open data Directive".

[2] For reference, see the detailed overview of legislation implementing the former PSI Directive in each EU Member State and the countries of the European Economic Area.

[3] Publicly available data is data that is available to the public whereas public sector information is information created or held by a public body.

[4] The Directive clarifies that an undertaking is considered "public" if public sector bodies may exercise "*a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it*", regardless of whether that is a direct or an indirect influence. The only relevant criterion is therefore

The Directive focuses on the economic potential of the re-use of information and encourages Member States to **make as much data available for re-use as possible**. Simply put, **all public sector content** (with some caveats) that can be accessed under national access to information rules is in principle **freely available for re-use**. Public sector bodies are not allowed to charge more than the marginal cost for the re-use of their data, except in very limited cases.

Moreover, the Directive takes positive steps to enhance the way that **publicly-funded research data** is made available, accessed, shared and re-used. Member States are required to **develop national policies** for open access to research data resulting from public funding, following the principle of **'open by default',** while new harmonised rules on re-usability are to be applied to all publicly-funded research data which is already made accessible via open **repositories**.

As such, the Open Data Directive is a **true testament to the EU embracing Open Access** and is a clear commitment to innovation and transparency. The Directive shows the EU's will to lift the barriers to the wide re-use of PSI and publicly funded research data. This is an enormous improvement over the previous situation where PSI was not always made openly available, with negative consequences on Europe's innovation sector and to the detriment of the fundamental principle of freedom of access to information.

2. Background: access to public data to enhance research and knowledge development and to boost innovation in the digital economy

As technology advances, our understanding of the power of data and information not just to support transparency in government, but also to drive innovation, create jobs and generate other public goods grows. The public sector - through the ministries, institutions and agencies that form it, through other organisations performing public tasks, and through funding research - is a major creator and holder of such information.

This can cover all sorts of subjects including markets, transport, environment, pollution, households, etc. Traditionally, data gathered, created and held by public bodies or undertakings is referred to as "public sector information" or PSI.

For academic **researchers**, PSI and publicly funded research data are a treasure trove of information that they can use to gain new insights into a plethora of issues and in turn generate new knowledge based on solid underpinnings. For businesses, and especially **start-ups and SMEs**, for example app builders, PSI is a resource with the great commercial potential to enable them to enter new, highly competitive markets and to explore emerging business opportunities for data-based products and services. As this data is opened up permanently to the public in an **open, standardised, machine-readable format**, researchers and businesses, along with any other member of the public and the public sector itself, can tap into these resources for great benefit.

The key innovation of the Directive is that it extends the logic of previous legislation on PSI to **publicly-funded research data**, which shares the "public" characteristic with PSI. In light of the above, the Directive aims to enhance the way that **PSI** and **publicly-funded research**

---

whether public sector bodies are able to exercise control over an undertaking. Once public undertakings make their data available, they have to comply with the principles of transparency, non-discrimination and non-exclusivity set out in the Directive and ensure the use of appropriate data formats and dissemination methods. They will still be able to set reasonable charges to recover the costs of producing the data and of making it available for re-use. See Article 2(3).

data are made available, **accessed and shared**, with a view to stimulating innovation in products and services in the Digital Single Market (DSM), to encourage Europe's digital economy and to foster academic research and the sharing of knowledge among researchers. Together, therefore, PSI and publicly-funded research data, insofar as they are made freely available to the public, are described as 'open data'. As such, the re-use of PSI and research data is approached according to the principle of **"open by design and by default."**

3. Provisions relevant to open access to research data

Several provisions are relevant to open access to research data; the most important one for research performing organisations, libraries and research funding organisations is Article 10. This section hence starts with an explainer on Article 10, then moves to other relevant articles, in numerical order.

Article 10: Research data

The main rules around **access and re-use of publicly-funded research data**[5] are found in Article 10.[6] Note that scientific publications[7] (i.e. books, journals, etc.) are excluded from the definition of research data.[8]

For the first time, the Open Data Directive makes accessing, sharing and re-using publicly funded research data a priority in EU law. Article 10(1) reads, in part: "*Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies') following the principle of 'open by default' and compatible with FAIR principles*." This is also in line with Horizon Europe.

Put differently, Member States need to support the availability of research data by adopting national **open access policies**. The Directive makes clear that these relate primarily to research data, rather than publications, although there is nothing that would exclude the broader approach of it applying to journals etc., as foreseen in the Competitiveness Council Conclusions of May 2016[9]. Such open access policies already exist in some countries such as the Netherlands, Slovenia, Spain, Lithuania or Norway, with others under development.[10] Such policies (and other relevant actions) would be addressed to research organisations and funders, i.e. those taking part in publicly funded research, either on the giving or receiving end.[11] Access to research data follows the principle of "**as open as possible, as closed as necessary**", according to the **FAIR principles** (findable, accessible, interoperable, reusable).

Open data is defined as data in an open format that can be freely used, re-used and shared by anyone for any purpose and as set out above, it applies both to research data as well as

[5] Research data is defined in Article 2(9) as "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results."

[6] In its article on Subject matter and scope (Article 1), the Directive refers to Article 10 and states that "in order to promote the use of open data and stimulate innovation in products and services, this Directive establishes a set of minimum rules governing the re-use and the practical arrangements for facilitating re-use of: … (c) research data pursuant to the conditions set out in Article 10."

[7] Article 2(9).

[8] For a complete list of excluded areas, see Article 1(2).

[9] http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf

[10] An *Analysis of Open Science Policies in Europe v4,* SPARC Europe Digital Curation Centre Report, 2019, https://zenodo.org/record/3379705#.XbKooJMzYW8

[11] Recital 28 states that: "…it is appropriate to set an obligation on Member States to adopt open access policies with respect to publicly funded research data and ensure that such policies are implemented by all research performing organisations and research funding organisations."

to information produced by public bodies as part of their public task.[12] **Open access** is explained as *"the practice of providing online access to research outputs free of charge for the end user and without restrictions on use and re-use beyond the possibility to require acknowledgement of authorship. Open access policies aim in particular to provide researchers and the public at large with access to research data as early as possible in the dissemination process and to facilitate its use and re-use."*[13]

The Open Data Directive does not apply to documents held by all types of cultural institutions. Indeed, some **cultural establishments** are **excluded** from the scope of the Directive, i.e. the rules of the Directive do not apply to them.[14] These are, for example, orchestras, operas, ballets and theatres, including the archives that are part of those establishments. These cultural establishments are mostly dealing with performing arts and almost all of their material is subject to third-party intellectual property rights.[15] However, documents held by libraries (including university libraries), museums and archives, are covered by the Open Data Directive where the holdings are in the public domain, or where the copyright and/or *sui generis* database rights belong to the institution.

Member States have the opportunity to promote in their national legislation the consistent use of data management plans (DMPs) to ensure longer term access to that data (this is recognized under Recital 23). Furthermore, Member States have the opportunity to rule that Application Programming Interfaces (APIs) also need to abide by the FAIR principles, using open protocols.

Nevertheless, account must be taken of "*privacy, protection of personal data, confidentiality, national security, legitimate commercial interests, such as trade secrets, and intellectual property."*[16] This opens a path for **exceptions** to the obligation of making data openly accessible. In other words, although the text of Article 10 stipulates that research data should be made "*as open as possible, as closed as necessary*" a clear delimitation of the situations in which data could not be made available is also provided for.

Indeed, it is not uncommon for government organisations not to disclose data about persons and enterprises in order to protect privacy or trade secrets. In that vein, Recital 28 states that *"research data which are excluded from access on grounds of national security, defence or public security should not be covered by this Directive."*

While it is important to include **exemptions** for personal data, third party intellectual property rights protection and security, other considerations such as knowledge transfer activities and commercial interests risk overly precluding the re-use of some public-sector information, making it necessary to recall that the **default is open**, free and meaningful access to publicly funded research data.[17] In other words, it is necessary to keep to a minimum, situations where beneficiaries of public funding for research are able to withhold this data from the public on the grounds of confidentiality or legitimate commercial interests.

This is an area that offers some flexibility for Member States at the national level to scope out the extent of such possible exemptions. In the interest of open access to research data, those exemptions should be as limited as possible. For example, a Member State could **develop a national rule** that when data is restricted, this must be clearly justified, recorded

[12] Recital 16.
[13] Recital 27.
[14] Article 1(2)(j).
[15] Recital 65.
[16] Article 10(1).
[17] Recital 28 also states that: "…under the national open access policies, publicly-funded research data should be made open as the default option."

and made public, ideally through public metadata in machine-readable form. A number of national open science (OS) policies and research funders already specify this in various ways as identified in an analysis of funder and national OS policies by SPARC Europe.[18]

While the scope of Article 10 is publicly-funded research data, *"certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidised by public funding or **co-funded by public and private-sector entities.**"*[19] The Directive is silent, however, as to exactly *which* obligations would extend to such mixed-funded research data.

There is thus an opportunity here for Member States to address this at the national level and **make the obligations incumbent on publicly-funded research data also mandatory for *partly* publicly funded research data.** However it must be borne in mind that applying the same obligations to partly private or majority private funded research is likely in many instances to disincentivise companies to invest in public-private partnerships. Member States thus need to find a balance between the need for data to be open and the economic realities.

Article 10(2) stipulates that **re-use** includes use for either **commercial or non-commercial purposes**, insofar as research data is publicly funded and have **already** been made publicly **available** through an institutional or subject-based **repository**.[20] Recital 28 similarly states that the Directive *"should apply only to such research data that have **already been made publicly available** by researchers, research performing organisations or research funding organisations through an institutional or subject-based **repository** and should not impose extra costs for the retrieval of the datasets or require additional curation of data"*. In other words, data needs to be deposited in a publicly accessible repository to fall under the obligations put in place by the Directive.

Nonetheless, Recital 28 states that *"Member States may extend the application of this Directive to research data made publicly available through **other data infrastructures than repositories**."*

In order to extend the scope of this rule and to ensure that as much research data is made openly accessible, Member States could **consider adding, after "repository": "or other local, national or international data infrastructures"** when developing their own national legislation.

Article 10(2) also provides that legitimate commercial interests, **knowledge transfer activities** and pre-existing intellectual property rights must be taken into account when making research re-usable.

Member States could more clearly **define what this implies in view of promoting open access and its positive impact on research, industry and society** while at the same time taking knowledge transfer activities into consideration.

Article 1: Database rights

Where databases fall under the scope of the Directive, the public sector body responsible for the database may not use the right to restrict the extraction of data under the Database

[18] SPARC Europe Analysis of funder and national OS policies on Licensing and Exceptions, August 2019: https://docs.google.com/document/d/1SWrRgN3ieqgEu0WNDgDkw5C9gNANs_rbWpd5mCyH_z4/edit?pli=1
[19] Recital 28.
[20] See also Recital 11.

Directive to prevent or restrict the re-use of documents.[21] This is a much welcome revision as the Database Directive provides exclusive rights to the owner of databases, which risk running counter to the principle of data openness.

Article 6: Charging of costs for cultural institutions

While the general rule is for public sector information to be available **free of charge**, cultural institutions, including libraries, archives and museums, are still allowed to *"charge above marginal costs in order not to hinder their normal running"*.[22] **Cultural institutions can charge a fee** as needed to recover the costs to make reproduction, provision and dissemination of documents as well as for anonymisation of personal data and measures taken to protect commercially confidential information, together with a reasonable return on investment.[23] These provisions existed in the previous version of the Directive.

Article 8: Standard licences

The Open Data Directive promotes the use of **standard licences** for the re-use of public sector information.[24] The Directive indicates that *"any licences for the re-use of public sector information should in any event place **as few restrictions on re-use as possible**, for example limiting them to an indication of source. Open licences in the form of standardised public licences available online which allow data and content to be freely accessed, used, modified and shared by anyone for any purpose, and which rely on open data formats, should play an important role in this respect."*[25] Standard licenses are defined in Article 2(5) as *"a set of predefined re-use conditions in a digital format, preferably compatible with standardised public licences available online."*

Standard open licensing is a positive step forward, however, there is a risk that if Member States remain unclear about which licences are acceptable for application to PSI law, it could create confusion and interoperability problems.

Member States are thus encouraged to promote the utilisation of **new or existing, interoperable standard licences or rights waivers for the sharing of research data subject to PSI law**.[26] Importantly, even though Member States are required to have in place standard licences for the use of PSI, public sector bodies are merely *encouraged* and not obliged to use them. Member States now have the opportunity if they choose to **make the use of standard licences an *obligation*** at the national level and to determine the characteristics of such licences.

See here for examples on open licensing in European national and international policy.

[21] Article 1(6) provides that "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 re-use of documents or to restrict re-use beyond the limits set by this Directive."
[22] Article 6(2)(b). Article 6(5) and Recital 38 are to the effect that "libraries… should be able to charge above marginal costs in order not to hinder their normal running." Income should not exceed costs and a reasonable return on investment.
[23] Article 6(5).
[24] Article 8(2) states that "in Member States where licences are used, Member States shall ensure that standard licences for the re-use of public sector documents, which can be adapted to meet particular licence applications, are available in digital format and can be processed electronically. Member States shall encourage the use of such standard licences."
[25] Recital 44.
[26] See for example the European Commission's "Guidelines on recommended standard licences, datasets and charging for the re-use of documents", online https://ec.europa.eu/digital-single-market/en/news/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use.

Article 9: Practical arrangements and preservation of documents

Article 9 lays out practical arrangements to **facilitate finding data**. Examples include the development of **tools and online portals** that make it easier for users to find and re-use data. Furthermore, for longer term access to data, in view of the threats posed by technological obsolescence and loss of data, Article 9(1) explicitly calls on Member States to facilitate the **effective preservation of documents** and digital information.[27] In relation to open licensing, it is important that search tools and repositories properly mark datasets and other documents with the **appropriately licensed metadata**, otherwise, users will not be able to find and know how they can re-use a particular resource.

As part of the practical arrangements for the preservation of documents, it is important for Member States to **ensure the correct implementation of standard open licences, implement meaningful preservation policies for public sector information and data, and provide instructions to bodies subject to PSI law and re-users alike on how to apply those licences**.

Article 12: General prohibition of exclusive arrangements and cultural heritage digitisation

The Open Data Directive generally aims to **prevent** the conclusion of agreements that could lead to **exclusive re-use** of public sector data by private partners, with some exceptions. Indeed, some public bodies strike complex deals with private companies, which can potentially lead to information being "locked in". Safeguards are put in place to reinforce transparency and to limit the conclusion of agreements which could lead to exclusive re-use of public sector data by private partners. In Article 12, the Directive strengthens the transparency requirements for public–private agreements involving PSI, limiting the possibilities to enter into exclusive arrangements.

As far as **cultural heritage institutions** are concerned and especially regarding **digitisation outsourced to private third parties**, non-standard rules on access are still applied as it is recognised that private players invest large amounts of money in the digitisation of public domain collections held by libraries and archives. Article 12(3) states that "*where an exclusive right relates to the digitisation of cultural resources, the period of exclusivity shall in general not exceed ten years."* This means that private companies engaged in digitisation are permitted to exclusively control access to re-use of these works for 10 years - and maybe longer - possibly keeping digitised works in the public domain under private exclusive control. This opportunity to restrict re-use is justified in the text in light of the necessity to give the private partner the possibility to recoup its investment, and so permit the digitisation in the first place.[28] The Directive provides that this window of exclusivity should be *"as short as possible, in order to respect the principle that public domain material should stay in the public domain once it is digitised."*[29]

Member States could thus provide in national legislation rules that limit the **duration of exclusive rights** to periods of a determined length, *as short as possible or a maximum of 10 years*, to enable wide and unrestricted use of cultural material in the public domain. This of course is subject to the cultural heritage institution having the necessary IT infrastructure in place to make the materials available online at the end of the period of preferential treatment. This is often not the case, and sometimes why such digitisation projects are undertaken in the first place because the institution does not have the required digital capabilities.

[27] Article 9(1) reads: "Member States shall also encourage public sector bodies to make practical arrangements facilitating the preservation of documents available for re-use."
[28] Recital 49.
[29] Recital 49.

Article 13: High-value datasets

In the Open Data Directive, a particular focus is placed on **high-value datasets**, defined as documents whose re-use is associated with important benefits for society and the economy, such as statistics or geospatial data. These datasets are of great interest to the research community and have a high commercial potential and can speed up the emergence of a wide variety of value-added EU-wide information products and services. They also serve as key data sources for the development of Artificial Intelligence (AI).

High-value datasets are subject to a separate set of rules ensuring their **availability across the EU free of charge, in machine readable formats, provided via APIs and, where relevant, as bulk download**.[30] The datasets themselves are not defined in the Directive, but their thematic scope is provided in an Annex.[31] The holdings of cultural heritage bodies are expressly excluded. Indeed, Article 14(4) states that: "*The requirement to make high-value datasets available free of charge... shall not apply to libraries, including university libraries, museums and archives."*

The Directive requires the adoption by the European Commission (via implementing acts following an impact assessment) of a list of high-value datasets to be provided free of charge and easily re-usable across the EU. The Commission is working together with Member States to define the list of specific high-value datasets.

Recitals 32 and 59: Cross-border uses

These recitals promote **access to data across borders** by stating that Member States are to make practical arrangements[32] to help find data and should encourage the preservation of documents for re-use; open APIs[33] should be used where possible. According to Recital 59, examples of such practical arrangements are "*assets lists, which should preferably be accessible online, of main documents… and portal sites that are linked to decentralised assets lists*."

At the national level, such practical arrangements should be more clearly defined to **ensure broad access to data across borders**. This can be better guaranteed by making research data open and by using common metadata standards for example**.**

[30] However, according to Article 14(3), these datasets cannot be made available free of charge if this would result in a distortion of the market.
[31] The thematic categories of high-value datasets, as referred to in Article 13(1), are: Geospatial; Earth observation and environment; Meteorological; Statistics; Companies and company ownership; and Mobility.
[32] Recital 59 provides examples of such practical arrangements: "assets lists, which should preferably be accessible online, of main documents (documents that are extensively re-used or that have the potential to be extensively re-used), and portal sites that are linked to decentralised assets lists."
[33] According to Recital 32, "An API is a set of functions, procedures, definitions and protocols for machine-to-machine communication and the seamless exchange of data."