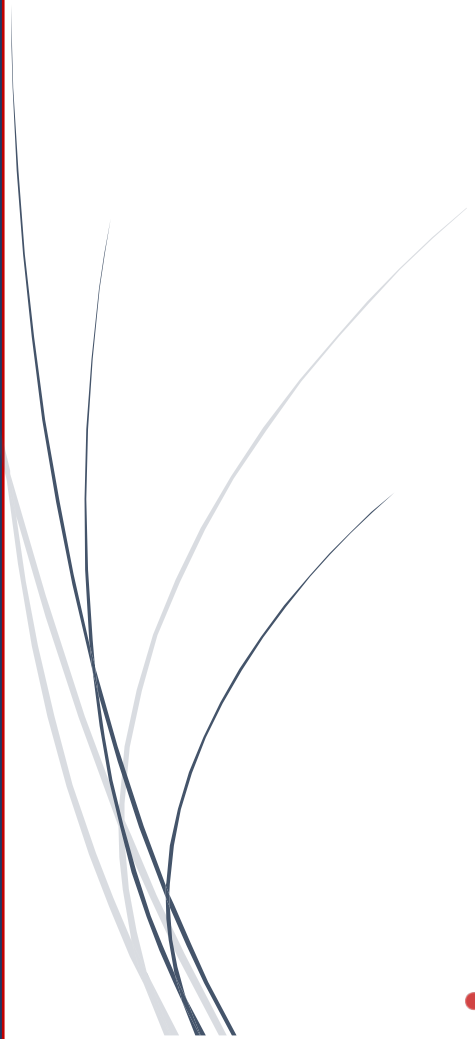
EOSC STUDY

LEGAL INTEROPERABILITY AND THE FAIR DATA PRINCIPLES

December 2020



X-officio

LEGAL | GOVERNANCE | PROCUREMENT

This study on legal and regulatory issues related to the application of the FAIR Principles was commissioned by the FAIR Working Group at the European Open Science Cloud (EOSC) Secretariat and delivered by:

Completion date	December 2020
Lead author	**Dr. Ohad Graber-Soudry**
Contributing authors	Prof. Dr. Timo Minssen Daniel Nilsson Dr. Marcelo Corrales Dr. Jakob Wested Bénédicte Illien
Contact	www.xofficio.eu

Disclaimer: The information and views set out in this study do not necessarily reflect the official position of the EOSC Secretariat. Neither X-officio nor any of the authors of this study or parts of it are responsible for the use made of the study.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

TABLE OF CONTENTS

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

EXECUTIVE SUMMARY

This study on legal interoperability has been commissioned by the EOSC FAIR Working Group. It aims to provide a systematic overview and analysis of the key issues in legal interoperability in connection with the implementation of the FAIR Principles within the context of the EOSC.

Data accessible through the EOSC will be governed by the FAIR Principles, embracing Open Science practices. Legal constraints such as ensuring a secure environment where privacy and personal data are protected and where users of the EOSC can be reassured about issues concerning data security, data sovereignty, intellectual property rights, liability risks and the like, will need to be addressed.

The recommendations provided in this study should contribute to the EOSC Interoperability Framework jointly developed by the EOSC FAIR and Architecture Working Groups, which addresses four interoperability layers (technical interoperability, semantic interoperability, organisational interoperability and legal interoperability). It is also hoped that this study will contribute to the broader discussion on legal interoperability, the FAIR Principles and the development of the EOSC.

The study is structured as follows:

Chapter one – provides an introduction and background to the study, including scope, limitations and a number of definitions intended to facilitate the discussion that follows.

Chapter two - addresses key thematic legal issues, including:

- Copyright (including database rights) and licences;
- Other intellectual property rights (e.g. patents, trade secrets, neighbouring rights);
- Privacy and data protection (GDPR);
- Other restrictions and legitimate reasons (e.g., protection of sovereign genetic resources and traditional knowledge); and
- Private law considerations.

Chapter three – adds to the analysis through a consideration of a number of enabling legal instruments.

Chapter four – concludes with a summary of the 31 recommendations which have been highlighted throughout the study.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

CHAPTER I: INTRODUCTION

1. Background

1.1. The European Open Science Cloud (EOSC)

With the exponential growth in the quantity and volume of data produced in the research lifecycle, new capacities to discover, access, and process data and information of different types, volumes and origins, are needed. Currently, too many datasets produced from public funds remain beyond the reach of many scientists, even scientists within the same discipline.

Against this background the EOSC – a European Commission initiative – is being developed as a globally accessible, multidisciplinary data infrastructure. The EOSC will federate the existing scientific data and digital infrastructures for data exploitation that are now spread across disciplines and European Union (EU) Member States. It aims at giving the EU a global lead in research data management by making science more efficient and productive.[1]

The EOSC will bring together institutional, national and European initiatives, data and service providers, research infrastructures and all relevant stakeholders to co-design and deploy a European Research Data Commons. This will be achieved as European researchers and professionals in science and technology will be offered a virtual and trusted federated environment where they will be able to deposit, find, access and reuse European scientific data through the EOSC and for the EU to make the transit from fragmented datasets to an integrated EOSC.[2] It is envisaged that the EOSC will stimulate and enable researchers to work collaboratively and practise Open Science and help solve the scientific and societal challenges of our time.[3] It is important to understand however that the EOSC is a federation, a 'catalogue of catalogues' and is not aiming, at least not at this stage, to become a one and only central point of access for researchers, even though it may have its own portal. In practice, it is more likely than not that researchers will continue to use the current services and repositories that are being used in their research community, and the EOSC will serve as a supporting environment for Open Science and not as an 'Open Cloud' for science.[4] This position may (or may not) change in the future, in particular once a common vision has matured and been agreed upon by all stakeholders.

For the purposes of this study, it is important to note that data accessible through the EOSC will be governed by the FAIR Principles, as will be explained further below, while legal constraints such as ensuring a secure environment where privacy and personal data are

[1] European Commission, *Commission Staff Working Document – Implementation Roadmap for the European Open Science Cloud*, SWD (2018) 83 final.
[2] Speech by former Commissioner Carlos Moedas in Amsterdam on "Open science: share and succeed", 4 April 2016.
[3] See Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), Version 0.818, October 2020, https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v08.pdf, accessed 17 November 2020.
[4] See European Commission, *Realising the European Open Science Cloud, First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud* (European Union 2016), https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf, accessed 17 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

protected and where users of the EOSC can be reassured about issues concerning compliance, data security, intellectual property rights, liability risks and the like, will have to be addressed.

1.2. The EOSC Secretariat

The EOSC Secretariat[5] is a consortium of European organisations from various scientific domains which was established to deliver support to the EOSC overall governance, while working openly and inclusively with communities to co-create an all-encompassing EOSC.

While remaining neutral towards the community it is serving, the EOSC Secretariat follows a pragmatic approach that is fully dedicated to realising the outcomes of the EOSC and addressing all the specific needs of the coordination structure expected by the EOSC.

The EOSC Secretariat retains a high degree of flexibility in its roll-out plan by adopting a co-creation approach, founded on a substantial budget left available for all upcoming, foreseen and unforeseen, challenges of introducing a truly operational EOSC serving all European stakeholders. This study is funded from the budget of the EOSC Secretariat.

1.3. The FAIR Working Group

The EOSC FAIR Working Group is one of six working groups[6] established in 2019 by the EOSC governance structure. Its role is to advise and provide recommendations to the Executive Board on the implementation of Open and FAIR practices within the EOSC.

The FAIR Working Group addresses cross-disciplinary interoperability, gathers requirements relevant to the EOSC services, and advises the EOSC governance bodies on FAIR-related matters.[7] It does so in close collaboration with the other five working groups. In particular, it coordinates with the Architecture Working Group to ensure that the FAIR Principles are supported by the technical arrangements while developing the interoperability framework, including legal interoperability – which is the subject matter of this study.

The FAIR Working Group will propose measures for increasing FAIR maturity to maximise sharing and re-use of data. Within the EOSC Working Group, a Task Force is currently working on identifying a set of common rules and general principles that should ensure data and services within EOSC support interoperability and drive the overall creation of the EOSC Interoperability Framework. The Task Force has organised these interoperability principles into four different layers: technical, semantic, organisational and legal interoperability. The conclusions of this study on legal interoperability will feed into the Interoperability Framework Report.

[5] EOSCsecretariat.eu, https://www.eoscsecretariat.eu/, accessed 17 November 2020.
[6] The other working groups are: Landscape, Architecture, Rules of Participation, Skills and Training, and Sustainability.
[7] See EOSCsecretariat.eu, FAIR Working Group, https://www.eoscsecretariat.eu/working-groups/fair-working-group, accessed 17 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

1.4. Scope and purpose of this study

Data-intensive science and innovation and the exponential growth in the quantity of research data requires that humans and machines can make better use of knowledge discovery, and access to integration and analysis of research data (including publications, digital objects, metadata and software). Most recently, the raging COVID-19 pandemic has dramatically shown the importance of this, as well as the ethical and legal dimensions of both Open and FAIR data sharing – i.e. making data Findable, Accessible, Interoperable, and Reusable.[8]

The Declaration on the European Open Science Cloud of 2017[9] emphasised, among other things, that research data should be open by default while taking into account the need to balance openness and the protection of scientific information, commercialisation and intellectual property rights, privacy concerns and security, following the principle "as open as possible, as closed as necessary".

The success of implementing the FAIR Principles, i.e. the ability of the research community to share, access, and reuse data, as well as to integrate data from diverse sources, requires, among other aspects, the free flow of research data without unnecessary constraints, allowing for discovery and interoperability within the EOSC.

Against this background, the FAIR Working Group has commissioned this study on legal interoperability on how, in the context of the EOSC, the application of FAIR Principles is enabled or blocked by legal constraints and impediments across the EU. The results of this study should be considered within the broader EOSC Interoperability Framework.[10]

The body of laws at EU level which directly affects legal interoperability and the implementation of the FAIR Principles within the context of the EOSC is fairly limited and includes intellectual property (in particular copyright, database rights), data protection and laws aimed at protecting sensitive or confidential data. There are however additional legal instruments and strategies that may also affect, directly or indirectly, legal interoperability. In parallel, there are certain 'enabling' legal instruments, soft law and policies that support and promote the application of either Open Data or FAIR Data Principles, or both, at least to the extent that the data is produced or funded by the public sector.

This study digs deeper into the scope and nature of such barriers to legal interoperability and the implementation of FAIR Principles. It aims at providing a systematic overview and analysis of the key issues in legal interoperability in connection with the implementation of FAIR Principles within the context of EOSC. By doing so, it is hoped that the study will contribute to the broader discussion on interoperability, the FAIR Principles and the development of EOSC.

[8] GO FAIR, Committee on Data (CODATA), Research Data Alliance (RDA), and World Data Systems (WDS), *Data Together COVID-19 Appeal and Actions* (2020) https://www.go-fair.org/wp-content/uploads/2020/03/Data-Together-COVID-19-Statement-FINAL.pdf, accessed 17 November 2020.
[9] European Commission, *EOSC Declaration* (2017) https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf, accessed 17 November 2020.
[10] Oscar Corcho et al., *EOSC Interoperability Framework (v1.0) – Draft for community consultation* (2020) https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf, accessed 17 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

1.5. Methodology

For the purpose of carrying out this study, a number of sources and methods have been used. The sources consulted and methodology carried out to achieve the objectives of this study are set out below.

1.5.1. Sources

The following sources have been used for carrying out this study:

- Desk research and analysis of legal norms, related literature, guidelines and web-based publications concerning FAIR Principles, open data and legal interoperability. A number of reports and academic texts on legal interoperability and research data have been produced over the last few years, focusing primarily on copyright and licensing issues in connection with the open sharing, access and reuse of publicly funded research data. This study builds upon such documents and expands them wherever relevant. However, the body of knowledge that has been reviewed is not intended and should not be perceived to be exhaustive of all knowledge of the FAIR Principles.

- Interviews conducted with data officers and experts from a selected number of research infrastructures and EOSC stakeholders. Individual interviews with data experts were instrumental in identifying both common and unique problems (depending on the discipline of the research infrastructure) faced by key stakeholders and to develop requirements in each aspect of legal interoperability. We are grateful to the following individuals who agreed to share their knowledge and thoughts with us in a series of interviews:

  - Carsten Thiel, CESSDA ERIC
  - Arnaud Gingold, OPERAS
  - Andreas Witt and Paweł Kamocki, CLARIN ERIC
  - Jonathan Taylor, European Spallation Source ERIC
  - Juan Bicarregui, UKRI
  - Sharif Islam, DiSSCo
  - Javier Quinteros, GEOFON

- Review of the results of the interviews conducted as part of a broader survey delivered by the Interoperability Task Force with key stakeholders from European Research Infrastructure Consortia (ERICs), ESFRI projects, service providers and research communities.

- Participation in the meetings of the Interoperability Task Force in the period from June to September 2020, in which relevant issues have been raised and discussed, as well as relevant symposia during the period of the study.[11] The authors would like to thank

[11] Most notably, the second ESFRI RIs-EOSC Workshop *Research Infrastructures shaping EOSC*, 6-7 October 2020, programme available at: https://www.esfri.eu/esfri-events/2nd-esfri-ris-eosc-workshop-research-infrastructures-shaping-eosc, accessed 17 November 2020; and the EOSC Governance Symposium 2020, 19-22 October 2020, programme available at: https://www.eoscsecretariat.eu/events/eosc-governance-symposium-2020, accessed 17 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Sarah Jones and Oscar Corcho for enabling our participation in the meetings of the Task Force on Interoperability and for their support.

1.5.2. Methodology

The study combined a number of methodological approaches to carry out the analysis:

- Legal Dogmatics Analysis: providing in-depth knowledge and an overview of current legal norms relevant to legal interoperability and clarifying the legal framework that is relevant and may be applicable to the implementation of FAIR Principles. The purpose of the legal dogmatic analysis is to describe, analyse, clarify, interpret, evaluate and systemise the content of valid current norms. The legal dogmatic method combines a literature review with the mapping of relevant legal sources (regulation, directives, national laws, case law, soft law, etc.).

- Pragmatic Analysis: adding a consideration of formal and practical implementation as well as broader ethical and social aspects to the legal findings, where relevant. The pragmatic analysis in the study seeks to identify the best criteria for the most efficient and realistic norm (e.g., policies, rules, measures, procedures or models). The study further identifies problems and proposes recommendations, taking into account guidelines, policy initiatives, projects, licensing arrangements and general considerations related to the FAIR Principles. The approach taken in this study capitalises on the diverse and interdisciplinary background of the project team, their international experience, and focuses on specific topics in order to evaluate existing models of governance and law to formulate recommendations.

- Comparative Analysis: wherever relevant, references to selected jurisdictions have been considered in order to illustrate specific issues and tensions between legal regimes (legal interoperability across jurisdictions) and in order to explore the variety of legislative setups for FAIR that have been adopted.

1.5.3. Structure

The main body of the study follows a thematic approach whereby relevant legal subject (themes) applicable to legal interoperability and the implementation of the FAIR Principles are considered and discussed. The key thematic legal issues addressed in the study and organised in separate chapters include:

- Copyright (including database rights) and licences;
- Other intellectual property rights (e.g. patents, trade secrets, neighbouring rights);
- Privacy and data protection (GDPR);
- Other restrictions and legitimate reasons (e.g., protection of sovereign genetic resources and traditional knowledge);
- Private law considerations.

A set of recommendations is provided throughout the study and is summarised at the conclusion of the study.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

1.6. Limitations and constraints

This study was prepared subject to certain conditions and constraints. Most notable is the fact that the EOSC is still in its early stages, only starting to emerge, and many issues remain unresolved as parties continue to discuss the future of the EOSC and the envisaged structure and shape it will form.

The nature of legal analysis, which is the subject matter of this study, requires an application of existing law to a clear set of facts in order to be meaningful. For this purpose, a good starting point for a study of this kind is to gain a thorough and precise understanding of what the EOSC is and how will it operate. However, as metaphorically outlined in the EOSC Strategic Implementation Plan,[12] the current understanding of the EOSC is best described as follows:

"*There is a parable of the blind men and the elephant, which originated in ancient India. It is the story of a group of blind men who have never come across an elephant before and who are to describe the elephant by respectively touching one - only one - different part of the elephant. Each blind man feels a different part of the elephant's body, such as the tail, the trunk, one leg. They describe the elephant based on their different experiences and of course, the descriptions are entirely different from one another.*

If you ask a room full of people what EOSC is, you'll get a room full of different answers. It's such a large-scale initiative and ambitious mission that it's like an elephant – everyone sees a different part and few see the big picture…"

Another constraint, which is not uncommon for studies of this kind, is the relatively strict timeframe within which this study had to be delivered. Unhelpful was also the fact that it was carried out during the COVID-19 pandemic, coinciding with significant lockdown restrictions in many of the EU Member States and beyond, and the consequences the pandemic had on the work environment and availability of some of the team members, as well as stakeholders, particularly those with families and young children.

Nevertheless, this study represents a significant effort to explain the legal issues that are relevant to legal interoperability within the context of the EOSC and to provide a set of recommendations. Since data is, by its nature, discipline-specific, it also means that some of the recommendations may be discipline-specific and, at the same time, some discipline-specific recommendations may be missing. By no means should this study be perceived as an attempt to provide a comprehensive analysis of all aspects of legal interoperability. Instead, it should be viewed as a living document with the potential to be supplemented and refined as new issues may be identified while the EOSC is being developed.

Finally, it is important to highlight that the purpose of this study is to make recommendations in relation to legal interoperability and how, in the context of EOSC, reusability of data is enabled or blocked by legal constraints. This study does not attempt to provide any value-judgement or policy recommendations in relation to the EOSC initiative or to open data and open science principles in general. For example, this study supports the assertion that, for legal interoperability to be maximised, the implementation of the FAIR Principles should go hand-

[12] See Sarah Jones and Jean-François Abramatic, *European Open Science Cloud (EOSC) Strategic Implementation Plan* (European Union 2019) https://op.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1, accessed 17 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

in-hand with efforts to make data open. This recommendation should be read in the context in which it is given and should not be understood to mean that it is desired to make research data available to everyone in general, or whether reciprocity should be a condition for access to non-EU users, or whether other EU competitiveness considerations or adequate IP protection plans should be taken into account as a matter of policy and to what degree.[13] A discussion on such policy recommendations falls outside the scope of this study and is thus not addressed here.

2. Data

A number of key concepts require clarification in order to avoid confusion and to facilitate a more precise understanding of the results of this study. In particular, within the context of the EOSC and the scope of this study on legal interoperability, the terms 'data', 'research data', 'digital objects' and 'research artefacts' are used interchangeably.

Given the variety of terms used in the existing literature and in the different disciplines to describe research data and its various direct and indirect components, in this study it is important to be clear on what is meant when discussing data as well as the scope and the meaning of the related terms. As has been highlighted elsewhere,[14] many research artefacts are discipline-specific, which means that FAIR practices will also be discipline-specific and so will the terms used to describe them. Moreover, taking the humanities as an example, it has been claimed that "the many things that would be seen as data in another discipline are often called something else in the humanities. We resist using the blanket term 'data' for the very good reason that we have more and precise terminology (e.g. primary sources, secondary sources, theoretical documents, bibliographies, critical editions, annotations, notes, etc.) available to us to describe and make transparent our research processes."[15]

It is impossible, within the scope of this study, to provide discipline-specific definitions to research data, and therefore recommendations made throughout this study may be more relevant to some disciplines and less so for others. We use the general term 'data' for data used in research in a broader sense, including anything which is a direct component of the research process, as further illustrated by the terms defined below. We primarily discuss research data that is made machine-actionable, in order to allow computational systems to find, access, interoperate and reuse the data.

2.1. Research data

The term 'research data' can be described as the evidence used to inform or support research conclusions. The UKRI, Engineering and Physical Sciences Research Council (EPSRC),

[13] See, for example, Frans Oort, *How to make open science work* (Science Business, 28 July 2020) https://sciencebusiness.net/viewpoint/how-make-open-science-work, accessed 17 November 2020; and Florin Zubașcu, *Viewpoint: the EU needs better IP protection of publicly-funded science* (Science Business, 3 December 2019), https://www.eoscsecretariat.eu/sites/default/files/viewpoint_the_eu_needs_better_ip_protection_of_publicly-funded_science_science_business.pdf, accessed 17 November 2020.
[14] Chue Hong et al., *Six Recommendations for Implementation of FAIR Practice* (2020) http://doi.org/10.5281/zenodo.4065549, accessed 17 November 2020.
[15] Jennifer Edmond and Erzsébet Tóth-Czifra, *Open Data for Humanists, A Pragmatic Guide* (2018) http://doi.org/10.5281/zenodo.2657248, accessed 4 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

defines research data as "recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created."[16]

The Concordat on Open Research Data[17] defines research data as "the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence ... [T]he primary purpose of research data is to provide the information necessary to support or validate a research project's observations, findings or outputs. Open research data are those research data that can be freely accessed, used, modified, and shared, provided that there is appropriate acknowledgement if required".

Research data may include various forms, such as 'raw data',[18] 'research ready',[19] published output datasets or published catalogue-type presentations of published output datasets[20] and it may also be divided into categories, such as observational, experimental, simulation, divided/complied, and reference/canonical.[21]

Examples of research data include items such as documents, spreadsheets, laboratory notebooks, fieldnotes, diaries, questionnaires, transcripts, codebooks, audiotapes, videotapes, photographs, films, test responses, sensor data, slides, specimens, samples, neuroimages, collections of digital outputs, data files, database contents (video, audio, text, images), models, algorithms, climate models, economic models, 3D models, scripts, gene sequence databanks, chemical structures, spatial data, chromatograms, contents of an application (input, output, logfiles for analysis software, simulation software, schemas), methodologies and workflows, standard operating procedures, protocols, and many more items of this kind.

2.2. Digital objects

The RDA Data Foundation & Terminology (DFT) Core Terms and Model[22] suggests that "a Digital Object is represented by a bitstream, is referenced and identified by a persistent

[16] UKRI Engineering and Physical Sciences Research Council, EPSRC Policy Framework on Research Data – Scope and Benefits, https://epsrc.ukri.org/about/standards/researchdata/scope/, accessed 19 November 2020.
[17] See HEFCE, RCUK, Universities UK and Wellcome Trust, *Concordat on Open Research Data* (2016) https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf, accessed 19 November 2020.
[18] Initially processed data produced at a research infrastructure or research facility such as a neutron or x-ray source for research in physical science, or an observatory (astronomy).
[19] Processed data which has been fully calibrated, combined and cleaned/annotated.
[20] See Andrew Burnham, *Research Data – Definitions* (2012) https://www2.le.ac.uk/services/research-data/old-2019-12-11/documents/UoL_ReserchDataDefinitions_20120904.pdf, accessed 19 November 2020.
[21] Ibid. The University of Southampton further suggests five ways to define research data, see Mark Scott and Simon Cox, *Introducing Research Data* (4th edn, University of Southampton 2016) https://eprints.soton.ac.uk/403440/1/introducing_research_data.pdf, accessed 19 November 2020.
[22] Gary Berg-Cross, Raphael Ritz and Peter Wittenburg, *RDA Data Foundation and terminology – DFT; Results RFC (v1.5)* (2015), https://www.rd-alliance.org/sites/default/files/DFT%20Core%20Terms-and%20model-v1-6.pdf, accessed 19 November 2020.

identifier, and has properties that are described by metadata". This definition was also adopted by the EOSC Interoperability Framework.[23]

Similarly, the Persistent Identifier (PID) Policy for the EOSC defines digital objects as "a bit sequence that can be stored in multiple repositories and is associated with a Persistent Identifier (PID) and metadata".[24]

The FAIR Expert Group[25] considers the concept of "FAIR Digital Objects" to mean "data, code and other research outputs. At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse."

2.3. Research artefacts and research objects

'Artefacts', 'research artefacts' and 'research objects' are terms that are gaining increased popularity in some disciplines. In the humanities, for example, research objects are used to denote supporting material which underpins or otherwise enriches the written output of research and may include numeric, written and audio-visual data, as well as software code, workflows and methodologies, slides, logs, lab books, sketchbooks, notebooks, etc.[26]

The EOSC Interoperability Framework uses the term "research artefacts" to mean software code, scientific workflows, laboratory protocols, open hardware designs, etc., that may be used in the context of research activity.[27]

2.4. Metadata

Metadata is data that contains descriptive, contextual and provenance assertions about the properties of research data or a digital object.[28] It provides systematic descriptions and attributes of data relevant to interpret what the data concerns. More broadly, it refers to all data about data, such as: structure and internal coherence, source references and licences, time-stamped changes to the data, quality, context, methods and techniques used as well as provenance and context relevant to the proper interpretation and reusability of the data.

[23] Supra note 10.
[24] European Commission, *A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC) - Report from the European Open Science Cloud FAIR and Architecture Working Groups* (European Union 2020) https://ec.europa.eu/info/sites/info/files/research_and_innovation/ki0420576enn.pdf, accessed 19 November 2020.
[25] European Commission, *Final Report and Action Plan from the European Commission Expert Group on FAIR Data - Turning FAIR Into Reality* (2018) https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1, accessed 19 November 2020.
[26] See presentation given at the University of Dundee on 28 October 2015, on behalf of the FOSTER project by Martin Donnelli, *Open Access and Open Data: what do I need to know (and do)?* (2015) https://www.slideshare.net/martindonnelly/open-access-and-open-data-what-do-i-need-to-know-and-do, accessed 19 November 2020.
[27] Supra note 10.
[28] Supra note 22.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Metadata is a key component within the FAIR Principles[29] and has a value of its own. Metadata and data are two separate things and should be treated as such. In particular, FAIR requires that metadata standards are articulated, and that metadata is made publicly available to the greatest extent possible, even if the data itself is not fully open or where the data is no longer available or destroyed.

Finally, for metadata to be useful, it must be standardised and both machine- and human readable to enable advanced research techniques. As will be argued further in this study, where access to data is subject to restrictions or certain conditions, this should be stated clearly and consistently in the metadata, to facilitate research, the combination of data and (legal) interoperability.

2.5. Terminology used in this study

To summarise, the vast number of definitions of data and data-related terms suggests that there is no one-size-fits-all definition of data, and it will often be a context-specific concept that cuts across a wide spectrum of disciplines. For the purposes of this study, we use the terms 'data', 'research data' and 'digital objects' interchangeably as a broad and encompassing term for all direct digital output of the research process, whether they are merely data in the traditional sense or research-oriented digital objects such as workflows, datasets, publications, software code or combinations of these. However, indirect components of the research process, such as teaching and training materials, are not included in our definition and will not be addressed in this study.

Due to the legal nature of this study, we may need to deviate from the above encompassing definition when we discuss specific legal instruments (e.g. a law or a regulation) that offer their own and different definition.[30] We will make reference to other definitions throughout the study where applicable.

3. FAIR Principles

The FAIR Data Principles, first coined in 2014[31] and published in 2016,[32] are a set of guiding principles that seek to increase the reusability of data and digital objects (including data-related algorithms, tools, workflows, protocols, services and other kinds of digital and research objects).[33] They put specific emphasis on enhancing the ability of machines to automatically find and use data, in addition to supporting its reuse by individuals. FAIR plays an essential role in promoting Open Science to improve and accelerate scientific research. According the

[29] Along with other components, most notably persistent identifiers.
[30] For example, the Open Data Directive defines 'research data' 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*". This definition is narrower than the definition we offer above.
[31] At a workshop at the Lorentz Center in Leiden, the Netherlands (https://www.lorentzcenter.nl/) by a community of scholars, librarians, archivists, publishers and research funders upon invitation by the Netherlands eScience Centre and the Dutch Techcentre for Life Sciences (DTL).
[32] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.
[33] Supra note 9, p. 1.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Declaration on the European Open Science Cloud of 2017,[34] the FAIR Principles are neither standards nor practices. The disciplinary sectors must develop their specific notions of FAIR data in a coordinated fashion and determine the desired level of FAIRness. FAIR stands for Findable, Accessible, Interoperable, and Reusable:

Data should be **Findable**	F1. (meta)data are assigned a globally unique and persistent identifier (DOI) F2. data are described with rich metadata F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource
Data should be **Accessible**	A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available
Data should be **Interoperable**	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data
Data should be **Reusable**	R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

In layman's terms,[35] the principle of *Findability* requires data to be identified, described and registered or indexed in a clear and unequivocal manner. Data should be assigned a unique and persistent identifier; that the main characteristics of data are systematically specified, ideally using standard formats; and that these are stored or indexed in a public resource such as a data archive or institutional repository.

The principle of *Accessibility* requires that datasets should be accessible through a clearly defined access procedure, ideally by automated means. This entails the establishment of authentication and authorisation procedures for access as well as the implementation of automated data retrieval protocols where appropriate. Metadata should always be accessible even if the underlying data is not or no longer available.

The principle of *Interoperability* requires that data and metadata are conceptualised, expressed and structured using common, published standards. This entails drawing on standard technical and semantic data formats, variables, ontologies and the like. Moreover, such standards should themselves be made FAIR, meaning at the very least that they are published, traceable and accessible.

The principle of *Reusability* lies at the core of the FAIR Principles, in particular within the context of legal interoperability. It requires that characteristics of the data and their provenance should be described in detail according to domain-relevant community standards, with clear

[34] Supra note 9.
[35] The following non-technical explanation of the FAIR Principles is borrowed from Boeckhout, M., Zielhuis, G.A. & Bredenoord, A.L. The FAIR guiding principles for data stewardship: fair enough?. *Eur J Hum Genet* 26, 931–936 (2018). https://doi.org/10.1038/s41431-018-0160-0.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

and accessible conditions for use. This entails providing and publishing accurate and relevant data descriptions, access and usage licences, the community standards which have been employed in the process as well as the associated provenance for each and every dataset.

FAIR is discussed within the context of Open Science – a movement that encourages researchers and the research community to be open, not only with their results, but also as they conduct their research.[36] However, the FAIR Principles apply to data regardless of whether access to the data is publicly available, i.e., whether the data is Open. FAIR data may or may not be (fully or immediately) Open and an access level to data may be set at "FAIR but not Open". The FAIR Principles do not restrict the recognition of legitimate and necessary reasons for shielding data and restricting disclosure. Examples include data that contains personal information under the General Data Protection Regulation (GDPR), sensitive data that requires access agreements (e.g. protection of endangered species, archaeological sites or aspects of national security), commercially confidential information or situations where there are ethical and moral reasons for restricting access and reuse of the data. The FAIR Principles do not address such legal limitations and techniques, nor do they address moral or ethical issues. The FAIR Principles merely require that data should be findable and the conditions of access and reuse are clearly set out, with the availability of contextual and supporting information (metadata), ideally through an automatic authentication and authorisation process. Key to the successful implementation and indeed to the success of the EOSC at large, is the minimum accepted standardisation for metadata, collaboration and data provenance.

In the context of the EOSC and the global drive towards Open Science, it is often argued that efforts should be made to maximise legitimate access and reuse, while ensuring that restrictions are justified and proportionate. The implementation of the FAIR Principles to publicly funded data needs to go hand-in-hand with the principle that data must be "as open as possible and as closed as necessary" and that the greatest benefits come when data is both FAIR and Open, as the lack of restrictions supports the widest possible reuse, and reuse at scale.[37] This study supports this assertion and, as will be demonstrated, legal interoperability will be far more difficult to achieve when data is not open (even if FAIR). This is particularly the case in circumstances where data (or parts of it) is protected by copyright.

Recommendation 1: Open access to research data is an enabler of legal interoperability. The promotion of FAIR Principles should go hand-in-hand with efforts to make data open in accordance with the principle that data must be "as open as possible and as closed as necessary".

Finally, it is also noteworthy in this context that FAIR data can easily be made Open but making Open Data FAIR could be a daunting exercise and therefore any new data produced should be 'FAIR by design'.

Recommendation 2: Regardless of whether the data is Open or not, all new data made available through the EOSC should be FAIR by design.

[36] For a short summary of what is Open Science and a few definitions of it, see supra note 3, Section 3.1.
[37] Supra note 25.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

4. Interoperability

The general term 'interoperability' is not a legal concept but rather a technical one. Furthermore, there is no one acceptable definition of interoperability.

The International Organization for Standardization (ISO) defines interoperability as the "ability of two or more systems or components to exchange information and to use the information that has been exchanged"[38] while the European Interoperability Framework Report[39] defines interoperability as "the ability of organisations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organisations, through the business processes they support, by means of the exchange of data between their ICT systems."

The EU legislator's definition of interoperability is context specific. A definition can be found in Recital 12 of the Computer Programs Directive[40] providing that interoperability is "the ability to exchange information and mutually to use the information which has been exchanged". The INSPIRE Directive[41] provides that interoperability is the "possibility for [spatial] data sets to be combined, and for services to interact, without repetitive manual intervention, in such a way that the result is coherent, and the added value of the data sets and services is enhanced".

As explained in the EOSC Interoperability Framework Report,[42] "achieving interoperability within EOSC is essential in order for the federation of services that will compose EOSC to provide added value for service users, no matter which scientific disciplines they work on" and "in order for the user systems to consume the digital objects provisioned by the EOSC services they must understand how to read and interpret them, what restrictions there are to use the object and what processes are involved in their production and consumption […] independent from the specific scientific discipline where the digital objects were created or are being consumed".

The role of interoperability frameworks is to define community practices for data sharing, data formats, metadata standards, tools and infrastructure, recognising the objectives and cultures of different research communities.[43] As further set out in the Turning Fair into Reality Report,[44] research that crosses international, legal and disciplinary boundaries relies on drawing together data from different domain repositories and therefore it is likely to face particular challenges because of the current lack of interoperability frameworks, which are needed to make use of similar mechanisms across boundaries.

---

[38] International Organization for Standardization, ISO/TS 27790:2009, https://www.iso.org/standard/44316.html, accessed 19 November 2020.
[39] European Commission, *New European interoperability framework - Promoting seamless services and data flows for European public administrations* (European Union 2017) DOI: 10.2799/78681.
[40] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs [2009] OJ L 111/16 (originally Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs [1991] OJ L 122/42) ("Computer Programs Directive");
[41] 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 [2007] OJ L 108 ("INSPIRE Directive").
[42] Supra note 10.
[43] Supra note 25.
[44] Supra note 25, p. 37.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

The *European* Interoperability Framework identifies four layers of interoperability: legal, organisational, semantic and technical. These have also been considered in the development of the *EOSC* Interoperability Framework.[45] This study focuses on legal interoperability in the implementation of the FAIR Principles within the context of the EOSC.

5. Legal interoperability

Within the context of the EOSC and the FAIR Principles, legal interoperability requires, in particular, that data should be reusable. It concerns the ability to combine datasets from multiple sources without conflicts among the restrictions that each dataset carries (i.e., support of one restriction inherently negates support of another).[46] For example, assume that Anna wishes to combine two resources X and Y in order to create a new derivative work.[47] Both X and Y carry a Creative Commons (CC) open licence but resource X carries an Attribution-Non-Commercial licence (CC BY-NC) while resource Y carries a Creative Commons Attribution-ShareAlike licence (CC BY-SA). If Anna assigns a non-commercial open licence to Z (for example the CC BY-SA-NC), she will breach the terms of the licence carried by Y. If Anna chooses a licence that allows the commercial use of Z (for example by using the CC BY-SA licence) she will breach the terms of the licence carried by X. In other words, the licences carried by X and Y separately cannot be combined and carried forward – in this example, they cannot be reused in Anna's derivative work Z.[48]

It also follows from the example above that the fewest restrictions contained in the source datasets will result in the fewest restrictions contained in the combined or derivative datasets. This implies that only where the data is free from any restrictions and is in the public domain (for example by assigning it a Creative Commons No Rights Reserved CC0 or PDDL dedication), will legal interoperability be maximised.

Legal interoperability also concerns situations where regulatory or policy measures restrict the disclosure of data, or that datasets may be made available only in certain jurisdictions or under certain conditions. Examples include legal restrictions based on intellectual property law, national security, the protection of endangered species or privacy regulations, such as the GDPR. A number of mechanisms are used in practice to restrict access to data where such regulatory or policy measures exist, e.g., embargo, data redaction[49], data generalisation, data anonymisation, or simply restricting any access to the data.

It was already mentioned that the FAIR Principles do not necessarily mean that data should be open and they do not restrict the recognition of legitimate reasons for shielding data. In such

[45] Note there are other proposed layers of interoperability, see for example, Elena Goldstein, Urs Gasser and Ryan Budish, Data Commons Version 1.0: A Framework to Build Toward AI for Good (*Berkman Klein Center*, 2018), https://medium.com/berkman-klein-center/data-commons-version-1-0-a-framework-to-build-toward-ai-for-good-73414d7e72be, accessed 19 November 2020.
[46] See Catherine Doldirina et al., Legal Approaches for Open Access to Research Data (2018) doi:10.31228/osf.io/n7gfa, p. 8.
[47] "Derivative work" is a creation or development which includes and/or is based on copyrighted material of the original work.
[48] Unless specific permission was granted for commercial use by the rights-holder of X.
[49] Data redaction is a masking technique that enables the possibility to mask data by removing or substituting all or part of the field value and thereby protecting specific (sensitive) data.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

cases where access to the data is restricted or subject to conditions, the FAIR Principles and legal interoperability require that the metadata enables discovery, that the conditions and authorisation for access and use are clearly and readily determinable through automated means, and that they do not conflict with each other.

It is also worth noting in the context of this study that there are a number of "enabling legal instruments" that support legal interoperability and the implementation of FAIR or Open Data Principles. Such enabling legal instruments may come in the form of EU directives and regulations, national laws, EU and national policies, international agreements, contractual agreements, individual or institutional policies and other forms of practice that may incorporate broader policy considerations. An example of such an enabling legal instrument is the Open Data Directive, which requires that research data generated by public sector bodies (and funded by the public) follows the principle of 'open by default' and is made available in a manner compatible with the FAIR Principles. However, there is a need to examine whether obligations or recommendations to use certain licences, in particular at the national level, are coherent with specific recommendations that are or may be adopted by the EOSC Interoperability Framework.

To summarise, legal interoperability covers the broader environment of laws, policies, procedures and cooperation agreements needed to allow the seamless exchange of information and reusability of data between different individuals, organisations and across jurisdictions. It occurs among multiple datasets from different sources when:[50]

- the legal use conditions are clearly and readily determinable for each of the datasets, typically through automated means;
- the legal use conditions imposed on each dataset allow the creation and use of combined or derivative products;
- users may legally access and use each dataset without seeking authorisation from data generators on a case-by-case basis assuming that the accumulated conditions of use for each and all of the datasets are met; and
- when access to the data is restricted, metadata is FAIR, i.e., using accepted standards to describe the data and use conditions and enabling their discovery.

[50] Supra note 46. See also Catherine Doldirina et al., *White Paper: Mechanisms to Share Data as Part of GEOSS Data-CORE*, https://www.earthobservations.org/documents/dswg/Annex%20VI%20-%20%20Mechanisms%20to%20share%20data%20as%20part%20of%20GEOSS%20Data_CORE.pdf, accessed 19 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

CHAPTER II: THEMATIC ANALYSIS

1. Introduction

This chapter, which forms the main body of the study, follows a thematic structure and looks at a number of legal and regulatory impediments to legal interoperability to the implementation of the FAIR Principles within the context of the EOSC. These are divided into the following topics, addressed in each of the sections in this chapter:

- Copyright (including database rights) and licences;
- Other intellectual property rights (e.g. patents, trade secrets, neighbouring rights);
- Privacy and data protection (GDPR);
- Other restrictions and legitimate reasons (e.g., protection of sovereign genetic resources and traditional knowledge);
- Private law considerations.

The recommendations provided throughout this chapter, and those provided in the other chapters, will be repeated and summarised at the conclusion of the study.

2. Copyright

2.1. Background

Copyright is an intellectual property right that grants authors or creators of an original "work"[51] the exclusive right to reproduce or otherwise communicate the work and also make adaptations and modifications to the work.

Copyright law is not fully harmonised at the EU or at the international levels and it remains a matter of national law. There is however a common international framework for copyright, set out in the Berne Convention for the Protection of Literary and Artistic Works.[52]

There is no exhaustive list containing the works that can be protected by copyright. However, there are several basic principles set out in the Berne Convention that are generally applicable in all jurisdictions. The first principle is that in order to be considered as "copyrightable work" the work must be an intellectual creation of the author.[53] The second principle of copyright protection in works is that only the "mode or form of its expression"[54] is protected, but not the idea behind the work.[55] Therefore, ideas, processes, as well as factual data are excluded from the protection of copyright and thus not copyrightable. The third principle is that the work has to be "fixed in some material form".[56] The fourth principle, which is critical for legal

[51] The term "work" covers any fictional or descriptive representations in writing or speech as well as computer programs, music, poetry and dramas. In the context of this chapter, we use the term "works" to describe any potentially copyrightable data.
[52] The Berne Convention for the Protection of Literary and Artistic Works (9 September 1886) (the "Berne Convention").
[53] The term used in Art. 2 of the Berne Convention is "literary and artistic works".
[54] Art. 2(1) Berne Convention.
[55] See, for example, Art. 1(2) the Computer Programs Directive which makes clear that ideas and principles underlying any element of a computer program, including those which underlie its interfaces, are not protected by copyright.
[56] Art. 2(2) Berne Convention.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

interoperability in the context of the EOSC, is that the copyright protection is obtained automatically, it arises from the moment the work is created and no registration or other formality is required. This means that if such automatic rights are not waived, or if there is no clarity regarding the legal conditions under which a work may be used, the ability of researchers to access and reuse the data is impaired, and so does legal interoperability.

Another important aspect of copyright is the moral rights authors have in their work. These rights are personal and the Berne Convention states that "independently of the author's economic rights, and even after the transfer of said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honour or reputation".[57] The moral right to a work means that an author has the right to be attributed as such or not to be attributed if they no longer wish to be associated with the work. Moral rights remain with the original author, usually even after the transfer of all their economic rights, although the extent to which moral rights may be transferred or waived is left to national law. In the EU, moral rights generally cannot be waived or transferable but some exceptions exist.[58]

There are limitations and exceptions to the protection of copyright, which allow for copyrighted works to be used without a licence or without prior authorisation from the copyright holder. Such limitations and exceptions are set out in national legislations and differ between jurisdictions. At the EU level there are a number of noteworthy exceptions:

(1) Directive 2001/29/EC (the Copyright Directive)[59] has a list of around 21 exceptions and provides, among others, an exception for "use for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved".

(2) The Copyright in the Digital Single Market Directive[60] provides for mandatory exceptions, which Member States are required to implement into national law. Relevant for this study are the following exceptions:

- Text and data mining for scientific research purposes;[61]
- Exception or limitation for text and data mining;[62]
- Use of works and other subject matter in digital and transborder educational activities;[63]

[57] Art. 6bis (1) Berne Convention.
[58] For example, in Luxembourg, moral rights, except for the right to oppose any offence against the author's reputation, can be transferred lawfully. In the UK moral rights may be waived.
[59] Art. 5 Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L 167/10 ("Copyright Directive").
[60] 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 [2019] OJ L 130 ("Digital Single Market Directive").
[61] Art. 3 Digital Single Market Directive.
[62] Art. 4 Digital Single Market Directive.
[63] Art. 5 Digital Single Market Directive.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

- Conservation of cultural heritage.[64]

(3) The EU Orphan Works Directive[65] provides for a deviation from the general rule of copyright protection. The directive focuses on the digitisation of orphan works, i.e. copyrighted data (such as books, newspapers and films) where the copyright holder is unknown or cannot be located.

(4) "Fair dealing" and "fair use" are related concepts concerning a user's rights under copyright law. Fair dealing is an exception to copyright infringement laid under copyright laws in common law jurisdictions such as the UK and Australia. The copyright acts of these jurisdictions state that fair dealing does not infringe copyrighted work if the dealing is for specific purposes as specified in the act. Similarly, "fair use" is a concept under US copyright law[66] that provides limitations on exclusive rights in works of authorship, according to which the use of copyrighted work is not an infringement of copyright if it falls within the list of "purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use) and factors to be considered in determining the fair use".[67]

In the EU, copyright is a matter of national legislation. The EU has adopted certain measures aimed at harmonising copyright laws to the extent that divergence from or differences in national law hinder the proper functioning and development of the internal market.

Copyright was first subject to limited and fragmented harmonisation through the adoption of a number of directives, including:

- Directives concerning certain works (the so-called "Computer Programs" and "Database" directives);[68]
- Directives concerning certain rights (the "Lending and Rental" or the "Cable and Satellite" directives[69] and also, later, the "Resale Right" directive);[70] and,
- Directives concerning certain rules or aspects (the "Duration", the "Enforcement" and later the "Orphan Works" and "Collective Rights Management" directives).[71]

[64] Art. 6 Digital Single Market Directive.
[65] Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works.
[66] Title 17 of the United States Code.
[67] Supra note 66 § 107.
[68] Supra note 40 ("Computer Programs Directive"); Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20 ("Database Directive").
[69] Directive 2006/115/EC of the European Parliament and of the Council of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property [2006] OJ L 376/28 (originally Council Directive 92/100/EEC of 19 November 1992 on rental right and lending right and on certain rights related to copyright in the field of intellectual property [1992] OJ L 346/61) ("Rental & Lending Right Directive"); Directive 93/83/EEC on the coordination of certain rules concerning copyright and rights related to copyright applicable to satellite broadcasting and cable retransmission ("Satellite & Cable Retransmission Directive") as amended by the Directive (EU) 2019/789 of the European Parliament and of the Council of 17 April 2019 laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes ("CabSat 2 Directive").
[70] Directive 2001/84/EC of the European Parliament and of the Council of 27 September 2001 on the resale right for the benefit of the author of an original work of art [2001] OJ L 272/32 ("Resale Right Directive").
[71] Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the terms of protection of copyright and certain related rights [2006] OJ L372/12 (originally Council Directive 93/98/EEC of

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

A more general, cross-cutting harmonisation has been initiated with the so-called "Information Society Directive",[72] supplemented by a directive on collective management,[73] and amended by a directive on one specific exception, which is also regulated for some aspects by a first copyright regulation,[74] and by the recent directive on copyright and related rights in the Copyright in the Digital Single Market Directive.[75]

2.2. Copyright and research data

The protection under copyright is, as stated above, the result of a creative endeavour. Facts, as such, are considered to be "discovered" rather than "created" and therefore factual data or datasets that contain merely facts are not eligible for copyright protection (although another type of protection may apply).[76]

However, the way in which facts or data are described may be copyrightable if the description contains a creative element and if there is more than a limited number of ways to describe the facts or the data. Even metadata may contain creative elements, for example, if the metadata contains a short summary or a creative description of the data, that part of the metadata may be automatically protected by copyright.

In addition, the compilation or collection of data into a dataset or database may be subject to copyright. In such a case it is the collection itself, not the facts or data in the collection, which is protected under copyright, provided that the collection is an intellectual creation of an original collection.[77] This will be further explained below.

---

29 October 1993 harmonising the term of protection of copyright and certain related rights [1993] OJ L290/9) ("Term Directive"), amended by Directive 2011/77/EU of the European Parliament and of the Council of 27 September 2011 amending Directive 2006/116/EC on the term of protection of copyright and certain related rights [2006] OJ L 265/1 ("Term Extension Directive"); Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works [2012] OJ L 299/5 ("Orphan Works Directive"); Corrigendum to Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights [2004] OJ L 195/16 ("Enforcement Directive").

[72] Supra note 59.

[73] Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market [2014] OJ L 84/72 ("Collective Rights Management Directive").

[74] Directive (EU) 2017/1564 of the European Parliament and of the Council of 13 September 2017 on certain permitted uses of certain works and other subject matter protected by copyright and related rights for the benefit of persons who are blind, visually impaired or otherwise print-disabled and amending Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society [2017] *OJ L 242*, and Regulation (EU) 2017/1563 of the European Parliament and of the Council of 13 September 2017 on the cross-border exchange between the Union and third countries of accessible format copies of certain works and other subject matter protected by copyright and related rights for the benefit of persons who are blind, visually impaired or otherwise print-disabled [2017] OJ L242/1.

[75] Supra note 60.

[76] Data may be protected by other rights or laws such as database rights, confidential information, privacy laws, security, etc. Multiple types of protection may apply to parts of the research data, while there may be other parts that have no legal protection.

[77] See Art. 5 WIPO Copyright Treaty (2002): "Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such. This protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation".

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Therefore, before reusing data, it is important that repositories, disseminators and users of data ascertain whether the data, parts of it, or other embedded elements in the data (e.g., pictures or flowcharts) are subject to copyright protection; who holds such rights and what the conditions are for reuse. If the data does not contain any creative element, and is merely composed of facts, it is not protected under copyright and it could be reused without risk of breaching copyright law. However, if the data or parts of it contain an element which falls under copyright protection, then that protected element(s) of the data require a licence or a waiver of rights before the data or dataset may be reused by third parties[78] – see **Recommendation 4**.

2.3. Copyright and metadata

As discussed in Chapter 1, metadata is ancillary information about the data or the dataset and is a key component required for the application of the FAIR Principles. FAIR requires to openly and richly describe the context within which the data was generated, to enable the evaluation of its utility, to explicitly define the conditions under which it may be reused and to provide clear instructions on how it should be cited when reused.[79] Metadata plays a crucial role in this process of making the data FAIR. In most cases, metadata will not include any copyrightable elements and will not be subject to copyright protection. As such, it could be made public and freely re-used by others. However, to the extent that metadata includes copyrightable information, for example a summary describing the data in a creative way, this may amount to copyrightable information. Any such description of the data that is more than a statement of facts can be viewed as an intellectual creation of an original work protected under copyright. In such cases, and unless other legal restrictions apply,[80] any applicable copyright to sections of the metadata should be waived and/or assigned a permissive licence with no restrictions, ideally by way of the CC0 or the PDDL. Applying a CC0 or PDDL to metadata means that if any copyrights exist, they will be waived, but if they do not exist, the CC0 or PDDL do not create any obligations.

Recommendation 3: Copyrightable **metadata** should be free from any restrictions and assigned a public domain waiver. The Creative Commons No Rights Reserved (CC0) or the Public Domain Dedication and Licence (PDDL), or an equivalent statement of rights should be preferred.

Often, the assessment whether parts of the metadata are subject to copyright protection will have to be made on a case-by-case basis, which in itself, becomes an impediment to legal interoperability.

[78] Note that other relevant considerations may apply when data is reused, such as liability, restrictions on the use of the data, etc. The discussion above only addresses copyright aspects.
[79] See GoFAIR, What is the difference between "FAIR data" and "Open data" if there is one? (GoFAIR) https://www.go-fair.org/resources/faq/ask-question-difference-fair-data-open-data/, accessed 19 November 2020.
[80] For example, a request for erasure of personal data on the basis of privacy rules, see further in section 4.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

CESSDA ERIC is the Consortium of European Social Science Data Archives. It is a distributed European infrastructure that brings together social science data archives and their trusted repositories from across Europe and aims at moving from the current fragmentation to a situation where data is easy to store, find and reuse. One of the services provided by CESSDA is a common metadata catalogue for finding, seamlessly accessing and reusing relevant data (FAIR data). For data to be findable it has to be included in as many search engines as possible, and for this purpose, the metadata should provide a description easily findable by search engines. CESSDA needs to assess if: (1) the existing description is "creative" enough to be subject to copyright, in which case they must seek approval of the original author. This may require a case-by-case analysis of approximately 40-50 thousand datasets; and, (2) if the description is creative, CESSDA may need to trace and contact the author or the right-holder to get the required approval, which may be very difficult or impossible if the description was produced many years ago.

As mentioned, metadata is all data about data. In order to facilitate the accessibility and re-use of the data described by the metadata, it is very important that the metadata entails clear information on any types of restrictions on the use of the data that it describes. Furthermore, for metadata to be useful and FAIR, it must be standardised and both machine and human readable to enable advance research techniques.

Recommendation 4: Metadata should include a standardised human and machine-readable statement of rights, legal restrictions, applicable licences, and, where relevant, additional conditions of use (including applicable jurisdictions) of the data that they are assigned to.

If also follows from Recommendation 4 that, in order to be effective, metadata should be standardised, to the greatest extent possible, in order to facilitate interoperability.

Recommendation 5: The EOSC should provide a mechanism, for example in the Rules of Participation, or by way of guidance, to facilitate the implementation of Recommendation 4 above in a harmonised manner.

2.4. Databases and research data

Databases play an important role in cloud computing federations. They are used to keep records and are essential for many types of activity, both commercial applications and scientific research such as genomic, biochemistry, astronomy, geology, etc. From a technical

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

perspective, a database is a "repository of data"[81] or a "set of resources" which includes the storing, maintenance and the provision of the input and output routines for the users.[82]

Examples of databases abound in the cloud. Scientific databases such as biological databases can best illustrate the legal and technical issues. They are evolving from private cloud models to more dynamic hybrid and federation scenarios. Genetic databases containing genetic information and images such as BBMRI-EU,[83] Euro-BioImaging[84] and the European Molecular Biology Laboratory (EMBL)[85] are all prime examples of complex database management systems.

From a legal perspective, database rights, also known as *sui generis* rights, are defined by Directive 96/9/EC (the "Database Directive")[86] 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."[87] Recital 17 of the Database Directive indicates that independent works, data or other material as such are not covered by the Database Directive, only the "collections" of such independent works, data or other materials are protected under the scope of the Database Directive. These may include collections of literature, art, music or material such as texts, sound, images, numbers, facts, and data, but it explicitly excludes independent works such as a recording, audio-visual, cinematographic, literary or musical work.

The Database Directive offers a double scheme of protection: (1) Copyright law: the selection or arrangement of the contents are the author's own intellectual creation, in which case copyright protection applies to the structure of the database (separately and in addition to any copyrightable content of the database). Within this scheme, the test is the same as under copyright and no other criteria need to be applied to determine whether the database is eligible for protection.[88] (2) The *sui generis* right, which protects the "investment" in obtaining, verifying or presenting the contents of a database (as a compilation of data). This database right offers protection in circumstances where copyright protection is not available based on the resources that database makers invest at the moment of creating, updating and presenting the content of databases.

Article 7 of the Database Directive grants exclusive property rights to database makers who can demonstrate that they have made a "substantial investment" measured in terms of quantity and/or quality in either the obtaining, verifying or presenting the contents of the database. In such cases the Database Directive imposes limitations and restrictions on the extraction and re-utilisation of the content of such databases. This could potentially be an additional layer of impediments to legal interoperability.

---

[81] Dan Simovici and Richard Tenney, *Relational Database Systems* (Academic Press, 1995), p.1.
[82] R. A. Frost, *Database Management Systems: Practical Aspects of their Use* (McGraw-Hill Book Company, 1984), pp. 3-5.
[83] BBMRI-ERIC is a European research infrastructure for biobanking. See https://www.bbmri-eric.eu, accessed 7 August 2020.
[84] Euro Bioimaging. See: https://www.eurobioimaging.eu, accessed 7 August 2020.
[85] European Molecular Biology Laboratory (EMBL). European Molecular Biology Laboratory (EMBL). See: https://www.embl.de, accessed 7 August 2020.
[86] Supra note 68.
[87] Art. 1(2) Database Directive.
[88] Art. 3(1) Database Directive.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

What constitutes a "substantial investment" is assessed on a case-by-case basis, taking into account aspects that are directed at obtaining, verifying or presenting the contents of the database, such as the use of financial resources, time spent, efforts made, human resources and the use of technical equipment. An example would be the special skills of a curator that are used in the creation of the database and the time spent in verifying the content of such database.[89]

While the definition of "substantial investment" is broad, it has some limitations. A way of identifying the extent of the legal protection afforded to a database is by determining the mechanisms – obtaining or creating – used for collecting the data.[90] In the case of *British Horse Racing Board (BHB) v. William Hill,*[91] the Court of Justice of the European Union (CJEU) held that the investment has to be in the obtaining, verification or presentation of the contents of the database. This is to be distinguished from the investment in the creation of data itself which cannot be taken into account when considering database rights. The CJEU therefore did not grant the legal protection to BHB because while making these types of (sports betting) databases, the data to organise the fixture of the sport events was "created" and not "obtained" from other sources.

Similar to copyright, there is no need for the registration of databases. Database rights arise automatically at the moment of their creation and last for 15 years,[92] which may be extended for another term given that any substantial modification has been made.[93] This means that almost any compiler of independent works, data or other materials who made a "substantial investment", may enjoy *sui generis* right protection unless expressly excluded or waived.[94]

The *sui generis* right causes some concerns regarding its practicality and flexibility to modern data processing technologies. Its failure to come to terms with new technological advances and with the onset of cloud computing services along with the open data movement may obstruct scientific research activities.[95] In the absence of an appropriate licence scheme or contractual framework database rights may lock up data because they grant protection to database makers. Certain legal issues could emerge, particularly if this is considered from a global cloud computing and open data perspective,[96] in particular with regards to *access* to information, *reuse* of data and legal interoperability.

[89] Arthur Lesk, *Introduction to Bioinformatics* (3rd edn, Oxford University Press 2008), p. 153.
[90] Jens Gaster, "Obtinere" of Data in the Eyes of the ECJ (2005) 6 *Computer Law Review International* 129-135, doi: https://doi.org/10.9785/ovs-cri-2005-129.
[91] Case C-203/02, *The British Horseracing Board Ltd and Others v. William Hill Organization Ltd* ECLI:EU:C:2004:695, [2004] ECR I-10415 ("the BHB case"). Similarly, see also, case C-444/02, *Fixtures Marketing Ltd. V. Organismos Prognostikon Agonon Podosfairou* ECLI:EU:C:2004:697 [2004] (Greece) ("the OPAP case"); case C-46/02, *Fixtures Marketing Ltd. v Oy Veikkaus AB* ECLI:EU:C:2004:694 [2004] (Finland); case C- 338/02, *Fixtures Marketing Ltd. v Svenska Spel AB* ECLI:EU:C:2004:696 [2004] (Sweden).
[92] Art. 10(1) Database Directive.
[93] Art. 10(1)(2) and (3) Database Directive.
[94] Art. 7(1) and Recitals 7, 13, 14, 17 and 40 Database Directive.
[95] Marcelo Corrales Compagnucci, *Big Data, Databases and 'Ownership' Rights of Data in the Cloud* (Springer 2019), pp. 2, 272.
[96] Ibid., pp. 20, 272.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

In a public consultation conducted by the EU Commission in August 2017,[97] many of the respondents considered that the Database Directive is outdated and requires modifications taking into account the recent technological and economic developments. One issue was whether the *sui generis* right struck a good balance between database makers and users. Views were polarised about the impact of database rights, especially with regard to the re-use of data.[98]

The Joint Institute for Innovation Policy led a study based on the data collected from the public consultation and the EU Commission released the Evaluation Report on the Database Directive in April 2018.[99] The Evaluation Report concluded that some of the provisions of the Database Directive are no longer fir for purpose, in particular in an increasingly data-driven economy. Arguments in favour of repealing the Database Directive were considered. If a full abolition is not possible, the EU Commission might consider modifying some of the provisions, in particular those related to machine-generated data.[100]

By and large, the main legal problem in connection with the Database Directive is the idea of conventional databases which have a fixed structure in which data is stored and gathered in a physical space, combined with the territorial scope of protection granted by the Database Directive. This is undoubtedly an anachronistic approach which is not consistent with the ubiquitous nature of cloud federations. Servers can be located in different countries, including countries outside of the EU/EEA, and databases can be easily replicated in virtual machines (VMs), which could create legal hurdles among the stakeholders involved.[101] Moreover, servers could be potentially "exported" to a jurisdiction outside of the EU/EEA offering no database right protection.

From an EOSC perspective, the protectionist approach of the Database Directive is problematic because it automatically frames *access* to data as a threat. This is counter to legal interoperability and in general to the Open Science idea. There should be more balance between the protection of databases on the one hand and access and re-use of data on the other.[102]

In practice this means that only users within the EU would need to obtain permission for re-using the whole compilation of the database, while users outside the EU would not be required to do so, due to the territorial nature of database rights. There is therefore a need to ensure that

[97] See European Commission, Summary Report of the Public Consultation on the Evaluation of Directive 96/9/EC on the Legal Protection of Databases (European Commission, 6 October 2017), https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-legal-protection-databases, accessed 18 July 2020.
[98] Ibid.
[99] European Commission, *Study in Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases* (European Union 2018) DOI: 10.2759/04895.
[100] Ibid. With regard to technological changes, the European Commission concluded the following: "(i) it is not (yet) clear how the *sui generis* right interacts; (ii) it could be advisable to clarify the notion of database maker; (iii) as far as possible, clarify the notions of substantial investment and substantial part including the notion of recorded and of created data; (iv) introduce a text and data mining exception; (v) as with European Commission's own conclusion to the Digital Economy Package, it is advisable to wait before proceeding to a legislative intervention in this respect." See also, Timothy Vollmer, 'Database Directive Study: Options for Neutralising the *Sui Generis* Right' (COMMUNIA, 10 May 2018), https://www.communia-association.org/2018/05/10/database-directive-study-options-neutralising-sui-generis-right/, accessed 10 July 2020.
[101] Supra note 95, p. 7.
[102] Supra note 95, pp. 7, 271.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

automatic database (*sui generis*) rights do not impose unintended restrictions on re-use for EU-based users (compared to non-EU users, that are not restricted by database rights due to the territorial nature of such rights).

Recommendation 6: Automatic database (*sui generis*) rights should be addressed in the licence chosen so that: (1) *sui generis* are covered by any permissive licence (or a waiver), and (2) they do not result in unintended restrictions on reuse of data by EU-based users compared to non-EU users (due to the territorial nature of such rights).

2.5. Who owns copyright or database (*sui generis*) right?

Copyright and database (*sui generis*) rights may be owned by an individual who created the work or by another entity (for example a funder, or an employer). In the case of database rights, these are owned by the database maker, i.e., the person that made the investment, in terms of time, technical equipment and human resources, and bore the financial risk in creating the database, usually the employer or the repository (if owned by someone else). Ownership of copyright may depend on funding conditions, applicable law, contractual arrangements or employment conditions between an institution and a researcher.

Equally, each dataset may, in practice, include different copyrightable assets ('embedded data'). For example, if a photo is embedded in a dataset, the photo may be subject to a separate licence and it may not always be easy to identify the correct copyright holder and applicable licence. Data producers and users may not be aware of the entire scope of copyright related to each component of a specific dataset and could therefore inadvertently infringe a third party's right when using, e.g., the photo.

The assessment of ownership of copyright and *sui generis* rights should be made on a case-by-case basis. There may be situations, for example, where the database (*sui generis*) right lies with the employing research institution who owns the repository, the copyright in the content of the data lies with the employee (or with another party if the employee transferred their copyright to that party) while copyright to other embedded data lies with a third party.

It is therefore important to verify the ownership rights in detail in relation to each component of the data and to clarify that the copyright in the different components lie with different parties and may be distributed under different licence terms.

Recommendation 7: All copyrightable components of the research data and their respective licences should be clearly identified in the metadata and assigned the correct rights-holder. In the case of database (*sui generis*) rights in repositories, the applicable (permissive) licence should be included in the terms of use of the repository.

2.6. Waivers and licences

2.6.1. General

As discussed in Chapter 1, FAIR data is not equivalent to Open data. The "A" in FAIR stands for "Accessible under well-defined conditions", implying that there may be legal or other legitimate reasons to retain access and re-use of data or parts of it restricted. However, from a

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

strictly legal interoperability perspective, the more the data is open, the less impediments to legal interoperability will occur. In the case of copyright or *sui generis* restrictions, the easiest and most effective way to achieve legal interoperability is to waive all rights and make the data part of the public domain, using the CC0 for example (as further discussed below). If the data needs to be licensed, it is recommended to use one of the least restrictive licences available. Licences such as the MIT License (for software) or the CC BY (for data other than software) are permissive licences that only require that credit is given to the author of the original work and are preferable in terms of legal interoperability over licences that introduce additional conditions or restrictions.

In practice, there will be cases where, for legitimate reasons, such as privacy, national security, sensitive information, etc., data needs to be shielded. The degree to which any such data is made available is at the discretion of the data rights-holder(s). In such cases, different methods may be used, such as generalisation of data, redaction of specific information, anonymisation[103], embargo periods, etc. However, there will be cases where no such methods would be sufficient and a standard open licence may not be suitable, so the data rights-holder(s) will need a customised licence or a specific contract regulating access to the data. This will most likely impair legal interoperability and it is therefore important to keep data "as open as possible, as closed as necessary".

Recommendation 8: Open and permissive licences, authentication and authorisation mechanisms and the use of restricted data access collections are preferred over the use of ad-hoc specific contracts entered into between a rights-holder and a data user. An additional contract should only be used if it is the least restrictive way to ensure compliance with legal restrictions or in other justified cases. In such cases smart contract solutions should be developed and put into use.

2.6.2. Waivers

When the intention is to make data (and metadata) available for reuse without restrictions, it is often conducted by way of dedicating the work to the public domain or waiving all rights to the data, to the extent permissible under law. The two most common forms for this purpose are:

- The Creative Commons No Rights Reserved (CC0) which means that the rights-holder(s) waives its copyrights in the data; or
- The Open Data Commons Public Domain Dedication and Licence (PDDL), which places the data in the public domain.

The PDDL is primarily used for databases and the CC0 is used for both databases and other copyrightable objects. As discussed, in many jurisdictions the moral right to a copyrightable object cannot be waived.

[103] In case that personal data is included, see section 4.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

The CC0 and the PDDL differ from each other in the manner in which they approach they restrictions on the possibility to waive moral rights. While the CC0 is silent on this issue,[104] the PDDL includes a licence that prohibits the possibility to invoke moral rights in cases when such rights cannot be waived under applicable national laws.

When the data or dataset has been dedicated to the public domain or when all rights to the data are waived, users are free to reuse the data or the dataset (subject to any non-waivable moral rights). Data in the public domain can be combined freely with any other data or dataset and the derivative work can be distributed in the manner decided by the creator of the derivative work.

There are other considerations that may influence the choice of the waiver. The CC0 licence provides no disclaimer or limitation of liability and, depending on the type of data and jurisdiction, the rights-holder may be exposed to liability risks in case the data is incorrect. In cases where such a concern exists, the data rights-holder(s) may consider a licence that includes a disclaimer of warranty, such as the Creative Commons Attribution (CC BY) licence or the MIT License for software, although there are other, arguably better, ways to address liability issues.[105]

Recommendation 9: Copyrightable data should be FAIR and, to the greatest extent possible, be made part of the public domain or assigned a permissive licence, unless legal or legitimate reasons apply. The Creative Commons No Rights Reserved (CC0) or the Public Domain Dedication and Licence (PDDL) or an equivalent statement of rights should be preferred. In cases where liability is a concern that cannot be addressed by other means, the CC BY 4.0 licence is an appropriate alternative.

2.6.3. Standard licences (common-use licences)

2.6.3.1. General

Licences lie at the core of legal interoperability and data reusability as researchers may need to combine different datasets from multiple sources in order to create an additional work. Depending on the type of licence used for each dataset, legal interoperability requires that: (1) the conditions of use of one dataset do not negate the conditions of use of another dataset, so the two (or more) datasets can be combined and carried forward; and, (2) that the accumulated restrictions carried forward under the combined dataset are not more restrictive than the initial conditions of use for one (or more) of the original datasets (the "lowest common denominator" effect).

In order to understand the meaning of these requirements it is first important to discuss the most commonly used licences.

The basic prerequisite for licensing is that the component of the data for which the licence is granted is copyrightable or protected by any other IP right. If the data is not protected by

[104] The CC0 clarifies that it only waives the rights as long as this is permitted under applicable law. Therefore, when moral rights apply, the CC0 will not affect them.
[105] See section 6.2.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

copyright or other IP right, there is no need for a licence and the data can be shared and re-used freely (subject to any other contractual or legal limitations).

Some jurisdictions, mainly common law jurisdictions, make a distinction between licences and contracts, based on the legal instrument used to enforce the relevant right (the licence or the contract). Civil law jurisdictions do not make a distinction between licences and contracts, and the licence is enforced in the same way as a contract.

A licence is in essence a contract under which the holder of the copyright (or any other IP right) grants permission for the use of the data to another person(s), within the limits set by the provisions of the licence.

There are a number of standardised open source licences which are described below.

2.6.3.2. Software licences

Open source licences are generally categorised as either permissive licences or non-permissive licences.

A permissive licence does not require the licensee to make the derivative work open or available in the public domain, or to provide the source code to downstream users. However, if the original work is also sub-licensed, then the original licence terms must be applied to the original work, while the derivative work may be distributed freely under any licence. The MIT License is the most common permissive licence used for software.[106]

Non-permissive licences generally include restrictions on the redistribution of derivative works. Any derivate work under a non-permissive licence must be made available under the same licence terms as the original work (referred to as "copyleft").[107]

Examples of non-permissive software licences include GNU GPL (v 2 and 3) and the EUPL 1.2.[108]

There is a large number of open source software licences. There are also many groups and organisations that review and approve open source licences, most notably the Open Source Initiative ("OSI")[109] and the Free Software Foundation.[110] Currently, there are over 100 open source software licences approved by the OSI.

[106] Apache and Berkley Software Distribution are also commonly used permissive licences. For comparison, the Creative Commons Attribution licence (CC BY) is a common permissive licence which is used for non-software works.
[107] Copyleft licences are open source software licences allowing the use, modification and distribution of software on the condition that its modifications (the derivative work) are distributed under the same terms and conditions as the original work.
[108] For comparison, the Creative Commons Share Alike (CC- BY-SA) is a non-permissive licence, which is used for non-software work.
[109] Open Source Initiative, https://opensource.org/, accessed 19 November 2020.
[110] Free Software Foundation, https://www.fsf.org/, accessed 19 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

According to the OSI, the following nine software licences are the most popular, widely used, or used by strong communities:[111] Apache License 2.0; BSD 3-Clause "New" or "Revised" license; BSD 2-Clause "Simplified" or "FreeBSD" license; GNU General Public License (GPL); GNU Library or "Lesser" General Public License (LGPL); MIT License; Mozilla Public License 2.0; Common Development and Distribution License; and, Eclipse Public License version 2.0.

According to GitHub[112] the most widely used standard open source licences are: MIT License; GPL v 2; Apache License 2.0; GPL v 3; and, BSD 3-Clause.[113]

Below we set out an analysis and assessment of the MIT License, the Apache License 2.0, GPL v 3 as well as the European Union Public Licence version 1.2 (EUPL 1.2). The analysis includes a description of the licence, as well as upstream and downstream compatibilities, which are critical to legal interoperability.[114]

2.6.3.2.1. The MIT License

The MIT License is the most commonly used open source licence. The licence is very permissive and only requires that the copyright notice[115] and the licence must be included in all copies or substantial portions of the software. There are no other restrictions (other than the copyright notice and the licence of the original work), which means that it can be distributed under any licence, open or commercial. There are no requirements on making source code available when the software is licensed in object code.

Since the MIT License is permissive the licence is downstream compatible with almost all licences, provided that copyright notice is given. It should be noted that there are a few modified versions of the MIT License which include minor additional requirements, for example the X11 License or the Expat License.

2.6.3.2.2. Apache License 2.0

The Apache License 2.0 is a permissive licence and imposes no requirements on the licensing of derivative work other than notification that changes have been made to the original work. If the original work is distributed by the licensee the original work must be licensed under the Apache License 2.0.

---

[111] Open Source Initiative, Licenses and Standards, https://opensource.org/licenses, accessed 19 November 2020.
[112] An open source software repository, see GitHub, https://github.com, accessed 19 November 2020.
[113] The list is in accordance to the order of popularity. The WhiteSource software platform identifies the same licences as being the most widely used but in a different order, as follows: MIT, Apache License 2.0, GPL v3, GPL v2 and BSD 3-Clause, see Ayala Goldstein, Open Source Licenses: Trends and Predictions (*WhiteSource*, 23 January 2020), https://resources.whitesourcesoftware.com/blog-whitesource/top-open-source-licenses-trends-and-predictions, accessed 19 November 2020.
[114] Upstream compatibility is used for determining which source code licence can be incorporated in source code from another licence and then relicensed under the latter licence. Downstream compatibility is used to determine under which other licences the derivative work for source code licensed under the given licence can be distributed.
[115] MIT.edu, https://www.mit.edu/~amini/LICENSE.md , accessed 19 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Unlike other permissive licences that apply only to copyrightable work, the Apache License 2.0 can be applied to both copyrightable work and to patents.[116]

The Apache License 2.0 is compatible with the GNU GPL v3 (but not with versions 1 or 2), which means that codes licensed under GPLv3 and/or Apache License 2.0 can be combined and licensed under the GPLv3. In such cases also the original work licensed under Apache License 2.0 can be distributed under GPL v 3.

2.6.3.2.3. GNU GPL v 3

GNU GPL v 3 is a copyleft licence, which means any derivative work must be made under the same licence. The purpose of this is to keep the work open source, but at the same time it restricts the licensee's freedom regarding the choice of licensing of the derivative work. It also means that the licensee is not allowed to mix different codes licensed under different copyleft licences.

For these reasons, and because copyleft licences are generally not compatible with each other, they may create impediments to legal interoperability (even though they are open source).

Similar to the Apache License 2.0, the GNU GPL v 3 includes a patent licence in order to further protect licences against infringement claims.

2.6.3.2.4. European Union Public Licence version 1.2

The EUPL v 1.2 is a licence created on the initiative of the European Commission[117] and is a non-permissive licence which includes a copyleft clause. This means that any derivative work[118] must be licensed under the EUPL 1.2 or later versions, unless expressly permitted.[119]

The licence also includes a compatibility clause, which is an exception to the copyleft clause, and allows the distribution of derivative work under another licence, as identified by the EUPL v 1.2 and under certain circumstances.

More specifically, this is limited to the situation where the derivative work is based both on original work licensed under the EUPL 1.2 and original work licensed under another compatible licence. In such situations the entire derivative work can be distributed under the compatible licence instead of the EUPL 1.2. The compatible licences are listed in an appendix

[116] The licence is terminated if the user sues anyone over patent infringement related to the work.
[117] Commission Implementing Decision (EU) 2017/863 of 18 May 2017 updating the open source software licence EUPL to further facilitate the sharing and reuse of software developed by public administrations [2017] OJ L 128/59.
[118] There is a discussion regarding "linking", i.e. an application that works through linking to different components, and if such a linking constitutes a derivative work or not. There are different opinions on this and to our knowledge no relevant case law within the EU Member States regarding this. The Computer Programs Directive (see footnote 40) addressed the issue of creating interoperability of independent software programs (Art. 6), and was interpreted by the CJEU in case C-406/10, *SAS institute Inc v. World Programming Ltd* ECLI:EU:C:2012:259 [2012]. However, for the purposes of this study we do not need to analyse the exact definition of derivate work but merely note the limits that a licence may have on the distribution of derivative works.
[119] EUPL V 1.2 section 5.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

to the licence and include, for example, GNU GPL v 2 and v 3, and CC BY-SA for works other than software.

The compatibility clause concerns the downstream compatibility, i.e. the possibility to merge work received under a EUPL into a work that is distributed under a compatible licence (as defined in the EUPL 1.2) and the combined work being distributed under the compatible licence.

In an upstream compatibility, the question to be assessed is whether code distributed under any other free or open source software licence may be licensed under EUPL 1.2 or incorporated in work licensed under EUPL 1.2. In such cases, the EUPL 1.2 allows for instances where such codes are distributed under the EUPL 1.2. This is the case, for example, with work licensed under LGPL (MPL).[120] However, GNU GPL v 2 or 3, on the other hand, is not upstream compatible with the EUPL v 1.2, meaning that work received under GNU GPL v 2 or v3 cannot be combined with work licensed under EUPL 1.2 and licensed under EUPL 1.2. This is a potential impediment to legal interoperability.

Recommendation 10: The use of Creative Commons licences is generally not recommended for licensing source code for software. Only open and permissive software licences such as the MIT License, the Apache License 2.0 or the equivalent should be used for software.

2.6.3.3. Non-software licences

2.6.3.3.1. Creative Commons

Creative Commons (CC)| is a not-for-profit organisation which makes a number of licences available for copyright (or database right) protected works. The CC licences are widely used for the licensing of data and databases, but they are not so commonplace for software and source code.

The structure of the CC licences is based on a number of building blocks, which may be combined in a number of different ways. The different building blocks are:

a) Attribution (CC BY) – users are required to give credit to the creator;
b) Share-Alike (CC SA) – any derivative work or adaptations must be shared under the same licence as the original work;
c) Non-Commercial (CC NC) – the licensed work may only be used for non-commercial purposes; and,
d) Non-Derivative (CC ND) – derivative work or adaptations are not permitted.

When using a CC licence the downstream user will automatically receive a licence to the original work according to the original licence. The new derivative work however, may be licensed differently from the original licence, although it must remain subject to the terms of the original licence.

[120] This licence is not analysed further in this study.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

For example, if the original data is licensed under the CC BY, a user can licence derivative work under any other licence, provided that credit for the original work is given to the creator of that original work. On the other hand, if the original data is licensed under the CC BY-SA the user must licence the derivative work using the same licence, CC BY-SA. There is thus a great restriction in the possibility of licensing derivative work. All CC licences include a disclaimer of warranty and a limitation of liability.

The following table provides a short description of each CC licence:

Licence	**Description**
Creative Commons Attribution (CC BY) 4.0	The CC BY 4.0 licence allows everyone to re-use, distribute and modify the licensed materials. The licence must however, in the case of distribution of the original licensed material, identify the creator of the work as well as any other rights-holder that according to the licence shall receive attribution, provide a copyright notice; make reference to the CC BY 4.0 licence and information about any modified content. Modified content, i.e. derivative work, can be distributed using any licence.
Creative Commons Attribution Share-Alike (CC BY-SA)	This licence is the same as the CC BY with the added restriction that any derivative work must be licensed under the same licence. The Share Alike has similar traits as copyleft under software licences.
Creative Commons attribution Non-Commercial (CC BY-NC)	This licence is the same as the CC BY with the added restriction that any derivative work must be for non-commercial use.
Creative Commons Attribution Non-Commercial Share-Alike (CC BY-NC-SA)	This licence is a combination of CC BY-SA and CC BY-NC, prohibiting the commercial use of the derivative work and also requires that the derivative work is licensed under the same licence.
Creative Commons Attribution Non-Derivative (CC BY-ND)	This licence is the same as the CC BY with the additional restriction that derivative work is not permitted.
Creative Commons Attribution Non-Commercial Non-Derivative (CC BY-NC-ND)	This licence is the same as the CC BY-ND with the additional restriction that commercial use is not allowed.

Two of the most commonly used assigning CC licences for public data and metadata records are CC0 (see above under waivers) and CC BY. The use of CC NC or CC ND licences for data

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

or databases intended for scholarly or scientific use is discouraged, as these licences are not completely open[121] and they may impede legal interoperability.

GEOFON attempts to facilitate cooperation in seismological research, earthquake monitoring and tsunami risk mitigation. It is a large seismological archive system with open data access for own data as well as data curated for partner institutions. GEOFON is one of the fastest data centres, providing earthquake location information, including magnitudes and moment tensors continuously and in real time. Leveraging on the work started more than a decade ago in data management by the GFZ library, GEOFON has initially archived and distributed data from GIPP experiments recommending a CC BY-SA licence. Recently, in the context of the Helmholtz Association, and taking advantage of national third-party projects fostering data FAIRness, the default suggested licence has been changed to CC BY 4.0. The latter is also the recommended licence by EPOS and other European initiatives. Ownership rights remain with the users of GIPP instruments, but they must accept to make data available to the community under a CC BY 4.0 licence after a short embargo period (max. 4 years).

A few noteworthy differences emerge when comparing the CC0 with the CC BY:

The CC BY includes a disclaimer of warranty and a limitation of liability, whereas the CC0 does not. The need for, and importance of, such a disclaimer varies, depending on the nature and type of data used and for what purpose – see **Recommendation 9**.

The CC BY licence requires attribution, meaning that users must give credit to the generator of the original work. The CC0 does not include such a requirement. In the scholarly or scientific communities, it is questionable if attribution should be required by way of a legal (licensing) obligation, or whether this should be achieved by other means. As stated by Doldirina (et al.):[122]

"*If attribution is desired by providers for research data, as is usually the case, the citation to parent datasets as a normative practice in the academic and scientific communities can fulfill an equivalent role. Data citation may be promoted through development of efficient technological means for facilitating the practice, inclusion of the practice in codes of conduct, and incorporation into the requirements of research publication outlets*".

Attribution is a legitimate requirement for recognition by producers of research data. However, using the CC BY is not the only way, and possibly not the best way, to achieve such a recognition. As discussed, not all the components of the data are likely to be protected by copyright. Requiring users to abide by the terms of a CC BY licence in relation to the entire data means that it will include non-copyrightable components, which will not be enforceable,

[121] According to the definition of open by the Open Knowledge Foundation. See: Open Knowledge Foundation, http://opendefinition.org/, accessed 19 November 2020. See also the Bethesda Statement on Open Access Publishing (20 June 2003), available at: http://legacy.earlham.edu/~peters/fos/bethesda.htm, accessed 19 November 2020; the Budapest Open Access Initiative (14 February 2002), available at: https://www.budapestopenaccessinitiative.org/read, accessed 19 November 2020, and the Berlin Declaration on Open Access (22 October 2003) available at: Open Access Max Planck Gesel, 'Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities', https://openaccess.mpg.de/Berlin-Declaration, accessed 19 November 2020.

[122] Supra note 46, p.33

but may be misleading, hence undermining the value of the licence. Furthermore, when data is licensed under CC BY and is distributed and re-used downstream several times in relation to derivative works, it may become difficult to keep track of what licence terms are applicable to each derivative work. The CC0 does not have the same problem. Using the CC0 instead of CC BY and relying on other forms of attribution such as the European Code of Conduct for Research Integrity or offering a citation by reference to Persistent Identifiers, etc. or simply asking for credit or attribution (rather than requiring attribution by legal means), is preferred from a legal interoperability perspective.

Recommendation 11: From a licence compatibility perspective, attribution should be pursued by means of moral and ethical obligations e.g., the European Code of Conduct for Research Integrity or the development of Persistent Identifiers, or by way of a standard form of acknowledgement, rather than by means of a licence such as the CC BY 4.0. The CC0 is, in general, preferred over the CC BY 4.0, although both are generally permissive.

2.6.3.3.2. Other non-software licences

Although the CC licences are the most frequently used licences there are several additional open licensing options.

Open Data Commons:
Similar to the CC, the Open Knowledge Foundation makes the licences under the Open Data Commons available for use to licence databases. The Open Data Commons Attribution License (ODC-BY) is a licence used for distributing databases, both protected under the EU database directive and as a copyrightable compilation of data.

It should be noted that the actual content of the database is not covered by this licence. In case the content of the database is copyright protected, it will have to be licensed separately. The structure of the ODC-BY is similar to the CC BY in that the database cannot be sub-licensed by a licensee. According to the terms of ODC-BY, if a licensee distributes the original database, the downstream sub-licensee will automatically receive a licence for the original work directly from the original licensor.

Open Use of Data Agreement (O-UDA):
The O-UDA is a result of an initiative by Microsoft to make it easier for individuals or entities that want to share or distribute data to do so, with no restrictions on its use. The O-UDA is a simple and useful alternative to the ODC-BY and CC licences. The licence object is defined as "materials you receive under the O-UDA" and does not specify in detail the different intellectual property rights that are covered by the licence. Therefore, databases under EU legislation, data included in the database and other copyrightable material can be licensed under the O-UDA.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Recommendation 12: The overall number of recommended licences for data available through the EOSC should be minimised to the greatest extent possible. In addition to the CC0, CC BY and MIT, permissive licences such as the ODC-BY or the O-UDA are preferrable along with information about their compatibility with other licences – see **Recommendation 14**.

2.6.4. National licences

A number of EU Member States have adopted their own national open licences for public data. Noteworthy among them are the UK Open Government Licence (OGL) and the French Open Licence 2.0, both of which are compatible with CC BY, ODC-BY as well as with each other.

A detailed examination of each of such national open licences exceeds the scope of this study. Within the context of the EOSC, it would be important to provide relevant information on the compatibility of each of the national licences with the licences recommended by the EOSC. This would help to avoid inadvertent breaches of copyright, support harmonisation and reduce the overall number of recommended licences throughout the EU and beyond – see **Recommendations 12, 14 and 30.**

2.6.5. Licensing specifically related to databases

As previously discussed, databases may include both copyright and *sui generis* rights, both of which require a licence.

The available licences may cover: (a) the database *sui generis* right *and* the actual copyrightable content of the database; or (b) only the database *sui generis* right but not the actual copyrightable content of the database.

Therefore, the better options for licences for databases are those that cover both the *sui generis* right *and* the actual copyrightable content of the database.[123]

For example, the O-UDA covers both aspects (as per (a) above) while the database licence ODC-BY only covers the database *sui generis* right (as per (b) above). In the case of type (b) licences, any content included in the database, such as a picture, will not be covered by the licence for the database and the specific data (i.e., the picture) will have to be licensed separately, in addition to the database *sui generis* right licence – see **Recommendation 6**.

The differences among licences also illustrate the need for data providers and users to have a certain degree of knowledge and training on copyright issues and licences in order to be able to make informed decisions regarding what type of licence they should use and the purposes they serve.

[123] Note again that the copyrightable elements of the database may include both the content of the database and the structure of the database, to the extent it qualifies for an intellectual creation.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

OPERAS is the European Research Infrastructure for the development of open scholarly communication in the social sciences and humanities (SSH). OPERAS identified that authors, including SSH researchers, often shy away from tackling legal issues involved in scholarly publishing on their own and, to a great extent, do not regard legal matters as part of their work. Because of this, researchers tend to sign most publishing contracts (contracts offered to them by commercial publishers) without giving sufficient thought to copyright issues, even if they are aware of the dire consequences. Often, researchers are led to believe that open licences such as CC0 or CC BY are irrelevant or unsuitable for their work. OPERAS' White Paper on Advocacy (July 2018) concludes that overcoming the lack of information and knowledge is a very important step on the way to achieve Open Access publishing.

Recommendation 13: The EOSC should develop or encourage the development of a centralised source of knowledge, guidance and support researchers on copyright and licences related issues so that researchers can make informed decisions on what licence they should assign to their data.

2.6.6. What licence should be used?

As has been demonstrated above, the most suitable licence will depend on the type of copyrightable data and its nature (e.g., database protected by *sui generis,* software, non-software, metadata, etc.). A wide variety of licences are available under each category, depending on the objectives that the licensor is trying to achieve.

From a legal interoperability perspective, the CC No Rights Reserved (CC0) or the Public Domain Dedication and Licence (PDDL), or an equivalent statement of rights, should be applied to all **metadata** as part of the Rules of Participation in the EOSC – see **Recommendation 3**. This is crucial to ensure legal interoperability in the implementation of the FAIR Principles.

The CC0 is, in general, preferred over the CC BY, although from a potential liability perspective, there may be instances where a clear disclaimer and limitation of liability clause is needed, such as the one provided by the CC BY or the MIT License – see **Recommendation 9**.

When it comes to the **data** itself, the above licences are desirable; however, this may not always be feasible. As outlined elsewhere in this study, there could be a number of legal restrictions on the distribution of data, such as personal data protection, protection of endangered species, sensitive data, etc. There may also be other legitimate reasons such as patent strategies, or simply a researcher's own preference for downstream control over the data.

In cases where other less-restrictive means such as an embargo period or redaction of data are insufficient, a customised licence may need to be drafted. This is likely to be on account of the

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

legal interoperability, as such customised licences or contracts are less interoperable and may not be compatible with other licences.

With the above issues in mind, it is impossible to recommend a specific licence for all data made available through the EOSC. Doing so may result in excluding significant data and thereby making the EOSC itself irrelevant, or it will come on the account of significant legal interoperability constraints. This said, the EOSC should encourage the use of open and permissive licences to the greatest extent possible. The most non-restrictive way to share data is to provide the data under a waiver such as CC0. However, any licence which is open and permissive such as (CC BY, U-ODA, MIT or other national licences) would be preferable to a customised licence and more supportive of legal interoperability in general – see **Recommendation 12**.

Equally important is to understand which licences are compatible with others. OpenMinTeD has a useful matrix presenting the compatibility of different licences.[124] Given the wide variety of available licences, it is recommended that the EOSC should provide relevant information on all permissive licences including on their compatibility with each other and any applicable restrictions.[125]

Recommendation 14: A list of EOSC-recommended licences and their compatibility with other licences, as well as with Member States' recommended licences, should be provided to data producers, rights-holders and users, in order to avoid an inadvertent breach of copyright and with a view to harmonise and reduce the overall number of recommended licences.

2.7. Expired, unknown or changes to copyright

Copyright and *sui generis* rights are limited in time. The default rule in the EU is the life of the author plus 70 years, while *sui generis* rights last for 15 years. There are also other related rights (e.g. sound recording, film fixations, non-original photographs, etc.) which may have different periods.

From a user's perspective it may be difficult to verify whether a copyright exists, and if it does, whether the protection has expired. Similarly, some historic copyrightable datasets and metadata may have no licence or unclear licence arrangements assigned to them ('orphan data').[126] This may occur, for example, in situations where the author is unknown, or deceased, leaving no locatable heirs, or where the holder of the copyright was a legal person but it has ceased to exist with no legal successor, e.g., due to liquidation, or where any records about copyright ownership have been lost. Without permission, a waiver, or specific exemption, no

[124] See Open Mining Infrastructure for Text and Data, 'OpenMinTeD Compatibility Matrix', https://openminted.github.io/releases/license-matrix/, accessed 19 November 2020.
[125] For example, the European Data Portal has developed a Licensing Assistant where a short description of many of the open licenses is provided, see European Data Portal, 'Licensing Assistant', https://www.europeandataportal.eu/en/training/licensing-assistant, accessed 19 November 2020.
[126] We use the term 'orphan data' in contrast to 'orphan works' which is used in the Orphan Works Directive 2012/28/EU. We consider that the Orphan Works Directive does not address the issue from a FAIR perspective completely because of its scope (the definition of 'orphan works' is relatively narrow) and because it has been criticised on the basis that it presumes that the reuse of orphan works should be restricted.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

one may use the dataset or sublicence the dataset and consequently, orphan data is often left unused due to the impossibility or the disproportionate cost involved in tracing the copyright holder.[127]

Finally, it is possible that user rights, restrictions and conditions of use may change over time and rights statements made in the past may no longer reflect the current rights-holder(s) claims regarding the data.

In such cases there is a need to clarify the status of historic copyrightable data or orphan data which have no licence or unclear licence arrangements or where the copyright holder is unknown or is not reachable.[128] Moreover, there is a need to allow for user rights, restrictions and conditions of use to be updated from time to time and repositories of data must allow for an easy mechanism of doing so, including an audit trail for any licence changes.

Recommendation 15: Instances of expired or non-existent copyright, or where data is already in the public domain, should be clearly marked, for example, by way of the Creative Commons Public Domain Mark (CC PDM) or equivalent.

Recommendation 16: The EOSC should encourage repositories (for example, through the Rules of Participation) to incorporate harmonised mechanisms to validate and allow for the update of restrictions, rights statements and conditions of use on data as these may change over time. For this purpose, the principle that data licences can only become more permissive not more restrictive after first being shared within the EOSC should be considered.

Recommendation 17: The EOSC should adopt a uniform set of recommendations or guidance on how to handle copyrightable datasets where the rights-holder(s) are unknown or not reachable and the data has no licence assigned to it. An 'orphan data' standardised notice and related legal implications could be considered for such cases.

[127] "*The British Library estimates that 40% of works in their collections are orphan and over 1 million hours of TV programmes from BBC archives are not used due to the impossibility or the disproportionate cost to trace rightholders – and the risk of a subsequent legal action is simply too great for this material to be made available online*". Neelie Kroes, former Vice-President of the European Commission responsible for the Digital Agenda, addressing the challenge in the context of the Orphan Works Directive: https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_11_163.

[128] The Rights Statements Working Group of the International Rights Statement Working Group has developed a number of rights statements that could be used to indicate, for example, if the rights-holder either cannot be identified or cannot be located, see: Speech 11/163, 'Neelie Kroes Vice-President of the European Commission responsible for the Digital Agenda Addressing the orphan works challenge IFRRO (The International Federation of Reproduction Rights Organisations) launch of ARROW+ (Accessible Registries of Rights Information and Orphan Works towards Europeana)' (European Commission, 10 March 2011) https://rightsstatements.org/files/180531recommendations_for_standardized_international_rights_statements_v1.2.2.pdf, accessed 19 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

3. Other forms of intellectual property rights

3.1. Patents

A patent is a form of intellectual property that gives its owner an exclusive right to prevent others from possessing, using, selling, manufacturing and importing the patented invention or offering to do any of these things within a definite geographical area. Unlike copyright, patents only arise if the inventor applies for and is granted a patent by a public authority, for a limited period of time (usually 20 years with potential extensions).

In principle, any invention directed at the functional or technical aspects of products or processes, in all fields of technology, may be patented. To qualify for patent protection the invention must meet certain conditions of "patentability", which are: (1) patentable subject matter, (2) novelty, (3) inventive step (non-obviousness); and, (4) industrial applicability.

It is important to note that premature disclosure of a potential patentable invention (prior to filing of a patent claim) may impair the requirement for *novelty* and the ability to acquire a patent protection for the invention.[129]

Not all kinds of inventions can be patented. European patent law provides for a number of exclusions, exceptions and limitations to patent rights. The European Patent Convention[130] offers a list of "non-inventions", including abstract ideas, scientific theories, mathematical methods and computer programs, that are excluded from patent protection.[131] Most research data as such is unlikely to be patentable, although specific applications and combinations of software and functional data with a technical effect may form part of patent claims.[132] In practice, patent eligibility may become a question of claim-language, where mixed claims could with clever drafting pass the patent eligibility threshold.

As explained elsewhere,[133] there are basically two cases in which disclosure of data might impact pending or future patent applications: (1) the data disclosure describes the invention and impairs the *novelty* requirement;[134] or, (2) the data disclosure fills a gap in other researchers' knowledge so that inventions that arise from the research are not described by the data, but are rendered "obvious" to a person skilled in the art by the disclosure. Once the

[129] With the application for a patent, the applicant must disclose the invention in a manner sufficiently clear and complete for the invention to be carried out by a person skilled in the art. This is the *quid pro quo* principle of patent protection, which distinguishes patents from e.g. trade secrets.
[130] Art. 52 (2) of the European Patent Convention, see European Patent Convention (European Patent Office, 16th edition, June 2016) https://www.epo.org/law-practice/legal-texts/epc.html, accessed 19 November 2020.
[131] The following in particular are not to be regarded as inventions: Discoveries, scientific theories and mathematical methods, aesthetic creations, schemes, rules and methods for performing mental acts, playing games or doing business, programs for computers, presentations of information.
[132] See Guidelines for Examination at the European Patent Office on "Programs for computers" (November 2019): https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_3_6.htm, and on "Data retrieval, formats and structures": https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_3_6_3.htm, both accessed 19 November 2020.
[133] See Carroll MW (2015) Sharing Research Data and Intellectual Property Law: A Primer. PLoS Biol 13(8): e1002235. doi:10.1371/journal.pbio.1002235.
[134] Subject to applicable grace period – a period of time preceding the filing of a patent application, during which certain types of disclosures do not destroy its novelty. In the EU, the grace period (granted under certain circumstances) is usually 6 or 12 months.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

invention becomes "obvious", it lacks the required *inventive step* needed to obtain patent protection.

In such cases, data generators may want to keep certain data secret, at least for a certain period of time (e.g., by way of embargo), or redact part of the data until a patent claim is filed or they may avoid disclosing the data altogether, due to strategic considerations relating to patent applications. This may particularly be the case if researchers are required to assign their rights to inventions to funders arising under a sponsored research agreement or due to other conditions of a grant from funding bodies that includes specific obligations regarding ownership of IP and technology transfer. Similarly, research organisations may be under an obligation to maximise the economic benefits of their research and seek to delay making research data open for an appropriate period.

Recommendation 18: In accordance with the principle of "as open as possible, as closed as necessary", EOSC policies should take into account commercial incentives and facilitate the seeking of IP protection in justified cases where the disclosure of the data may compromise the ability, for example, to file for patents or protect trade secrets.

Furthermore, in cases where patentable data such as software or functional data had been disclosed, there is a risk that specific reuses of the data may infringe patent claims if no appropriate licence has been granted. In such cases, it will be important that the metadata indicates reusability restrictions due to pending or exiting patent claims as well as information about the relevant patent proprietors, licensing options and any other relevant information.

Recommendation 19: Metadata should indicate reusability restrictions on software or data due to pending or existing patent claims or when data had been redacted due to commercially confidential information. The metadata should also provide information about the relevant patent proprietors, licensing options and any relevant additional information.

To the greatest extent possible, data users should still be able to access protected data, through the implementation of proper authorisation procedure and safeguards, utilising smart contract solutions, in order to guarantee the confidentiality of the data and compliance with terms of use and applicable restrictions.

Recommendation 20: The EOSC should facilitate easy and intelligible platform solutions, e.g. through smart contracts, that allow scientists and their institutions to acquire licences where necessary, or to achieve "freedom to operate" confirmations where research exemptions, experimental use exemptions and/or patent pledges apply.

3.2. Trade Secrets

Trade secrets offer protection for know-how and business information that cannot be protected by conventional Intellectual Property Rights (IPR) or that is otherwise elected not to be protected under conventional IPR. This could, for instance, be an invention that does not fulfil the requirements to obtain patent protection, such as the eligibility, *novelty*, or *inventive step*

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

criteria. Trade secrets also protect know-how or information that is not described in a patent application but that remains secret and is crucial to optimise the use of an invention.

Data may be treated as a trade secret if it derives economic value from not being generally known or readily ascertainable, and if the data has been subject to "reasonable measures" to keep it secret. The Trade Secrets Directive[135] defines trade secrets as any information that: (i) is secret, i.e., not generally known or readily accessible to people in a wider community than those who typically deal with that information, (ii) has an actual or potential commercial value because it is secret, and (iii) has been subject to reasonable steps under the circumstances to keep it secret.[136] While the Directive's broad definition does not specify exactly what category or content of information or data qualifies as a trade secret and the details of what exactly constitutes such "reasonable measures" remain debatable, most research data already meets this definition in the early stages of collection or generation. Accordingly, many scientific researchers hold trade secrets in their research data and follow-on data for some period of time, even if they are unaware of this fact.[137]

The organisation and ways of operating a database could also be considered trade secrets, since this would have an important impact on the quality of science-related data and the services provided. Trade secrets could, for instance, relate to the (systematic) approach chosen to collect, store, label, process and track data or to the algorithm used to analyse the collected data. Researchers that, for example, access collections of data, could be required to contractually commit themselves to respect the confidentiality of trade secrets. A third party would not only have to copy the database content or structure or software, but would also need access to certain confidential know-how or information to establish and operate the database or software in an optimal manner. In this respect, trade secrets offer an additional protection of databases or software.

Researchers and scientists may use various techniques to keep certain data secret due to the development of IP protected innovative products and processes, technology transfer, or with a view to protect data and know-how as trade secrets where patents are not available.[138] The Trade Secret Directive requires that the trade secret has been "subject to reasonable steps under the circumstances to keep it secret". In the absence of a trusted mechanism to enable the shielding of specific data, and integrating different levels of security, there is a risk that complete sets of data will not be shared at all due to the fear of leakage of trade secrets or that researchers feel that not all reasonable steps have been taken to protect the data.

Recommendation 21: The EOSC should create or encourage the creation of a trusted environment with reliable access control and authorisation procedures which will accommodate different techniques for shielding data in order to enable researchers to share data while providing them the possibility to protect IP and trade secrets related to innovative products and processes.

[135] Directive 2016/943 of the European Parliament and the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] O.J. L 157.
[136] Art. 2 Directive 2016/943.
[137] See supra note 133.
[138] Indeed, researchers may also be prevented from disclosing certain data due to funding requirements, organisational policies or other related reasons.

3.3. Regulatory (data) exclusivities: a sector-specific reuse limitation

Another form of protection that has become increasingly important to the commercial aspects of research data in the health and life sciences is regulatory market and data exclusivities, which are available for the protection of clinical trial data. The European rules governing regulatory exclusivity[139] introduced the "8 + 2 + 1" year rule.[140] Under this rule, which applies to both small-molecule drugs and biologics, data exclusivity applies during the first eight years from the grant of the innovator company's marketing authorisation. This implies that if a producer of a generic or biosimilar intends to apply for marketing authorisation of a generic drug or biosimilar and wants to cross-reference to existing preclinical or clinical data, the authorities may during this period not accept such references due to data exclusivity restrictions.

Regulatory data exclusivity acts as a barrier that prevents a specific reuse of the existing clinical trials data required for marketing authorisation of a similar medical products. In that sense, regulatory data exclusivity operates in much the same way as an IPR (e.g. patents) during the exclusivity period. To the extent that clinical trials data is made available through the EOSC, data users, such as scientists involved in publicly funded clinical trials, must be made aware of possible data exclusivity-based reuse restrictions which may prevent them from using the data, unless a specific licence is given – see **Recommendations 19 and 20**.

3.4. Other rights

As mentioned, there are many other forms of protection that may become relevant in the context of making data available through the EOSC, and which need to be considered before reusing the data. Another example is the protection granted in Germany to photographs (*Lichtbild*) under a copyright-related right (or a *neighbouring right*), which is different from "photographic works" that are protected by the German Copyright Act. A specific section in the German law affords separate protection to photographs and it does not require creativity on the photographer's part, only a minimum level of "personal intellectual effort." The scope of protection afforded however, is very similar to copyright, but, as explained, it is based on a different section of the applicable law and may be extended, depending on the timing and intensity of the publication of the photograph.[141]

Common to all such additional rights is that users must be made aware of them and any associated reuse restrictions which may prevent them from using the data, unless a specific licence is given – see **Recommendations 19 and 20**.

[139] Directive 2001/83 of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use [2001] O.J. L 311.
[140] Regulation 726/2004, of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency [2004] O.J. L 136.
[141] See decision of the German Federal Court of Justice ('der Bundesgerichtshof') in case I ZR 104/17, 20 December 2018.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

4. Privacy and data protection

4.1. Background

This section examines the EU General Data Protection Regulation (GDPR)[142] which introduces a number of constraints and stricter protocols on the processing of EU residents' personal data. The emerging challenge is to devise ways of integrating GDPR compliance into EOSC operations (including products – hardware and software – and service design) without compromising legal interoperability.

The benefits of being able to make use of big and open data in scientific research and other areas are acknowledged by the GDPR. This acknowledgement comes with certain rules for the use and reuse of these data owing to increasing threats such as cyberattacks and data leaks. Data protection and security matters are at present under intense scrutiny and the design of a better legal interoperability framework is a significant part of discussion in legal spaces.[143]

4.2. Personal data

Personal data is defined by the GDPR as "any information relating to an identified or identifiable natural person." 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".[144]

In the context of the GDPR, the typical actors are: i) the "data subject" – the "identified or identifiable natural person" about whom the data in question relates;[145] ii) the "data controller" – the entity who "defines the means and purposes of the processing";[146] and, iii) the "data processor" – the entity which process personal data on behalf of the data controller if the controller did not process personal data directly themselves but outsourced the task.[147]

The GDPR restricts the processing of personal data. "Processing" is broadly defined to mean "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".[148] "Sensitive personal data"[149] is subject to stricter

[142] 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 [2016] OJ L 119 ("General Data Protection Regulation").
[143] European Data Protection Supervisor, Interoperability, https://edps.europa.eu/data-protection/our-work/subjects/interoperability_en, accessed 19 November 2020; Dan Lohrmann, Why You Need the Cybersecurity Framework (*Government Technology*, 20 May 2018), https://www.govtech.com/blogs/lohrmann-on-cybersecurity/why-you-need-the-cybersecurity-framework.html, accessed 10 November 2020.
[144] Art. 4(1) GDPR.
[145] Art. 4(1) GDPR.
[146] Art. 4(7) GDPR.
[147] Art. 4(8) GDPR.
[148] Art. 4(2) GDPR.
[149] "Sensitive personal data" is any data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human

processing conditions and requires explicit consent from the data subjects.[150]

As discussed in chapter I, the FAIR Principles emphasise accessibility, reusability and "machine-actionability,"[151] i.e., the capacity of computational systems to "find, access, interoperate, and reuse data with none or minimal intervention."[152] Within the context of the EOSC, data will be uploaded to repositories, stored in a data archive and made available to users for further reuse. If such data includes personal data, this will qualify as "processing" – thus, from a GDPR perspective, it represents certain legal risks, which are increased by the volatile nature of cloud federations and the additional developments with regard to the subjects of big data analytics, data mining and AI.

Therefore, to the extent that GDPR requirements are applicable to research data, they become potential barriers to legal interoperability as they may lead to: (1) legal impediments to making certain data reusable, Open and/or FAIR; or, (2) EOSC participants (such as data producers, service providers and users) inadvertently breaching the GDPR.

4.3. Identifying personal data

As mentioned, the definition of personal data is very broad and includes items such as names, surnames, telephone numbers, Internet Service Provider (ISP) addresses, fingerprints, etc. In the context of EOSC, instances of personal data can be found in all levels where data is made available including research data, metadata and Persistent Identifiers (PIDs).[153]

In practice, identifying whether personal data is included in research data, metadata or PIDs is not always straightforward and requires knowledge and a case-by-case assessment. It is more likely that in specific disciplines (such as social sciences, biotechnology and medicine) research data will almost inevitably include information that qualifies as personal data, but other disciplines are not excluded. In such cases it will be the data generator who is best placed to identify when research data contains personal data and to take the necessary steps to ensure compliance with the GDPR or take the necessary measure as required by the GDPR.

Recommendation 22: Encourage the development of basic guidelines on GDPR issues for researchers on identifying personal data and on implementing the "Privacy by Design and by Default" approach.

4.4. Consent

The GDPR prohibits the processing of personal data unless a valid legal basis for the processing exists. One such legal basis, which is frequently used, is the consent of the data subject. For GDPR purposes, consent should be distinguished from other consent requirements that fulfil ethical standards or procedural obligations. The GDPR requires that consent is "given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication

being; health-related data; data concerning a person's sex life or sexual orientation. See Art. 4(13)-(15); Art. 9 and Recitals 51-56 GDPR.

[150] Art. 9 (1) (2) GDPR.

[151] Go FAIR, FAIR Principles, https://www.go-fair.org/fair-principles/, accessed 12 November 2020.

[152] Ibid.

[153] See, for example, para. 1.3, supra note 24.

of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement".[154]

However, obtaining informed consent for each and every dataset is not always practical, especially in cases of the secondary use of the data, due to the impossibility or the disproportionate cost to trace each data subject individually. Moreover, even in cases where the data subject can easily be traced, there will be cases when it would not be appropriate or even legal to base processing on consent, due to questions such as whether the consent is given freely.[155] It is therefore important to base the personal data processing on one of the other legal bases.

4.5. Other legal bases

The GDPR also allows the processing of personal data relying on a number of other valid legal bases or in relation to special categories of personal data. In addition to consent, these generally include legal bases relating to: performance of a contract; legitimate interest; vital interest; legal requirement; and, public interest.

One example is the special category of "public interest in the area of public health"[156] which is of particular interest in light of the current COVID-19 pandemic.

> The COVID-19 pandemic requires collaborative and international research efforts. A significant source of information for researchers is the large amount of digital health data that is continuously being collected in cloud-based databases. The GDPR explicitly allows the processing of sensitive personal data for scientific research purposes if it is "necessary for reasons of public interest in the area of public health." The European Data Protection Board (EDPB) has recently released its "Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of COVID-19 outbreak" and supports research and data sharing under the appropriate legal framework. The EDPB recommends, inter alia, the use of pseudonymisation techniques to reduce the risk of deidentification or tracking of individual persons. Even in these cases, the EDPB states that any data processing has to be transparent and that the data should be processed with sufficient privacy safeguards in place and not shared with third parties without authorisation.

Another example is in relation to "scientific research purposes". In this context, it is not always possible to fully identify the purpose of personal data processing at the time of collection. Recital 33 of the GDPR[157] provides a degree of flexibility in terms of specification and

[154] Recital 32 and Art. 4(11) GDPR.
[155] See, for example, the discussion in section 3 ("elements of valid consent") in the European Data Protection Board, *Guidelines 05/2020 on consent under Regulation 2016/679* (2020) https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf, accessed 5 November 2020.
[156] Art. 9(2)(i) GDPR.
[157] Recital 33 GDPR reads: "It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for

granularity of consent required, and some secondary use of the data could fall within the category of scientific research purposes. However, in practice, open access and open data cannot be generally justified on the basis of scientific research purposes, as making data *generally* available for research purposes through the EOSC is not connected to a *specific* research purpose. Additional measures and safeguards are required[158] in particular in relation to the principles of data minimisation and purpose limitation.

Common to the use of other legal bases for the processing of special categories of personal data is that, in the absence of the individual's consent or anonymisation, potential problems may still arise when relying on such grounds for reuse of research (containing personal) data,[159] due to the various principles and requirements that the GDPR introduces. These principles and requirements are summarised below.

4.6. Principles and requirements

There a number of important principles and requirements set out in the GDPR in relation to data processing which could be summarised as follows:

Lawfulness, Fairness and Transparency: Data "shall be processed lawfully, fairly and in a transparent manner in relation to the data subject".[160] This principle requires that "any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualization be used".[161]

***Purpose limitation*:** As a general rule, data should be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes" however "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes."[162]

***Data minimisation*:** Data processing should only use as much data as is necessary to achieve a certain task. Data collected for one specific purpose cannot be repurposed without further consent.[163]

---

scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose."

[158] See European Commission Art. 29 Working Party, *Guidelines on consent under Regulation 2016/679* (2017) https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051, accessed 19 November 2020. In addition, the controller may apply further safeguards in such cases. Art. 89(1), for example, highlights the need for safeguards in data processing activities for scientific or historical or statistical purposes. These purposes "shall be subject to appropriate safeguards, in accordance with this regulation, for the rights and freedoms of data subject." Data minimisation, anonymisation and data security are mentioned as possible safeguards as described further in this study below.

[159] See also RDA COVID-19 Working Group. Recommendations and Guidelines on data sharing. *Research Data Alliance, 2020*. DOI: https://doi.org/10.15497/rda00052.

[160] See Art. 5, 6 and 9 GDPR.

[161] Recital 58 GDPR. According to Art. 29 Data Protection Working Party, 2018, 5 "the concept of transparency in the GDPR is user-centric rather than legalistic." This highlights the central role of the comprehensibility and presentation of the information.

[162] Art. 5(1)(b); 6 and 26 GDPR.

[163] Art. 5(1)(b) GDPR.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Accuracy: Data should be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay".[164]

***Storage limitation*:** Data storage should be proportionate to the length and the purpose of the data collection. Longer periods are permitted "insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes."[165] In addition, appropriate technical and organisational measures to safeguard data subjects' rights and freedoms are required.[166]

Integrity and Confidentiality: Data shall be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures."[167]

***Data Portability*:** Data subjects have the right to obtain data that had previously been submitted to the data controller. This data must be delivered to the data subject in a "structured, commonly used and machine-readable format," and, if required, to send those data to another controller;[168]

***Right to be Forgotten (right to erasure)*:** Data subjects are able to request that a data controller delete their personal data and to cease from activities involving the processing or distribution of that data to third parties;[169]

***Breach Notification*:** The GDPR requires disclosure if there has been a data breach. Specifically, data controllers are required to notify a supervisory authority of data breaches within 72 hours. Data processors are required to disclose breaches to the controllers;[170]

***Privacy by Design and by Default*:** Embedding of both privacy and data protection requirements in the design of information technology products and systems. The controller must implement the appropriate technical and organisational measures – such as pseudonymisation – to ensure that only personal data which is necessary for each specific purpose of the processing are processed, e.g. the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility.[171]

***International Data Transfers*:** Cross-border transfers of EU nationals' personal data to third parties outside of the EU requires additional safeguards[172] – see further below.

[164] Art. 5(1)(d) and 16 GDPR.
[165] Art. 5(1)(e) GDPR.
[166] Art. 5(1)(e); 89(1) GDPR.
[167] Art. 5, 24 and 32 GDPR.
[168] Art. 20 GDPR; see also supra note 158.
[169] Art. 17 GDPR.
[170] Art. 33 GDPR.
[171] Art. 25(1) GDPR.
[172] Art. 46 GDPR.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

***Extra-Territorial Scope*:** The GDPR broadens the territorial applicability of privacy protections on data controllers and processors who may be based within the EU as well as outside of it;[173]

***Data Privacy Impact Assessment (DPIA)*:** A DPIA is required when the data controller begins to process personal data in a way that is likely to involve a "high risk" (e.g., where special categories of sensitive data are processed);[174]

***Data security*:** Data controllers and data processors are required to process personal data securely by means of "appropriate technical and organisational measures."[175]

Personal sensitive data: Extreme caution and necessary measures must be taken to protect personal data, in particular personal sensitive data.[176]

Recommendation 23: The EOSC Rules of Participation (RoP) should include a requirement for users, repositories, data and service providers (and any other data controllers and data processors) to implement appropriate measures to ensure compliance with the GDPR, for example, by requiring repositories to indicate the conformity of their data and services with the GDPR.

4.7. Anonymisation and pseudonymisation

As mentioned, there are certain measures that may be taken in order to ensure compliance with the GDPR. Both anonymisation and pseudonymisation are useful methods that service providers and EOSC operators (data controllers and processors) may use, for example, to reduce the likelihood that data breaches result in the leakage of private information and subsequent penalties. In order to adhere to GDPR obligations, anonymisation and pseudonymisation are recommended. While both techniques allow for personal data to be securely processed, within the scope of the GDPR these two methods are substantially dissimilar. The key difference is with regard to whether data subjects are re-identifiable.[177]

When data is **anonymised**, the link between personally identifiable information and the aggregate dataset is removed.[178] This aspect makes it impossible to re-identify a data subject. Data anonymisation is the most appropriate approach when protecting personal data since it can be shared for secondary purposes – such as scientific research – without placing individual

[173] Art. 3, 4(7) and 4(8) GDPR.
[174] The GDPR describes the requirement of DPIA as follows: "Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks." See Art. 35(1) GDPR.
[175] Recital 78 and Art. 32 GDPR.
[176] Art. 9 GDPR.
[177] Robert Walters, Leon Trakman and Bruno Zeller, *Data Protection Law: A Comparative Analysis of Asia-Pacific and European Approaches* (Springer, 2019), p.90.
[178] A number of tools are available to help anonymise personal data. See, for example: Amnesia, https://amnesia.openaire.eu/, accessed 30 October 2020.

privacy at risk. The GDPR does not apply to anonymised data. Specifically, it clarifies that when data is not related to an identified or identifiable natural person the processing and storage of anonymised data fall outside the GDPR's applicability.[179]

A conclusion by the European Commission's Article 29 Working Party mentioned that in order to be able to meet the current standards for anonymisation, the natural person linked to the data cannot be identified by using "all means likely reasonably to be used" whether by the data controller or a third party. In effect, the process by which the anonymisation occurs must be clearly laid out from the beginning in order to reach the level of anonymisation required.[180]

It is essential within the anonymisation process that data controllers must make use of a mechanism which renders re-identification impossible. This typically involves stripping the data of enough characteristics to cripple any potential process to identify the individual that the data represents. While there may not be a singular approach that is usable for all situations, the Working Party advises that the anonymisation choice should be made after assessing each situation individually and by utilising any number of techniques that may be applicable.

A major drawback of anonymisation[181] is that rendering data anonymous decreases its quality. For example, voice recordings may lose their sound, thus rendering it non-usable in the context of language research. Anonymisation also decreases its value within the context of data analytics and data mining tools. Techniques which are employed to mask data also have the effect of distorting its quality for use in analytical procedures. The main goal of data processing is often to parse the data in the search for patterns within the aggregate data. Data and metadata on aggregate can, for example, aid a healthcare institution in planning for a disease outbreak or devise improved treatment options.[182]

Pseudonymisation by comparison is a de-identification technique whereby personal data is substituted with artificial identifiers rendering data linkage to an individual very difficult or impossible. The GDPR recommends that those who either control or process personal data should implement pseudonymisation as a general good practice. It defines pseudonymisation as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".[183] Pseudonymisation is regarded as having a number of strengths enabling sensitive data to be protected but still manageable.[184]

---

[179] Recital 26 GDPR.
[180] Namely, randomisation and generalisation. In particular, the opinion examines noise addition, permutation, differential privacy, aggregation, k-anonymity, l-diversity and t-closeness. See Article 29 Data Protection Working Party, *Opinion 05/2014 on Anonymisation Techniques* (10 April 2014), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, accessed 24 August 2020.
[181] Interviews with the Consortium of European Social Science Data Archive (CESSDA) and the European Research Infrastructure for Language Resources and Technology (CLARIN).
[182] Marcelo Corrales Compagnucci et al., Homomorphic Encryption: The 'Holy Grail' for Big Data Analytics & Legal Compliance in the Pharmaceutical and Healthcare Sector? (2019) 3 *European Pharmaceutical Law Review* 144. DOI: https://doi.org/10.21552/eplr/2019/4/5.
[183] Art. 40(2)(d) GDPR.
[184] Heidelinde Hobel et al., 'Anonymity and Pseudonymity in Data-Driven Science' in John Wang (ed) *Encyclopedia of Business Analytics and Optimization* (IGI Global, 2014), p.128.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Although effective, pseudonymised data is still considered to be personal data that must be protected and as a result lies under the purview of the GDPR. One key concern with pseudonymisation is that in order to make use of the underlying data it must be decrypted. The decryption process is what then makes the data potentially vulnerable to cyberattacks. A malicious third party is potentially able to expose and intercept the data during this decryption process.[185]

To summarise, anonymising or pseudonymising is achieved by employing the use of different methods and procedures. These approaches regularly involve robust algorithms to provide a high level of security. Nonetheless, they have the undesirable effect of lowering the quality of data for subsequent analysis, or they reveal the data in the process of decrypting it for analysis, rendering it vulnerable.[186]

Recommendation 24: Where possible, data anonymisation is an appropriate approach to protect personal data. When done correctly, anonymised data can be shared for secondary purposes – such as further scientific research – without placing individual privacy at risk.

4.8. Cross-border data transfers

Free flow of personal data to and from countries outside of the EU is essential for the expansion of international cooperation and research, which is crucial for some EOSC stakeholders.[187] Data transactions which involve transfers to third countries have to satisfy the data protection standards outlined in the GDPR. Ensuring an adequate standard of protection and a transparent legal framework for international data transfers is therefore fundamental.

The EU Commission distinguishes between countries with adequate levels of protection and third countries that do not provide sufficient data protection. Third countries deemed to have adequate protections are those for which the EU Commission has confirmed, by way of an adequacy decision,[188] a level of privacy "essentially equivalent" to that of the GDPR.

Transfers of personal data to a third country with adequate level of protection can be conducted without the need to obtain any further authorisation. However, transfers to third countries that do not have an adequate level of protection require the use of additional safeguards, such as EU approved standard contractual clauses (SCCs) and binding corporate rules (BCRs).

[185] Steve Touw, 'Homomorphic Encryption Alone is Security, Not Privacy' (Immuta, 14 September 2018), https://www.immuta.com/homomorphic-encryption-alone-is-security-not-privacy/, accessed 10 July 2020.
[186] Homomorphic Encryption (HE) can be a useful tool in the context of data analytics. The HE method of analysing the data does not expose any private information since it is still encrypted at rest and while in transit. This brings an additional layer of protection, granting possibilities for using the data for scientific research and for secondary use. See supra note 182.
[187] Interview with the Consortium of European Social Science Data Archive (CESSDA).
[188] The European Commission has recognised that the following countries provide an adequate level of protection: Andorra, Argentina, Canada (commercial organisations), Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland, Uruguay and the United States of America (limited to the EU-US Privacy Shield framework which was recently invalidated by the CJEU on 16 July 2020. See European Commission, Adequacy Decisions, https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en, accessed 24 August 2020. Adequacy discussions are ongoing with South Korea.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Data "exporters" in collaboration with data "importers" have the duty to assess the law and practice of the country to which data will be transferred. The EDPB recently released its recommendations on the measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data.[189]

Stricter rules resulting from the GDPR mean that transferring the personal data of EU nationals to third parties outside of the EU requires additional safeguards. Changing practices regarding the implementation of these rules, such as the recent invalidation of the EU-US Privacy Shield Framework by the CJEU[190] have exacerbated this problem.

Recommendation 25: The EOSC should address the issue of cross-border data transfer (outside the EU/EEA) and require the implementation of additional security and organisational safeguards in the SCCs where international data transfers take place. Where possible, this should be done in a machine-actionable manner such as in a Service Level Agreement (SLA) and the data management system.

[189] Recommendation 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data (10 November 2020), available at: https://edpb.europa.eu/sites/edpb/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf accessed 19 November 2020.

[190] Case C-311/18, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems* ECLI:EU:C:2020:559 [2020] ("Schrems II case").

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

5. Other restrictions and legitimate reasons

In the previous chapters we have reviewed copyright (including *sui generis*), the General Data Protection Regulation (GDPR), and other forms of intellectual property rights that must be taken into consideration when making research data reusable and the impact of such legal instruments on legal interoperability. There are, however, other legal constraints, conventions or legitimate reasons that restrict the disclosure of, access to, or reuse of data. These may be in connection to, for example, the protection of endangered species, traditional cultural resources, national security, strategic resources, sovereign genetic resources, traditional knowledge, etc.

While Open Data Principles, in parallel to the FAIR Principles, are required to maximise legal interoperability, they also need to be balanced against such legal restrictions and legitimate interests. Measures used to restrict access to sensitive data include the generalisation of data, redaction of specific information (such as location of an endangered species), embargo periods, authorisation procedures, specific contractual arrangements, etc. In order to remain FAIR, any restrictions or conditions applicable, must be set out clearly in the metadata, which should be made available in a human and machine-readable manner – see **Recommendation 4**.

In line with the principle of "as open as possible as closed as necessary", there is a need to ensure a correct balance between legal interoperability and the shielding of data, so that restrictions on access and reuse are only imposed if they have a legitimate basis, are proportionate, and do not go beyond what is necessary and required. When data is protected due to law and regulations, the position is quite clear, but this is not the case when data is or should be protected due to ethical or legitimate grounds, which are not always clearly defined and agreed upon.

Recommendation 26: Restrictions on the access and reuse of data should be proportionate and applied in cases of applicable legislation or legitimate reasons. The EOSC should consider preparing, in consultation with stakeholders, guidance on what are considered to be 'legitimate reasons' that go beyond existing legislation and which could justify the introduction of additional restrictions on access to and reuse of data. The guide could also provide a proposed list of such legitimate reasons.

In addition, necessary measures that are taken to protect data, such as an embargo period, generalisation of data or the redaction of specific information, need to be effective and proportionate. On the one hand, removing complete sets of data due to a remote risk connected with patent strategies will have to be tested against the principle of proportionality. On the other hand, when redacted or generalised data is combined with other sets of data and carried forward, there is a risk that sensitive information which was initially protected (for example regarding the location of endangered species), may nevertheless be deducted by way of analogy, if it is combined with other sets of related data.

Finally, as further demonstrated below, specific laws that restrict access to sensitive data may only be applicable in one jurisdiction but not in others. Equally, certain data, such as traditional knowledge, may be afforded protection by applicable intellectual property law in one jurisdiction but will not be afforded similar protection under intellectual property law, or indeed any other law, in other jurisdictions.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

The following sections demonstrate how such considerations may come into play by reference to a couple of legal instruments and related legitimate considerations.

5.1. Traditional knowledge, traditional cultural expression and sovereign genetic resources

There is no universally agreed definition of Traditional Knowledge (TK), Traditional Cultural Expressions (TCE) and Genetic Resources (GR). Instead, a number of definitions are provided in the various international conventions and agreements as well as in the national and local legislations. In general, these terms can be understood as follows:

TK is technical know-how, practices, skills, and innovations developed by a traditional regional, indigenous or local community.[191] As such, it is embedded in the knowledge system of a traditional community and can have a specific meaning in its culture. For instance, some forms of TK can be considered sacred and can be used only by certain members of a traditional community.

TCE are cultural manifestations such as music, art, designs, symbols and performances. The nature of TCEs primarily subjects them to copyright, trademark, certification of marks, industrial design and appellations of origin.[192]

GR are genetic materials of actual or potential value found in plants, animals and micro-organisms. GR are subject to access and benefit-sharing regulations, in particular within the international frameworks defined by the Convention on Biological Diversity and its Nagoya Protocol, as well as by the International Treaty on Genetic Resources for Food and Agriculture of the United Nations Food and Agriculture Organization.[193]

GR are encountered in nature and, since they are not created by humans, they cannot be directly protected by intellectual property law. However, inventions based on or developed using GR may be patentable or protected by plant breeders' rights. At the European level, this means that the access to and use of genetic resources will also be regulated by the Directive on the legal protection of biotechnological inventions[194] and Regulation 511/2014 implementing the Nagoya Protocol.[195] These create an obligation for users of genetic resources data to "exercise due diligence to ascertain that genetic resources and traditional knowledge associated with genetic resources which they utilise have been accessed in accordance with applicable access and benefit-sharing legislation or regulatory requirements, and that benefits are fairly and equitably shared upon mutually agreed terms, in accordance with any applicable legislation or regulatory requirements".[196] This will, for example, require users to collect the previous

[191] World Intellectual Property Organization, Intellectual Property and Genetic Resources, Traditional Knowledge and Traditional Cultural Expressions (WIPO 2015) p. 13.
[192] Ibid., p. 15.
[193] Ibid., p. 18.
[194] Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the Legal Protection of biotechnological inventions [1998] OJ L 213.
[195] Regulation (EU) No 511/2014 of the European Parliament and of the Council of 16 April 2014 on compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization in the Union [2014] OJ L 150.
[196] Art. 4 Regulation (EU) No 511/2014.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

informed consent given by the traditional community and respect the mutually agreed terms when reutilising the data.

TK, TCE and GR may be protected for various reasons such as "enabling indigenous and local communities as well as governments to have a say over the use of their traditional knowledge by others",[197] preventing the misappropriation of elements of the community's identity, and enabling communities to control and benefit collectively from their commercial exploitation.

TK, TCE and, to a lesser extent, GR, are not afforded the same level of protection globally and in some jurisdictions, they are not protected at all.

Some resources are protected by communities' customary rules which may not be recognised and enforced by the country in which the community lives and therefore may not be considered as IP or otherwise protected material (which means that restrictions on reusability cannot be enforced by the modern IP system or by law).

Some countries do protect such resources under their IP system, either by qualifying them as a special category of protected subject matter of an IP system[198] or without qualifying them as a special category of protected subject matter of IP.

Some countries protect them as a special category of rights outside their IP system.[199]

The place of TK and TCE, in particular within conventional intellectual property structures, is therefore very uncertain. There is also a great deal of variability between jurisdictions in the way that TK, TCE and GR are being protected, if at all. When protected, it may be by the modern IP model or by a traditional customary model, but the two systems are not necessarily compatible with each other.

Some of these problems have been addressed by the development and use of traditional knowledge licences (some rights reserved models), labels and notices, for example by the organisation Local Contexts[200] to promote legal and respectful use of data.

5.2. Endangered species

Each species is part of an ecosystem and contributes to its functioning in some ways. The extinction of a species has an impact on the natural balance and functioning of the ecosystems

[197] World Intellectual Property Organization, 'Traditional Knowledge and Intellectual Property – Background Brief', https://www.wipo.int/pressroom/en/briefs/tk_ip.html, accessed 19 November 2020.
[198] An example is Cameroon where "folklore" is expressly protected by copyright or neighbouring right. Law No 2000/011 of December 19, 2000 on Copyright and Neighboring Rights Law, s 5 (Cameroon) at WIPO, Collection of Laws for Electronic Access - Cameroon, https://www.wipo.int/edocs/lexdocs/laws/en/cm/cm001en.pdf, accessed 19 November 2020.
[199] This is for instance the case in Guatemala which protects tradition and custom in its Law for the Protection of the Nation's Cultural Heritage. See Decreto Numero 26-97 y sus reformas, Ley para la Protección del Patrimonio Cultural de la Nación (Guatemala), https://wipolex.wipo.int/en/text/235791, accessed 28 August 2020.
[200] Local Contexts, https://localcontexts.org/, accessed 28 August 2020. See also, Intellectual Property Issues in Cultural Heritage, 'Traditional Knowledge Licensing and Labeling Website 1.0', https://www.sfu.ca/ipinch/project-components/community-based-initiatives/special-initiative-traditional-knowledge-licensing-an/, accessed 28 August 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

it lives in. This impact can have limited consequences as well as devastating ones when it leads for example to the extinction of other species.

In order to avoid the disruption of ecosystems and thus the potential for unpredictable consequences, species at risk of extinction ("endangered species") benefit from special protections laid down in international and national law. By way of example, the following legal acts are relevant to the protection of endangered species:[201]

- International conventions related to natural science, e.g., Convention on Biological Diversity (CBD), Convention on International Trade in Endangered Species of Wild Fauna and Flora, Convention on the Conservation of Migratory Species of Wild Animals, etc.;
- Nagoya protocol on access and benefit sharing to CBD;
- The Open Data Directive;
- The INSPIRE Directive;
- The Habitats Directive (92/43/EEC), Birds Directive (amended, 2009/147/EEC) and protected areas (Natura 2000).

The legal mechanisms used to ensure the protection of species and their related data consist of various measures such as collection permits, the signing of mutually agreed terms and the generalisation or redaction of location data. If location data is released by way of open access, or even if released under controlled access but without appropriate safeguards, there is a tangible risk that such location data will be used for illegal and destructive purposes such as poaching leading to the weakening of the species.

Equally, if redacted or generalised data is combined with other sets of data and carried forward, there is a risk that sensitive information may be deducted by way of analogy. Rights-holders of sensitive data related to endangered species may be reluctant to share their data if they do not have sufficient safeguards that the data will be sufficiently protected in situations where the combination of datasets can generate additional information which would otherwise be considered as sensitive in one country but not in another country.

Mechanisms are therefore required to ensure that such new data which is produced from the combination of a number of separate datasets and carried forward will not inadvertently generate information which is considered to be sensitive under the terms of use, of one or more of the parent datasets. More generally, it is important that in such cases of deliberate or inadvertent breach of legal restrictions, a procedure for monitoring or reporting violations of use conditions is in place and corrective measures can be taken.

Recommendation 27: Adopt procedure for monitoring or reporting violations of use conditions and leakage of sensitive data.

[201] Hannu Saarenmaa et al., 'Open access implementation guidelines for DiSSCo. Deliverable D6.5.' (2019) https://zenodo.org/record/3465285#.X5a8xVko90s, accessed 28 August 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

6. Private law considerations

In addition to the regulatory impediments discussed thus far, there are several private law considerations that may become relevant to the discussion on legal interoperability within the context of the EOSC. It will not be possible to set out a comprehensive list of all such private-law-related obstacles to legal interoperability, but the issues identified below are likely to be relevant to the general operation of the EOSC.

6.1. Terms of use

The possible model for EOSC is "a pan-European federation of data infrastructures built around a federating core and providing access to a wide range of publicly funded services supplied at national, regional and institutional levels, and to complementary commercial services".[202]

As such, it is envisaged that data made available through the EOSC will not originate from the EOSC, but rather from other repositories. Users will be registered and access data through such repositories, while accepting the applicable repositories' terms and conditions of use. This poses the question of legal interoperability between different terms of use, e.g., on the EOSC portal and the third-party providers. There is a risk that repositories may use different terms and conditions, which may not be fully compatible with each or that impose unintended restrictions for data to be combined with other data from different repositories and carried forward.

Recommendation 28: The EOSC should seek to harmonise participating repositories' terms of use to the extent possible so as to avoid conflicting terms of use where data is combined from different disciplines and repositories.

6.2. Liability

Concerns have been expressed by stakeholders in relation to potential liability resulting from inaccuracies, misuse and breach of privacy laws connected with the (re-)use of data.

As discussed under section 2.6.3.3.1, the CC BY includes a disclaimer of warranty and a limitation of liability, whereas the CC0 does not. The need for, and importance of, such a disclaimer varies, depending on the nature and type of data used and the purpose it is used for. In some cases, the CC BY may provide a higher degree of comfort for some researchers (see **Recommendation 9)**, although, in general, liability issues could also be resolved by other means, such as in the terms of use for repositories or by ethical obligations. Liability does not have to be addressed by the licence. Furthermore, as was pointed out elsewhere, "a license or a waiver that states that a dataset comes without warranty, does not give a wildcard for being able to put "alternative facts" in datasets. e.g., when a document you publish is the authoritative source for something, you are responsible, as part of your job description, to correctly represent

[202] As described in supra note 1, p. 9.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

the real world. Responsibility for a dataset should not be som ething that depends on an open data license."[203]

6.3. Data sovereignty

A report from September 2019 prepared by the RDA-CODATA Interest Group on Legal Interoperability of Research Data[204] highlights a number of concerns raised by stakeholders in relation to researchers' diminished authorial control connected with the use of open and permissive licences. This relates in particular to the use of attribution-only licences (such as the CC BY), waivers or when the data is dedicated to the public domain.

In line with the principle of data sovereignty, some service providers and licensors require information on what part of their data is shared with whom and under what conditions; who had been granted access to the data and for what purposes the data was used for. They may also wish to exclude certain applications of the data for purposes not supported by the licensor, such as military or commercial purposes. The licensors wished to be notified of any reuse of their data and possibly limit the applicable licence to certain geographical areas or to certain markets.

In some specific cases, licensors expressed a concern that the entire database or sets of data may be downloaded and duplicating somewhere else (either for research or for commercial purposes or indeed for other purposes).

Finally, the licensors also wished to have control of the duration of any licence given as well as termination possibilities.

Recommendation 29: The EOSC should assess stakeholders' data-sovereignty-related concerns and consider whether any harmonised "data sovereignty" clauses could be developed and recommended for use by repositories.

[203] Pieter Colpaert, 'CC0 is the best open data license' (Pieter Colpaert, 23 February 2020) https://pietercolpaert.be/open%20data/2017/02/23/cc0.html, accessed 20 November 2020.
[204] See RDA-CODATA Interest Group on Legal Interoperability of Research Data, *Proposed Re-Charter (Revision Sept. 2019)*, https://www.rd-alliance.org/sites/default/files/2019-09-20_IG-Charter-2019-Post-PHIL_0.pdf, accessed 20 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

CHAPTER III: ENABLING LEGAL INSTRUMENTS

1. Introduction

While the focus of this study thus far has been on impediments to legal interoperability connected with the application of the FAIR Principles, it is also relevant, for completion, to look at how certain legal instruments and the general move towards FAIR and Open Data serves as *enabler* of access and reuse of data within the context of the EOSC.

Legal instruments including rules or policies can facilitate FAIR data, Open Data and legal interoperability by introducing 'top down'[205] requirements for publicly funded research data to be Open or to be FAIR or both, and by harmonising rights concerning research data. Some legal instruments may introduce direct obligations, e.g. requiring all data to be Open and FAIR, while others may have an indirect impact, e.g. setting a framework or conditions that encourage the adoption of general access and reuse policies modelled around the FAIR Principles.

Enabling legal instruments that are focused on implementing access and reuse of data, particularly Open Data, may need to navigate through potentially conflicting legal requirements in relation to, for example, IPR, data privacy, protection of national interests or sensitive data. In addition, harmonisation in requirements related to the access and reuse of data may be achieved by different ways and does not necessarily mean uniformity, bearing in mind the principle of subsidiarity.

Enabling legal instruments may vary and include both hard and soft law instruments. Several EU directives address data governance and have their genesis in international conventions. Thus, their context extends beyond the EU.[206] Soft law instruments such as policies and guidelines are used by both international institutions, EU institutions and in national context, for example, by universities and research infrastructures or research performing organisations.

This chapter explores some[207] such enabling instruments and discusses their implications to legal interoperability within the context of the FAIR Principles and the EOSC. By doing so, it also provides some examples for how the regulator addressed some of the issues concerning the implementation of Open or FAIR data principles in contexts other than the EOSC, and the manner it chose to balance the different legal interests involved.

2. EU Directives

At the EU level it is useful to look at three different directives in order to illustrate the contribution of a top-down approach to Open or FAIR data and legal interoperability.

---

[205] RDA-CODATA Legal Interoperability Interest Group, *Legal Interoperability Of Research Data: Principles And Implementation Guidelines* (2016) https://www.rd-alliance.org/rda-codata-legal-interoperability-research-data-principles-and-implementation-guidelines-now, accessed 20 November 2020, Principle Five.

[206] For example, the INSPIRE directive is an implementation of the Aarhus Convention and the Open Data Directive is strongly influenced by the G8 Open Data Charter.

[207] The instruments covered below are by no means exhaustive and have only been included for purposes of illustration. For example, competition law aspects under Article 101 and 102 TFEU and the possibility that data platforms could be obliged to disclose essential data sets in specific circumstances are not discussed in this study.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

2.1. The INSPIRE Directive[208] (IND)

The IND was created to address the basic challenge of having access to data relevant to the development and coordination of environmental policies across the EU Member States. Spatial data was often missing or incomplete as was the metadata and documentation of the spatial data. It was difficult to combine spatial datasets and there was no system to find, access and use spatial data compatible with other datasets. Furthermore, cultural, institutional, financial and legal barriers prevented and delayed the sharing and reuse of existing spatial data.[209]

The IND establishes a legal framework for an infrastructure for the access and use of spatial information. The purpose is broader than the EID (see further below) and facilitates third-party value-added services.[210] The IND addresses *data* as well as data and network *services* and *technologies* provided by the Member States to enable the search for and access to data.[211] The IND applies to spatial data held by public authorities that relates to specific themes listed in annex I-III of the IND, e.g. buildings, soil, land use, geographical names, administrative units and geology. To ensure that the spatial data infrastructures of the Member States are compatible and usable in the EU and transboundary context, the IND requires that common implementing rules are adopted in a number of specific areas (metadata, data specifications, network services, data and service sharing and monitoring and reporting).[212]

The IND does not affect or take precedence over existing ownership and IPR held by public authorities.[213] Member States are required to provide an infrastructure for environmental data and services comprising discovery, view, download and transformation services and finally a service to invoke spatial data service.[214] Limitations on public access to all of the above service levels may be imposed if access would adversely affect international relations, public security or national defence. If the adverse effects relate to the adverse effect of a different interest such as IPR, data privacy or commercial confidentiality, limitations may only be imposed on the latter four service levels (i.e. not on discovery services).[215]

Grounds for limiting access must be interpreted in a restrictive way, taking into account the public interest of access weighted against the interests of limiting or conditioning access.[216]

Information on the conditions that apply to the access and use of datasets and services must be included in the metadata as well as limitations on public access and the reasons for such limitations.[217] Services must be made available to the public free of charge, unless a charge secures the maintenance of the spatial datasets and corresponding services.[218]

---

[208] Supra note 41.
[209] See George Cho and Joep Crompvoets. 'The INSPIRE directive: some observations on the legal framework and implementation' (2019) Survey Review 310, DOI: 10.1080/00396265.2018.1454686
[210] Recital 26 and Art. 1 IND.
[211] Art. 3(1) and Art. 11 IND.
[212] EU Commission, Inspire Knowledge Base, https://inspire.ec.europa.eu/inspire-directive/2, accessed 19 November 2020.
[213] Art. 2(2) IND.
[214] Art. 11(1) litra a-e IND.
[215] Art. 13(1) and (2) IND.
[216] Art. 13(2) IND.
[217] Art. 5(2)(b) and (e) IND.
[218] Art. 14 (1) and (2) IND.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

The section on data sharing in Article 17 of the IND provides general principles governing the sharing of data between public authorities, institutions and bodies of the community both within Member States, with public institutions of other Member States and organisations established by international agreements. Conditions on access and reuse may not contain restrictions likely to create practical obstacles at the point of use[219] and bodies of the community must be granted access to data on harmonised conditions.[220] However, the IND does not provide guidance on e.g. choice of licences or model clauses for access and reuse.

2.2. The Open Data Directive[221] (ODD)

The ODD applies to all public sector information and introduces a fairly comprehensive legal framework for the governance of information that complements the IND.[222] While the ODD should be interpreted consistently with the IND, a general problem of the IND was the very different implementation of licensing regimes as well as significant variation in fees in the different Member States.[223]

The ODD specifically addresses research data[224] and refers to the FAIR Principles in Article 10, which reads as follows:

1. *"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 the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations.*
2. *Without prejudice to point (c) of Article 1(2), research data shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV, insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. In that context, legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account."*

The ODD does not apply if there are competing legal requirements such as IPR, the protection of privacy or sensitive information. An exception to this is the rights in the database directive, which may not be exercised by a public sector body.[225]

---

[219] Art. 17(2) IND.
[220] Art. 17(8) IND.
[221] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information [2019] OJ L 172 ("ODD").
[222] Art. 1(7) ODD. See the section below for on the INSPIRE Directive.
[223] Frederika Welle Donker, From Access to Re-use: a user's perspective on public sector availability (D.Phil thesis Delft University of Technology, 2016).
[224] The directive defines research data 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*".
[225] Art. 1(6) ODD and Art. 7 Directive 96/9/EC.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

Reuse without conditions is the default rule in the ODD and conditions may only be introduced if they can be sufficiently justified, are objective, non-discriminatory and proportionate. Critics have argued that the EU Commission should have taken a step further and prescribe which licences they endorse in order to provide an even stronger degree of harmonisation to the benefit of openness and legal interoperability.[226] However, adoption of specific EU licences is likely to leave too narrow a margin of appreciation for Member States and, from even a broader viewpoint, could also create other legal interoperability difficulties with data of non-EU origin.[227]

Introduction of a general framework rather than embracing specific open licences was considered to strike a better balance between the value of proximity of decisions to the direct stakeholders (in accordance with the principle of subsidiarity), the prospects of harmonisation and the global nature of knowledge collaborations and data sharing. Furthermore, there was no consensus on what licensing schemes would be best suited for the job. Leaving this decision to national institutions gave room for the bottom-up development of, hopefully, a more robust licensing regime for the ODD, compatible with the principles of open science.

The degree to which this decision may interfere with legal interoperability will depend on the degree of consensus in the research community and acceptance of instruments such as the Creative Commons licensing framework or similar solutions. There is a risk with the bottom-up approach that the different Member States may adopt different licensing regimes which may not be fully compatible with each other (or with the EOSC-recommended licences) and therefore compromise legal interoperability.

Recommendation 30: Ensure that open licensing recommended at Member State level are coherent and compatible with licensing recommendations provided by the EOSC.

Exclusive licensing is another issue that is addressed in the ODD. Some public institutions have made agreements with private sector companies regarding the information they have, for example in exchange to having their collections digitalised.[228] Some of these arrangements have taken the form of exclusive licences. The ODD introduces some constraints and time limitations on the use of exclusive licences for public sector information, requiring regular review of the agreement, that the exclusive arrangements must be in the public interest, that

[226] Stephen Abbott Pugh, Missed opportunities in the EU's revised open data and re-use of public sector information directive (*Open knowledge foundation*, 6 July 2019) https://blog.okfn.org/2019/07/09/missed-opportunities-in-the-eus-revised-open-data-and-re-use-of-public-sector-information-directive/ , accessed 20 November 2020.

[227] Ruffus Pollock and Danny Lämmerhirt, Open Data Around the World - European Union in Tim Davies, Stephen B. Walker, Mor Rubinstein, and Fernando Perini (eds.), *The State of Open Data: Histories and Horizons* (African Minds and International Development Research Centre 2019), DOI:10.5281/zenodo.2677762.

[228] See e.g. the collaboration between British Library and Google Books: British Library, *The British Library and Google to make 250,000 books available to all*, https://www.bl.uk/press-releases/2011/june/the-british-library-and-google-to-make-250000-books-available-to-all, accessed 19 November 2020).

the public sector information part is made publicly available and that the exclusivity agreement does not run for more than 10 years.[229]

2.3. The Environmental Information Directive[230] (EID)

The EID implements the minimum standards agreed upon in the Aarhus Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters.[231] The purpose of the EID is to ensure access to information across sectors, inform decision-making and enhance awareness and public participation in environmental issues[232] by means of establishing a right to access to information. Environmental information is broadly defined as any information in any form that may be coupled with environmental issues including measures or activities affecting or likely to affect the environment or designed to protect it. Note that software is not covered by the definition of environmental information in the EID.[233]

The EID applies to public institutions at all levels and to information held by other entities on behalf of public institutions, as well as institutions performing functions on behalf of public institutions.

Restrictions on access to information are only allowed under specific circumstances, and they must be interpreted narrowly.[234] A refusal of access must be based on a balancing of the public interest in disclosure against the interest served by a refusal. IPR, protection of privacy, national security and protection of endangered species are acknowledged as legitimate grounds for refusal of access or redaction of the information. Some of these grounds for exempting information from disclosure may cause uncertainty regarding the scope, e.g. when justice is "adversely affected"[235] or "protection of international relations".[236]

The EID establishes a reactive right to access, i.e. there is no obligation for public authorities to make information publicly available without a request, e.g. in an open database. Access may be subject to reasonable fees. The EID does not provide any specifications on the conditions or terms of use that should apply to the information. Finally, the EID does not include any requirements for FAIR data, due to the fact that it predates the adoption of the FAIR data principles.

2.4. International law

Conventions and bilateral agreements between states often touch upon the topic of access and reuse of research data, requiring that it is kept with the public domain. In the Antarctic Treaty

[229] Art. 12 ODD.
[230] Directive 2003/4/EC on public access to environmental information repealing council directive 90/313/EEC [2003] OJ L 41 ("EID").
[231] United Nations Economic Commission for Europe, Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (1998) ("Aarhus Convention").
[232] Recital 1 EID.
[233] See definition of environmental information at Art. 2 EID.
[234] Art. 4 EID.
[235] Art. 4(2)(c) EID.
[236] Art. 4(2)(b) EID.

from 1959, for example, it is established that "scientific observations and results from Antarctica shall be exchanged and made freely available."[237]

The Nagoya Protocol (NP) on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilisation[238] governs the use of physical samples and associated traditional knowledge, except human genetic resources. It establishes a clearing house model for approving the use of these resources and model clauses are provided. However, the scope of NP does not extend to digital sequence information which may be shared with third parties. It is currently a hotly debated topic whether the principles enshrined in the NP should extend to encompass sequenced data and transfer to third parties.[239] From a legal interoperability perspective it would be crucial that such initiatives carefully consider the agreements and e.g. model clauses that would apply.

From a *human rights perspective*, access to data touches upon the right to information and the right to science. The right to information is a central ingredient in the safeguarding of democratic principles such as accountability and transparency that is intended to ensure the legitimacy of public institutions and the decisions they make. This right is modified by opposing rights such as the right to privacy[240] and data protection[241] and the right to intellectual property.[242] The right to science is stipulated in Article 27 of the Universal Declaration on Human Rights. In a statement from UNESCO, Open Science and the access and reuse of data are deemed instrumental to "*increase scientific collaboration and access to networks, strengthening scientific culture, enhancing the involvement of citizens in research activities and increasing the access to scientific data and information for communities, policy and decision makers*".[243]

The *environmental agenda* has also been an important driver of the access to data initiative, and the EID from 2003 is the implementation of the Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters ("Aarhus Convention").[244] This convention has a dual goal of enhancing public awareness and discourse of environmental matters and to provide a basis for better policy development and coordination given the transnational nature of environmental problems. Access to data is instrumental to both of these goals. However, EU Member States that have adopted the convention have also interpreted it differently.[245] Thus, the EU legal framework is instrumental

[237] Antarctic Treaty (1959), Art. III, (1)(c), available at: https://documents.ats.aq/keydocs/vol_1/vol1_2_AT_Antarctic_Treaty_e.pdf, accessed 19 November 2020. For further references for bilateral and multilateral agreements, see e.g. supra note 46.
[238] Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity, https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf, accessed 19 November 2020.
239 Jon Ambler et al., 'Including Digital Sequence Data in the Nagoya Protocol Can Promote Data Sharing' (Trends in Biotechnology, 9 July 2020) DOI:https://doi.org/10.1016/j.tibtech.2020.06.009.
[240] Art. 7 Charter of Fundamental Rights of the European Union ("EU Charter") and Art. 8 European Convention on human Rights ("ECHR").
[241] Art. 8 EU Charter/ Art. 8 ECHR.
[242] Art. 17(2) EU Charter.
[243] UNESCO statement on open science at 'UNESCO Takes the Lead in Developing a New Global Standard-setting Instrument on Open Science' (UNESCO , 28 November 2019) https://en.unesco.org/news/unesco-takes-lead-developing-new-global-standard-setting-instrument-open-science, accessed 19 November 2020.
[244] Supra note 231.
[245] Supra note 223, p. 76.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

to harmonising the interpretation of the obligations and implementation required by the convention.

The G8 Open Data Charter,[246] which is one of the key influencers of the ODD, has an economic perspective on access and reuse of data. The G8 Open Data Charter emphasises access and reuse of data as a key driver of all stages of innovation from discovery to research and development as well as a key element in service innovations.

3. Soft law instruments

3.1. National policies

There are many initiatives on the national level to introduce and implement policies on the access and reuse of data, many of which relate to the general underlying principles of open government data that promote transparency, accountability and value creation.[247] Most EU countries have joined the Open Government Partnership,[248] while the EU legal framework has provided specific instruments such at the ODD, the EID and the IND discussed above.

Following the enactment of the predecessor of the ODD, the Public Sector Information (PSI) Directive of 2013, the EU Commission issued a notice with guidelines on the recommended licences to use for open data.[249] The headline in the guideline reads "standard licences", but it also considers the use of notices as well as drafting of individual, national licences as alternative to existing standard licences such as the Creative Commons.

Creative Commons licences are used to govern PSI in many EU jurisdictions and beyond.[250] The UK Open Government Licence (OGL) is the default licence choice where nothing else is stated and it is compatible with the CC BY. However, different licences are allowed for the datasets made available via the government data portal.[251] In Belgium, data available from the government's open data website is free to use by default unless otherwise stated.[252]

The ODD, from 2019, requires EU Member States to adopt open access policies, aiming at making public sector information openly available by default, while publicly funded research data must be made available in accordance with the FAIR Principles by July 2021.[253] It should thus be expected that more detailed national policies on access and reuse of research data will be published within the next year.

[246] International Open Data Charter, https://opendatacharter.net/principles/, accessed 19 November 2020.
[247] OECD, Open Data Government, https://www.oecd.org/gov/digital-government/open-government-data.htm , accessed 19 November 2020.
[248] Open Government Partnership, https://www.opengovpartnership.org/, accessed 19 November 2020.
[249] European Commission, Commission notice on guidelines to standard licences, datasets and charging for the reuse of documents [2014] OJ C 240.
[250] See e.g. Austria: http://data.gov.au/about; Italy: http://www.dati.gov.it/ and Japan: https://www.data.go.jp/terms-of-use/terms-of-use/, accessed 19 November 2020.
[251] UK Government, https://data.gov.uk/terms; Canada is an example of another country with their own Open Government Licence: Open Government Licence – Canada (Government of Canada, 18 June 2019) https://open.canada.ca/en/open-government-licence-canada, accessed 19 November 2020.
[252] Data.gov.be, Terms of use, https://data.gov.be/en/terms-use, accessed 19 November 2020.
[253] Art. 10 ODD.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

More specifically to FAIR, the 'FAIR in Practice Task Force' observes eight main approached towards introducing policies on FAIR practices in Europe.[254] These fall into three general categories: national approaches, funding or infrastructure requirements and community/local approaches. The Netherlands is an example of a country that has adopted national approach to implementation of FAIR Principles via the Dutch National Plan for Open Science[255]. Ireland[256] and Norway[257] are other examples of national strategies implementing the FAIR Principles although not always mentioning them explicitly. France, Belgium and Germany provide examples of an approach to the implementation of FAIR Principles via funding, infrastructure or compliance requirements, where, for example, non-profit funders require compliance with FAIR Principles[258] or compliance with FAIR Principles is a requirement for national data policies.[259] Finally, examples of community/local approaches are found in, for example, the UK with the Concordat on Open Research Data,[260] exemplifying a more bottom-up approach to the implementation of the FAIR Principles. Furthermore, universities and other research institutions across Europe have started to encourage or require compliance with the FAIR Principles.[261]

A recent study on Open Science policies in Europe[262] suggests that Member States' national policies provide good coverage of key Open Science elements including providing a definition for data, recommending data sharing, encouraging the production of data management plans and addressing intellectual property issues. Areas that are less well covered include expectations around data citation, providing data availability statements as well as costs associated with research data management and making data FAIR.

3.2. Other funders or institutional policies

Many policies on the governance of data have been developed on an institutional level. Institutions such as the WHO and the OECD, European Research Infrastructures[263] and

[254] Supra note 14.
[255] W.J.S.M van Wezenbeek et al., *National Plan Open Science* (Ministerie van Onderwijs, Cultuur en Wetenschap 2017) https://doi.org/10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65.
[256] Government of Ireland, National Open Research Forum, *National Framework on the Transition to an Open Research Environment* (July 2019). http://norf-ireland.net/wp-content/uploads/2019/07/NORF_Framework_10_July_2019-2.pdf, accessed 19 November 2020.
[257] Norwegian Ministry of Education and Research, *National Strategy on Access to and Sharing of Research Data* (2018) https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/, accessed 19 November 2020.
[258] See, for example, Welcome Trust, *Good research practice guideline* (April 2018) https://wellcome.org/grant-funding/guidance/good-research-practice-guidelines, accessed 19 November 2020.
[259] Ouvrir la Science, *National plan for Open Science* (4 July 2018) https://www.ouvrirlascience.fr/national-plan-for-open-science-4th-july-2018/, accessed 19 November 2020.
[260] Supra note 17.
[261] See for example, Utrecht University, Research Data Management Support, How to make your data FAIR (2019) https://www.uu.nl/en/research/research-data-management/guides/how-to-make-your-data-fair; TU Delft, *TU Delft Research Data Framework Policy* (August 2018), https://d1rkab7tlqy5f1.cloudfront.net/Library/Themaportalen/RDM/researchdata-framework-policy.pdf, accessed 19 November 2020.
[262] Vanessa Proudman, Thordis Sveinsdottir and Joy Davidson, *An Analysis of Open Science Policies in Europe* (v6) (2020) http://doi.org/10.5281/zenodo.3689450.
[263] See, for example, the PANOSC data policy addressing raw data, metadata and results, available here: *PaN-data EuropeDeliverable D2.1Common policy frameworkon scientific data* (2011) panosc.eu/wp-content/uploads/2019/05/PaN-data-D2.1_PolicyFramework.pdf, accessed 19 November 2020.

national institutions such as universities have adopted their own policies on data sharing and reuse.

A WHO-initiated survey on the existence of open data policies in international institutions such as the World Bank, UNICEF and WIPO shows a heterogeneous landscape regarding the existence of policies on access to data as well as their content.[264]

In Europe, the EU Commission and the European Research Council (ERC) have had an active role in promoting the adoption of open access policies via funding mechanisms introducing requirements of open access to research data.[265] Beneficiaries of funding under Horizon 2020 must observe Article 29.2 and 29.3 of the model grant agreement, requiring open access to scientific publications and to research data.[266] The model's grant agreement requires beneficiaries to deposit both data necessary for validation of the results and other data including metadata, according to FAIR Data Management Plans (DMPs).[267] The data must be made available for third parties to access, mine, exploit, reproduce and disseminate free of charge. Information about tools and instruments necessary to validate the data must be disclosed, and where possible made available to third parties. While both FAIR and Open access to research data has become the default in Horizon 2020, the Commission also recognises that there are good reasons to keep some or even all research data generated in a project closed, taking into account considerations such as confidentiality obligations, commercialisation and IPR, privacy concerns, security, etc.

European Research Infrastructure Consortia (ERICs) are required to adopt a data policy. In the guidelines provided by the EU Commission for ERIC data policies, adoption of open access and open source regimes is encouraged[268] but no specific guidelines for example on licences, reuse and metadata are provided. There are other challenges that research infrastructures may face in the implementation of the FAIR data principles.[269] Initiatives such as the PANOSC

[264] See in this regard the *Policy on use and sharing of data collected in Member States by the World Health Organization (WHO) outside the context of public health emergencies*, Annex 2 22 (August 2017) https://www.who.int/docs/default-source/documents/data-sharing-policy-collected-by-member-states-outside-of-public-health-emergencies.pdf?sfvrsn=69f8e65_2, accessed 19 November 2020.

[265] See in relation to ERC grants: European Research Council, Open Access, https://erc.europa.eu/managing-project/open-access, accessed 19 November 2020.

[266] As well as a requirement for a machine-readable electronic copy of the published version. See: *EU Grants: H2020 AGA —Annotated Model Grant Agreement:V5.2* (26 June 2019) https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf#page=245, accessed 19 November 2020. See also, European Commission, *H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020* (version 3.2) (2017) https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf, accessed 19 November 2020.

[267] See European Commission, *H2020 Programme - Guidelines on FAIR Data Management in Horizon 2020* (version 3.0) (2016) https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, accessed 19 November 2020.

[268] See also EU Commission, ERIC Practical Guidelines - Legal Framework for an European Research Infrastructure Consortium (European Union 2015) https://op.europa.eu/en/publication-detail/-/publication/c6647f05-874e-4cdd-af70-22ade4759930, accessed 19 November 2020, Annex II, Art. 26.

[269] Jonathan Taylor, 'The hows' and 'whys' of data' (Physics World, October 2017) http://live.iop-pp01.agh.sleek.net/physicsworld/reader/#!edition/editions_neutron_2017/article/page-22463, accessed 19 November 2020.

project[270] are actively trying to address these issues despite the lack of more specific guidelines form the ERIC legal framework.

In conclusion, there are numerous enabling legal and policy instruments that provide incentives or oblige publicly funded research to adhere to the Open and FAIR data principles. However, it is important to secure coherency between the requirements of such legal instruments and the general recommendations for the EOSC.

Recommendation 31: The EOSC should encourage developments in the implementation of EU and national enabling legal instruments but at the same time coordinate directly with relevant entities and Member States to ensure that implementing national policy recommendations are harmonised and coherent with general recommendations provided by the EOSC.

[270] The Photon and Neutron Open Science Cloud, https://www.panosc.eu/, accessed 19 November 2020.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

CHAPTER IV: RECOMMENDATIONS

1. Open access to research data is an enabler of legal interoperability. The promotion of FAIR Principles should go hand-in-hand with efforts to make data open in accordance with the principle that data must be "as open as possible and as closed as necessary".

2. Regardless of whether the data is Open or not, all new data made available through the EOSC should be FAIR by design.

3. Copyrightable **metadata** should be free from any restrictions and assigned a public domain waiver. The Creative Commons No Rights Reserved (CC0) or the Public Domain Dedication and Licence (PDDL), or an equivalent statement of rights should be preferred.

4. Metadata should include a standardised human and machine-readable statement of rights, legal restrictions, applicable licences, and, where relevant, additional conditions of use (including applicable jurisdictions) of the data that they are assigned to.

5. The EOSC should provide a mechanism, for example in the Rules of Participation, or by way of guidance, to facilitate the implementation of Recommendation 4 above in a harmonised manner.

6. Automatic database (*sui generis*) rights should be addressed in the licence chosen so that: (1) *sui generis* are covered by any permissive licence or a waiver, and (2) they do not result in unintended restrictions on reuse of data by EU-based users compared to non-EU users (due to the territorial nature of such rights).

7. All copyrightable components of the research data and their respective licences should be clearly identified in the metadata and assigned the correct rights-holder. In the case of database (*sui generis*) rights in repositories, the applicable (permissive) licence should be included in the terms of use of the repository.

8. Open and permissive licences, authentication and authorisation mechanisms and the use of restricted data access collections are preferred over the use of ad-hoc specific contracts entered into between a rights-holder and a data user. An additional contract should only be used if it is the least restrictive way to ensure compliance with legal restrictions or in other justified cases. In such cases smart contract solutions should be developed and put into use.

9. Copyrightable **data** should be FAIR and, to the greatest extent possible, be made part of the public domain or assigned a permissive licence, unless legal or legitimate reasons apply. The Creative Commons No Rights Reserved (CC0) or the Public Domain Dedication and Licence (PDDL) or an equivalent statement of rights should be preferred. In cases where liability is a concern that cannot be addressed by other means, the CC BY 4.0 licence is an appropriate alternative.

10. The use of Creative Commons licences is generally not recommended for licensing source code for software. Only open and permissive software licences such as the MIT License, the Apache License 2.0 or the equivalent should be used for software.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

11. From a licence compatibility perspective, attribution should be pursued by means of moral and ethical obligations e.g., the European Code of Conduct for Research Integrity or the development of Persistent Identifiers, or by way of a standard form of acknowledgement, rather than by means of a licence such as the CC BY 4.0. The CC0 is, in general, preferred over the CC BY 4.0, although both are generally permissive.

12. The overall number of recommended licences for data available through the EOSC should be minimised to the greatest extent possible. In addition to the CC0, CC BY and MIT, permissive licences such as the ODC-BY or the O-UDA are preferrable along with information about their compatibility with other licences – see Recommendation 14.

13. The EOSC should develop or encourage the development of a centralised source of knowledge, guidance and support researchers on copyright and licences related issues so that researchers can make informed decisions on what licence they should assign to their data.

14. A list of EOSC-recommended licences and their compatibility with other licences, as well as with Member States' recommended licences, should be provided to data producers, rights-holders and users, in order to avoid an inadvertent breach of copyright and with a view to harmonise and reduce the overall number of recommended licences.

15. Instances of expired or non-existent copyright, or where data is already in the public domain, should be clearly marked, for example, by way of the Creative Commons Public Domain Mark (CC PDM) or equivalent.

16. The EOSC should encourage repositories (for example, through the Rules of Participation) to incorporate harmonised mechanisms to validate and allow for the update of restrictions, rights statements and conditions of use on data as these may change over time. For this purpose, the principle that data licences can only become more permissive not more restrictive after first being shared within the EOSC should be considered.

17. The EOSC should adopt a uniform set of recommendations or guidance on how to handle copyrightable datasets where the rights-holder(s) are unknown or not reachable and the data has no licence assigned to it. An 'orphan data' standardised notice and related legal implications could be considered for such cases.

18. In accordance with the principle of "as open as possible, as closed as necessary", EOSC policies should take into account commercial incentives and facilitate the seeking of IP protection in justified cases where the disclosure of the data may compromise the ability, for example, to file for patents or protect trade secrets.

19. Metadata should indicate reusability restrictions on software or data due to pending or existing patent claims or when data had been redacted due to commercially confidential information. The metadata should also provide information about the relevant patent proprietors, licensing options and any relevant additional information.

20. The EOSC should facilitate easy and intelligible platform solutions, e.g. through smart contracts, that allow scientists and their institutions to acquire licences where necessary, or

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

to achieve "freedom to operate" confirmations where research exemptions, experimental use exemptions and/or patent pledges apply.

21. The EOSC should create or encourage the creation of a trusted environment with reliable access control and authorisation procedures which will accommodate different techniques for shielding data in order to enable researchers to share data while providing them the possibility to protect IP and trade secrets related to innovative products and processes.

22. Encourage the development of basic guidelines on GDPR issues for researchers on identifying personal data and on implementing the "Privacy by Design and by Default" approach.

23. The EOSC Rules of Participation (RoP) should include a requirement for users, repositories, data and service providers (and any other data controllers and data processors) to implement appropriate measures to ensure compliance with the GDPR, for example, by requiring repositories to indicate the conformity of their data and services with the GDPR.

24. Where possible, data anonymisation is an appropriate approach to protect personal data. When done correctly, anonymised data can be shared for secondary purposes – such as further scientific research – without placing individual privacy at risk.

25. The EOSC should address the issue of cross-border data transfer (outside the EU/EEA) and require the implementation of additional security and organisational safeguards in the SCCs where international data transfers take place. Where possible, this should be done in a machine-actionable manner such as in a Service Level Agreement (SLA) and the data management system.

26. Restrictions on the access and reuse of data should be proportionate and applied in cases of applicable legislation or legitimate reasons. The EOSC should consider preparing, in consultation with stakeholders, guidance on what are considered to be 'legitimate reasons' that go beyond existing legislation and which could justify the introduction of additional restrictions on access to and reuse of data. The guide could also provide a proposed list of such legitimate reasons.

27. Adopt procedure for monitoring or reporting violations of use conditions and leakage of sensitive data.

28. The EOSC should seek to harmonise participating repositories' terms of use to the extent possible so as to avoid conflicting terms of use where data is combined from different disciplines and repositories.

29. The EOSC should assess stakeholders' data-sovereignty-related concerns and consider whether any harmonised "data sovereignty" clauses could be developed and recommended for use by repositories.

30. Ensure that open licensing recommended at Member State level are coherent and compatible with licensing recommendations provided by the EOSC.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644

31. The EOSC should encourage developments in the implementation of EU and national enabling legal instruments but at the same time coordinate directly with relevant entities and Member States to ensure that implementing national policy recommendations are harmonised and coherent with general recommendations provided by the EOSC.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644