Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782   Start Date of Project: 01/01/2019   Duration: 40 months

Deliverable 8.4 Report on ethical and legal issues and implications for EOSC

Dissemination Level	PU
Due Date of Deliverable	28/02/2022 (M38)
Actual Submission Date	15/03/2022
Work Package	WP8 - Governance/Sustainability/Quality Assurance
Task	Task 8.3 Legal and Ethical Issues
Type	Report
Approval Status	Approved by EC - 04 May 2022
Version	V1.0
Number of Pages	p.1 – p. 40

Abstract:

The report identifies ethical and juridical issues for open access/data sharing and addresses how these issues can have implications for the EOSC.

History

Version	Date	Reason	Revised by
0.1	01.11.2021	First draft	Marita Helleland, Ina Nepstad, Håkon Jørgen Tranvåg, Mathilde Steinsvåg Hansen, Trond Kvamme, Siri Tenden, Ingvild Eide Graff, Marianne Høgetveit Myhren (CESSDA/NSD (Sikt))
0.2	05.11.2021	Peer Review	Emiliano Degl'Innocenti (CNR), Maurizio Sanesi (CNR)
0.3	12.11.2021	Revised draft	Marita Helleland, Ina Nepstad, Håkon Jørgen Tranvåg, Mathilde Steinsvåg Hansen, Ingvild Eide Graff, Marianne Høgetveit Myhren, Vigdis Namtvedt Kvalheim (CESSDA/NSD (Sikt))
0.4	29.11.2021	Review by Coordinator	Ivana Ilijasic Versic (CESSDA), Martina Drascic Capar (CESSDA)
0.5	30.01.2022	Revised draft	Marita Helleland, Ina Nepstad, Håkon Jørgen Tranvåg, Mathilde Steinsvåg Hansen, Marianne Høgetveit Myhren, Vigdis Namtvedt Kvalheim, Eva Payne (CESSDA/NSD (Sikt))
0.6	08.02.2022	Peer Review	Lorna Ryan (ESS ERIC), Axel Boersch-Supan (SHARE ERIC), Franciska M. G. de Jong (CLARIN ERIC) and Erzsébet Tóth-Czifra (DARIAH ERIC)
0.7	09.02.2022	Revised draft	Marita Helleland, Ina Nepstad, Håkon Jørgen Tranvåg, Mathilde Steinsvåg Hansen, Marianne Høgetveit Myhren, Vigdis Namtvedt Kvalheim, Eva Payne, Trond Kvamme (CESSDA/NSD (Sikt))
1.0	14.02.2022	Final version submitted	CESSDA/NSD (Sikt)

Author List

Organisation	Name	Contact Information
CESSDA/NSD (Sikt)	Mathilde Steinsvåg Hansen	mathilde.hansen@sikt.no
CESSDA/NSD (Sikt)	Ina Nepstad	ina.nepstad@sikt.no
CESSDA/NSD (Sikt)	Marita Helleland	marita.helleland@sikt.no
CESSDA/NSD (Sikt)	Håkon Jørgen Tranvåg	hakon.tranvag@sikt.no
CESSDA/NSD (Sikt)	Trond Kvamme	trond.kvamme@sikt.no
CESSDA/NSD (Sikt)	Siri Tenden	siri.tenden@sikt.no
CESSDA/NSD (Sikt)	Ingvild Eide Graff	ingvild.graff@sikt.no
CESSDA/NSD (Sikt)	Marianne Høgetveit Myhren	marianne.myhren@sikt.no
CESSDA/NSD (Sikt)	Eva J.B. Payne	eva.payne@sikt.no
CESSDA/NSD (Sikt)	Vigdis Namtvedt Kvalheim	vigdis.kvalheim@sikt.no
University of Michigan, ESS ERIC REB Chair	Julie de Jong	jjosefos@umich.edu
CLARIN ERIC CLIC	Paweł Kamocki	kamocki@ids-mannheim.de
CLARIN ERIC CLIC	Andreas Witt	witt@ids-mannheim.de
CESSDA/TAU-FSD	Arja Kuula-Luumi	arja.kuula-luumi@tuni.fi
CESSDA/GESIS	Elizabeth Lea Bishop	ElizabethLea.Bishop@gesis.org
ALLEA Permanent Working Group on IPR	Alain Strowel	alain.strowel@uclouvain.be
Dublin Institute for Advanced Studies, Vice President of ALLEA and chair of the ALLEA Task Force on Open Science	Luke Drury	ld@cp.dias.ie
Norwegian University of Life Sciences and member of the ALLEA WG on Science and Ethics	Deborah Oughton	deborah.oughton@nmbu.no
ALLEA Permanent Working Group on Science and Ethics	Bertil Emrah Oder	boder@ku.edu.tr

Chair of the ALLEA Working Group on E-Humanities	Maciej Maryl	maciej.maryl@ibl.waw.pl
ALLEA Scientific Policy Officer	Mathijs Vleugel	vleugel@allea.org

Executive Summary

Background: The main purpose of SSHOC is to create the social sciences and humanities area of the European Open Science Cloud (EOSC), thereby facilitating access to flexible, scalable research data and related services streamlined to the precise needs of the Social Science and Humanities (hereinafter SSH) community [1, 2]. The ambition of EOSC is to develop a 'web of FAIR data and services' for science in Europe, by building on existing research infrastructure and services supported by the EC, Member States and research communities. It brings these together in a federated 'system of systems' approach by aggregating content and enabling services to be used together [1, 2].

Main aim: This report examines possible ethical and legal issues with open access/sharing of social science and humanities research data and possible implications for the EOSC.

Scope of the report: The report identifies some ethical and juridical issues for open access/data sharing and addresses how these issues can have implications for the EOSC. The anticipated implications identified in this report relates to use and governance. The information presented has mostly been identified based on published reports and documents. A selection of experts in ethics and IPR has been asked to provide input on ethical issues and IPR issues for data sharing/open access. These parties were chosen as they had relevant experiences with data sharing and ethics and intellectual property rights (IPR).

Main outcome: Promoting data sharing among the scientific community is important and is a political goal. There are policy guidelines and incentives on sharing research data, both at international and national levels [3]. It has become an essential practice, whether between different groups within the same organisation, between partners in larger platform endeavours, or even, as in growing open data movements, with the public. Data sharing contributes to the advancement of science in small increments, in addition to ensuring replicability and increased scientific rigour of individual studies. Data sharing enables new insights from existing data, and lets organisations make full use of this core resource. However, it also introduces ethical and legal issues.

Conclusion: As part of identifying ethical issues, the report identifies a risk of research subjects losing control over their personal data, a risk that scientific data might be used for commercial interests, that sharing data sources potentially can be coupled to persons can increase the risk of harm for the research subjects. In addition, the report identifies issues with informed consent, and the challenges of ensuring that each research subject fully understands what participation in a project entails, and what storage and access via an open cloud platform entails. As part of identifying legal issues, the report finds that the content of contracts between a researcher and research subjects in study can have implications for the aggregated 'system of systems' approach of the EOSC, as it can affect how data can be shared in a federated and common EOSC catalogue. Furthermore, the report finds that this environment may have implications for the handling of intellectual property rights (IPR). These issues can be considered differently depending on the context of scientific research and the influence of private actors/commercial

undertakings and must be resolved before data are shared and managed in each project. Results from this report indicate that further work on how to resolve/handle ethical and legal issues related to open access and sharing of research data should be actioned. It will be helpful to envision a mechanism able to verify that legal and ethical issues are considered and resolved, before making data available through a federated EOSC catalogue.

Highlighted recommendations in the report:
- The EOSC should provide overarching, transnational ethical guidelines for data sharing and open access, and services connected to the EOSC should ask researchers or research institutions who wish to share data to comply with such guidelines. Other reports have also recommended that the EOSC provide templates for data sharing agreements, national support services and best practice bench marking.
- The EOSC should evaluate whether it is a service open and available to all, including purely commercial agents or entities, or if it should be a service that prioritises researchers and academic institutions. Efforts should be made to ensure, through clear guidelines and/or requirements that services connected to the EOSC have provided the data owners and depositors an opportunity to opt out of giving commercial entities future access to data that is indexed by and can be found through the EOSC.
- As a federated system, the conditions, and safeguards of the EOSC should have clear and explicit requirements that cover the protection of vulnerable groups, and such requirements should be set for the services that wish to join the EOSC.
- Regarding existing datasets, the EOSC should take steps to safeguard datasets have been evaluated as being scientifically and ethically sound before they are shared. The services connected to the EOSC might put in place requirements for the ethical considerations that should be made by researchers, and demand confirmation that these requirements have in fact been met before data can be shared.
- A plan should be made on how to ensure IPR licensing arrangements are in place, when sharing data through the EOSC.
- The EOSC should also have a plan on how to verify which data can be indexed, discovered, and shared through the cloud, to act in accordance with agreements made with participants.

Abbreviations and Acronyms

ALLEA	All European Academies (the European federation of Academies of Science and Humanities
EOSC	European Open Science Cloud
ESS ERIC	European Social Survey, European Research Infrastructure Consortium
ESS ERIC REB	European Social Survey, European Research Infrastructure Consortium, Research Ethics Board
CLARIN ERIC CLIC	Common Language Resources and Technology Infrastructure, European Research Infrastructure Consortium, CLARIN Legal and Ethical Issues Committee
EC	European Commission
FAIR	Findable, Accessible, Interoperable, Reusable
FEK	The Norwegian National Research Ethics Committees
GDPR	General Data Protection Regulation
HUNT	The Hunt Study – a longitudinal population health study in Norway
IPR	Intellectual property rights
NESH	The National Committee for Research Ethics in the Social Science and Humanities, (Norway).
OECD	Organisation for Economic Co-operation and Development
SERISS	Synergies for Europe's Research Infrastructures in the Social Sciences
SSB	Statistics Norway
SSH	Social Sciences and Humanities
SSHOC	Social Sciences & Humanities Open Cloud
TDM	Text and data mining
WP	Work Package

Table of Contents

1 Introduction

1.1 General

The main purpose of SSHOC is to create the social sciences and humanities area of the European Open Science Cloud (hereinafter EOSC) thereby facilitating access to flexible, scalable research data and related services streamlined to the precise needs of the Social Science and Humanities (hereinafter SSH) community [4, 5]. The ambition of EOSC is to develop a 'web of FAIR data and services' for science in Europe, by building on existing research infrastructure and services supported by the EC, Member States and research communities. It brings these together in a federated 'system of systems' approach by aggregating content and enabling services to be used together [1, 2].

SSHOC Task 8.3, Legal and Ethical Issues, has been assigned with work to support the main purpose of SSHOC, to be achieved as part of Work Package (hereinafter WP) 8, Governance/ Sustainability/ Quality Assurance. In the current document, Task 8.3 presents a report on ethical and legal issues with open data access, and implications for EOSC.

Promoting data sharing among the scientific community is important and has become an essential practice, whether between different groups within the same organisation, between partners in larger platform endeavours, or even, as in growing open data movements, with the public. Data sharing helps the advancement of science in small increments rather than through single studies, and ensures replicability, thereby helping to confirm the findings of a study. Data sharing enables new insights from existing data, and lets organisations make full use of this core resource. But it can also introduce new legal and ethical issues or risks.

This report first describes several ethical issues that can arise with open access and data sharing and addresses how these can have implications for the EOSC. The anticipated implications identified in this report relates to use and governance. Some of the information on which the report is based stems from published reports and documents. In addition, the report also presents input provided by some selected experts that have extensive knowledge about research data and ethical-/IPR issues.

The report also identifies some legal issues related to intellectual property law and contractual law that may arise with data sharing and reuse of research data. Legal issues related to data protection and how these can have implications for the EOSC, have been identified in a former SSHOC report, *"Report on the impact of the GDPR and its implications for EOSC"* and will therefore not be further elaborated in this document [4]. However, a short summary of the findings of the *"Report on the impact of the GDPR and its implications for EOSC"* is presented [4].

1.2 Scope of the report

The focus of the report is to identify ethical and legal issues that can arise in relation to open access and data sharing and address how these can have implications for the EOSC.

In Chapter 1.3, a short description is given of what the FAIR principles are, what we mean with research data, and how ethics is interpreted in this report. This information has been included in the beginning of the report, as it can be considered background information relevant for contextualisation of the findings. The information also describes how the FAIR principles are understood in the report, how research data is understood and how ethics has been interpreted. The report presupposes that the EOSC, likely through a distributed network of repositories, will store research data containing both personal and anonymised information. Some of the identified ethical and legal issues, and their implications for the EOSC, will follow from these interpretations.

It can be hard, if not impossible, to separate legal and ethical concerns. However, it is worth noting that even though legal and ethical concerns may overlap, acting in accordance with the law does not necessarily mean acting ethically. An act may be both legal and unethical at the same time. This issue is not elaborated in the report but is touched upon in Chapter 2.6. Chapter 2 addresses ethical issues, whilst Chapter 3 focuses on legal aspects. The identification of data protection issues as a part of legal issues was submitted in a former SSHOC report, "Report on the impact of the GDPR and its implications for EOSC" [4]. This document will therefore mainly focus on ethical issues. However, some legal issues related to contractual law and intellectual property rights are mentioned in Chapter 3.

Selected experts on research ethics and IPR have been consulted. The contributions on ethical issues are presented in Chapter 2, whilst IPR issues are placed in Chapter 3. The aim of these consultations was to gather input on ethical- and IPR issues that can be linked to data sharing, especially to data sharing solutions such as EOSC. The parties were asked questions reflecting that topic. Some of the contributions have been provided in writing, whilst others have been gathered during interviews through a digital channel. The parties were chosen as they were considered to have relevant experiences that could be of value for this report. It is important to stress that findings from these consultations cannot be generalised to represent the view of all experts on ethics and IPR in general.

Although this report identifies several issues, it is important to clarify that this document does not necessarily present an exhaustive list of the ethical or legal issues related to open access and data sharing, and which could have implications for the EOSC. Additional ethical challenges, e.g., issues related to working with cultural heritage data, and issues associated with different flavours of the Creative Common licenses, should be elaborated in further work. In addition, some of the ethical and legal issues referred to in this report could be further elaborated in future work. For instance, and as suggested in

an EOSC-Nordic report, "Open Science in the Nordics: Recommendations on Legal Issues", a dedicated EOSC Task Force could be created to support knowledge about legal issues on software, patents, copyright, access to data, competition law, state aid rules and similar issues [6].

A central take away of the report is that putting legal and ethical support structures in place, which are available for scholars across Europe is crucial in the transition towards FAIR and open research culture in the SSH. The recommendations given in this report is not directly given to researchers, but instead to federated services as EOSC. The purpose is therefore not to provide practical examples to be used by researchers, e.g., on difficulties of translating legislation or Code of Conducts. Instead, the report intent to identify issues that will apply in general, and which must be acted on at high levels. This will, hopefully, also facilitate to improve the everyday life of research and sharing of data.

Within this report, a research subject in SHH can be defined as the individual about whom the SHH researcher conducts their research. The research subject may actively and voluntarily participate in research, or they may be analysed without actively participating. The definition covers both identifiable and non-identifiable individuals.

A contextual aspect of data sharing is the fact that the understanding of the concept of 'research data' may be different in different contexts. There seem to be a rather clear consensus on the basic elements of the definition of the concept (the EOSC defines research data as: "Data collected or produced in the course of scientific research activities and used as evidence in the research process, or commonly accepted in the research community as necessary to validate research findings and results"). When addressing research data in this report, it is narrowed to social science and humanities data. Data within the SSH disciplines are heterogenous and may consist of many types of data, both quantitative and qualitative, like datasets, text, audio, video and images, and the analytical units may be individuals, organisations or geographical units like regions and countries. These data may be collected through a wide range of methods like for instance interviews, surveys, observations, measurements, tests, content coding, aggregations, and web harvesting; and the data sources may vary from language corpuses, social media content, historical and archeological artifacts, to literature, music and other cultural expressions.

It is important to stress that different research institutions within the social science and humanities environment will be providing different data types to the EOSC and the legal issues related to contract and IPR, and research ethics issues will vary accordingly.

This report will not identify ethical and legal issues related to biological data. However, as this has been addressed in the reports "D5.3 Data access protocol for accelerometer data, linked to survey data, conforming FAIR principles (Access to biomedical data)" and "D5.1 Guidelines for ethics considerations in making biomedical survey data FAIR (Access to biomedical data)", a short summary of the main findings of these reports are presented in Chapter 2.7 [7, 8].

1.3 FAIR principles and ethics in research

FAIR principles

FAIR is an acronym for the four principles: Findability, Accessibility, Interoperability, and Reusability. The goal is to formulate concrete guidelines that will "serve to guide data producers and publishers as they navigate around […] obstacles, thereby helping to maximise the added-value gained by contemporary, formal scholarly digital publishing" [9].

The first principle states that both data and metadata should be Findable [10, 11]. In order to find the data, they must have a reliable resource identifier. This identifier should be both unique and persistent, so that it will always identify the same data resource in a consistent way. Furthermore, the data must be described properly. This entails applying rich metadata, because a "digital resource that is not well-described cannot be accurately discovered". The metadata should also include the identifier of the data source. Finally, and perhaps obviously, the data should be registered in a searchable resource, such as registries or other similar applications [12].

The second principle states that, once the data has been found by the user, it must also be Accessible [10, 11] without un-needed limitations affecting implementation, documentation and human intervention.[9]. The protocol for access should be free and open, but the principle of accessibility does not mean that the data itself must be open and freely accessible in all cases. Some data require constraints, for ethical or legal reasons. The access protocol should therefore allow for an authentication and authorisation procedure for the requester [12].

The third principle states that the data must be Interoperable [10, 11], i.e., making sure that the data can be integrated with other data, both from the same research field, but also from other fields of research. The aim of interoperability is to support the exchange and interpretation of data between different systems and contexts for both computers and humans, without the need for mappings and other translation algorithms. To ensure a high level of interoperability for digital resources it is critical to rely on a shared, accessible, and globally understood language for knowledge representation, and a vocabulary following the principles of FAIR [12].

The last principle states that data must be Reusable [10, 11]. This means the data should be properly documented and curated, that they conform to certain community standards, with clear terms and conditions for how they should be reused. Both the legal and licensing status of the data should be clear and readable for humans and machines. This will not only facilitate validation of previous results from an original study, but also facilitate new projects based on these original findings.

Ethics in research

When addressing ethical issues related to open access and data sharing, it may be appropriate to start by defining and placing the concept 'ethical' in the given context. Broadly speaking, addressing what is

'ethical' in research means addressing matters that have to do with accepted principles or norms of conduct for researchers and research institutions. Research within the social sciences and humanities (hereinafter SSH), like in all other areas of research, requires the knowledge and implementation of both scientific methods and ethical norms of conduct, in order to safeguard sound scientific practice. Above in Chapter 1.2, a short description of the characteristics of SSH data is presented.

Ethical norms or principles within research may dictate, for example, that voluntary consent from the research subject is essential, that the envisaged research be necessary and beneficial for society, that research should avoid harm, that the risk taken in research should be proportional to the human importance of the problem to be solved, or that the research subject should have the right to withdraw at any time [13]. Furthermore, The Declaration of Helsinki (1964) puts forward that science or society's need for new knowledge can never justify exposing the research subject to unnecessary risks: the goal of generating new knowledge "can never take precedence over the rights and interests of individual research subjects" [14]. The declaration also states that within research with vulnerable groups and individuals, the vulnerable groups or individuals should receive specifically considered protection, and they should stand to benefit from the results of the research [14].

Although such fundamental, normative principles for research ethics are generally agreed upon across the board, each research community or research discipline may still follow their own internally interpreted and accepted variations of fundamental ethical norms. This variation between different research disciplines and their respective ethical norms for conduct can lead to challenges related to interdisciplinary cooperation. The same can be said for cooperation across national borders: Different EU member states may have differing ethical norms of conduct as well as different systems for handling ethics in research. See for example the SERISS report "Synopsis of Policy-Rules for Collecting Biomarkers in Social Surveys", in which experiences with ethical and legal requirements for research are documented for twenty different EU member states [15]. According to the SERISS report, all twenty countries have established some type of system relying on research ethics committees, i.e., that research committees or similar advisory organs provide advice or formal reviews of ethical issues in research projects. But the report also concludes that the actual organisation of such committee systems differs between each member state, thus making it complicated to gain ethical approval in transnational research projects [15].

However, some overarching European ethical codes of conducts exist, for example, *The European Code of Conduct for Research Integrity* (2017) by the European Federation of Academies of Sciences and Humanities (hereinafter ALLEA), which serves as a reference document for research integrity in EU funded research projects. By 'research integrity', ALLEA refers to how research can be conducted in accordance with accepted ethical frameworks as well as with sound scientific practice [16]. The code identifies four fundamental principles of research integrity: reliability, honesty, respect and accountability [16]. These principles of research integrity, the code states, should guide researchers when they engage with practical, ethical and intellectual challenges within research [16]. See Chapter 2.1 below, on how such ethical principles for research relate to open access and data sharing.

Seeing as ethical guidelines are precisely *guiding*, and not necessarily given by law as injunctions or prohibitions, the individual researcher must be able to both interpret ethical guidelines as well as make actual ethical evaluations and apply ethical norms in particular situations. For example, the researcher who is about to share data may have to find a balance between ethical guidelines stating both that research material should be made available to other researchers for secondary analysis and further use, and that the researcher is obligated to safeguard privacy and confidentiality for their research subjects. In other words, the researcher must balance ethical norms of transparency with ethical – or judicial – norms of privacy and confidentiality and make an evaluation before sharing data.

A data sharing environment such as the EOSC should both support researchers' ethical obligations and share the responsibility for following ethical norms for research. This can, for instance, be achieved by the EOSC providing overarching, transnational ethical guidelines for data sharing and open access, and by asking the researchers or research institutions who wish to share data to comply with such guidelines. This echoes another SSHOC report, "Open Science in the Nordics: Recommendations on Legal Issues", which stresses that the EOSC should guide institutions towards secure data sharing by providing templates for data sharing agreements [6]. In addition, the report mentions that "providing some national support services and best practice benchmarking could speed up the establishment/adoption of OS/EOSC support services at RPOs."[6]

This is also supported in an EOSC pilot white paper, "White Paper 4: Ethics Supporting Document to D3.3 Draft Policy Recommendations", which states that "... EOSC also has an important role to play in raising awareness about ethical issues within scientific communities, as well as in wider ethical debates, for instance around the interplay of scientific knowledge and society". In addition, it mentions that the input might be introduced as the EOSC evolves, and the resources have been determined [17]. The paper states that awareness of ethical issues and its relevant responses must not be considered as a peripheral topic or an addition to main scientific activity. Instead, it must be integrated into the core of the EOSC [17].

2 Ethical issues and implications for the EOSC

Data sharing is becoming an essential part of research. This includes sharing data between different groups within the same organisation, between partners in larger platform endeavours, or even, as in growing open data movements, with the public. Sharing enables the gaining of new insights from existing data, but it also introduces new ethical risks. In the following, the report presents several ethical issues related to open access and data sharing. However, as stressed in Chapter 1.2, it is important to clarify that an exhaustive list of the ethical issues that could have implications for the EOSC will not necessarily be presented. In chapter 2.6, contributions from experts on ethical issues are presented.

This report will not identify ethical and legal issues related to biological data. However, a summary of findings in the reports "D5.3 Data access protocol for accelerometer data, linked to survey data, conforming FAIR principles (Access to biomedical data)" and "D5.1 Guidelines for ethics considerations in making biomedical survey data FAIR (Access to biomedical data)" are presented in Chapter 2.7.

2.1 Why is open access and data sharing desirable?

"Science is a social enterprise: independent and collaborative groups work to accumulate knowledge as a public good" [18].

Following both ethical principles and scientific ideals of transparency, a researcher is obligated to disseminate knowledge. This obligation implies that researchers should make their research readily available both internally within the research community, and externally, to society. Keeping research open and accessible safeguards communication between the research community and society. Giving research "back" to society can even be seen as a democratic imperative within research.

During the last two decades, the obligation to disseminate knowledge has been increasingly tied to publishing through open access. The Berlin Declaration (2003) defined open access as "a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community" [19]. As such, open access can be seen as one of the ways the research community fulfils its obligation to make knowledge available for society.

Transparency is not just a scientific ideal which should be valued as an abstract, perfect goal [18]. Access to open data could also have practical implications when seeking to improve scientific progress. Over the recent decade, there has been a growing attention towards a "replication crisis" in social and medical sciences. Replicability – the ability to replicate the evidence of a previously confirmed hypothesis – is the "cornerstone of science" [18, 20-22].

The possible reasons for the replication crisis in science are multiple and complex and will not be covered in detail in this report [23, 24]. However, it is worth noting that several of the suggested solutions to the problem deal directly with open access and data sharing, which is relevant for the EOSC.

The lack of successful replication could be a result of differences in study design. Open access to all data material of a published study is therefore vital for "conducting direct replications and accumulating scientific knowledge" [22].

However, making the data underlying a published study available will not only improve the quality of replication studies, but it will also "improve data aggregation methods and confidence in reported results" [25]. Increased transparency can add protection against fraud, misuse and misrepresentation of a finding, can reduce the error rate of both data and analysis, and can facilitate further interpretation of already reported findings or "[establish] new findings that are not included in the original published report" [26].

It is worth noting, though, that in a survey of 600 researchers in psychology, only 10% of the respondents had ever shared their primary data in an online repository. More than 60% said that there is no "suitable repository" to share their data, and 71% said that it would be "somewhat likely", "likely" or "very likely" that they would share their data if there was "a suitable place to share" [27]. This illustrates both the lack of appropriate (or well-known) data repositories, and the fact that there is a desire among researchers to share their data, and shows the potential for a depository such as EOSC.

It is thus of vital importance to facilitate the sharing of data, and in particular, FAIR data. In this way, the EOSC would be a beneficial tool for improving scientific knowledge and progress. In some situations, however, data cannot be made publicly available, for instance due to ethical and/or legal regulations [28]. In such cases, one solution could be to only make available information that describes the data, and not the data in itself. Such a solution is suggested in "The ethical principles of research with human research subjects and ethical review in the human sciences in Finland" [28]. The EOSC can also be a beneficial tool for improving scientific knowledge and progress by sharing such descriptive data.

2.2 Respect for research subjects and risk of harm

A key ethical issue when it comes to data sharing, especially through open access, can be identified by asking 'Who has an interest in the sharing of data?'. This report has touched upon both external, societal interests in data sharing, i.e., the availability of research for the public good, as well as internal interests within research itself, i.e., sharing data for the purpose of replication and verification and thereby safeguarding scientific standards for research. However, ethical principles also dictate research's relationship to the persons or cultures included in research. See for example ALLEA's formulation of *respect* as a fundamental principle of research integrity, i.e., "Respect for colleagues, research subjects, society, ecosystems, cultural heritage and the environment." [16] The need for sharing data must be balanced against the potential harm implicated for the persons that data have been collected from. In other words, the risk taken in sharing data should be proportional to the societal benefit of sharing such data.

It can be helpful to look at The General Data Protection Regulation (EU) 2016/679 (hereinafter GDPR) for an analogy in this ethical issue: According to GDPR[1], the body which, alone or jointly with others, determines the purposes and means of the processing of personal data (hereinafter controller) must consider "the risks of varying likelihood and severity for the rights and freedoms of natural persons", risks which are connected to potential "physical, material or non-material damage" for these persons [29, 30]. Furthermore, when a data protection impact assessment is needed, the GDPR[2] states that "[w]here appropriate, the controller shall seek the views of data subjects".

2.3 The loss of control of personal data

As addressed above in Chapter 1.2, it can sometimes be hard to draw a clear line between the legal and the ethical frameworks for sharing data. The two are partly intertwined and - when discussing the ethical implications for the subject at hand - it is natural to bear in mind the principles of the GDPR. These fundamental principles of the GDPR are first lawfulness, fairness and transparency; second purpose limitation; third data minimisation; fourth accuracy; fifth storage limitation; sixth integrity and confidentiality; and finally, accountability [31].

Of these, the principles of transparency and purpose limitation are perhaps the most vital when discussing the issue of data sharing. As previously mentioned, the GDPR requires us to take the view of the research subjects. It is the individuals' data, and they have the right to know exactly how their data are being processed and for what purpose. However, it is worth noting that the GDPR is not designed to prevent the sharing of personal data: one of the objectives of the GDPR is precisely to secure the free movement of personal data within the Union [32].

For the individuals to remain in control of their personal data, an open cloud such as the EOSC must remain committed to a design with strict, yet transparent protocols for access. This will not go against the principles of FAIR data: Accessibility does not equal open access. As the H2020 program guidelines on FAIR data noted, data should be "as open as possible and as closed as necessary", and sound data management is an essential part of best practice in research [33]. Does this mean it is possible that sharing data can be reconciled with the individual's right to control?

Researchers who are collecting data must provide the individuals with information, not just about their current research project, but also about any potential future use. This means that the individuals must be informed about data-sharing plans, such as where they plan to deposit the data. Furthermore,

---

[1] See Art. 24 GDPR; Responsibility of the controller and Recital 75; Risks to the Rights and Freedoms of Natural Persons

[2] Art. 35 GDPR; Data protection impact assessment (number 9)

researchers must inform the research subjects of the purposes for which the data might, and might not, be reused in the future [34].

This need for information does not have to restrict future research. Article 5 no. 1 letter b of the GDPR specifies that "further processing for (…) scientific or historical research purposes (…) shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes". In other words, the GDPR encourages the re-use of personal data for future research even if those purposes are not articulated at the point of initial processing [35].

The research subjects must also receive information about who will have access to the data in the future. The GDPR adopts a broad definition of "research," and the line between academic research and commercial interests is often blurred. Nonetheless, it is recommended that the EOSC evaluates whether it is a service open and available to all – including purely commercial agents or entities – or if it should be service that prioritises researchers and academic institutions. At least, efforts should be made to ensure, through clear guidelines and/or requirements, that services connected to the EOSC have provided the data owners and depositors an opportunity to opt out of giving commercial entities future access to data that is indexed by and can be found through EOSC.

2.4 Vulnerable groups and the risk of re-identification

Before sharing data, the risk of harm for research subjects must be considered. This is particularly pertinent when it comes to sharing sensitive data or data about vulnerable groups or individuals, or, as is often the case, data that are both sensitive and concern vulnerable groups or individuals. Such data should be given special consideration in research, for the purposes of avoiding harm. It can be argued that the persons from whom sensitive data are collected should stand to benefit from their data being shared.

For example, it is important to make a balanced ethical evaluation of whether the sharing of data about ethnic minorities could involve risks for the minorities, either currently or in the future, or whether the minority will stand to gain from their data being shared. Members of ethnic minorities may be more vulnerable to re-identification through their data being shared, but at the same time they might have an interest in having their data being widely shared and re-used, considering that their ethnic group might be underrepresented in research. It is worth noting that open research data can reduce the need for other researchers to gather the same type of data, which in turn may reduce pressure on selected small population groups [28].

Paradoxically, sensitive data can be worthy of protection and at the same time be of high value for further research. Historically, groups of ethnic or religious minorities, or patient groups with certain diseases, may have been either under- or over-studied in many fields of science. Today, the under-studied groups might have an interest in participating in research and might pursue an active role in promoting more research among their members. This means that data containing special categories of personal data, and

which might be hard to anonymise, is of great value not only for the research community but also for the individuals themselves.

Still in line with the analogy from the GDPR, it can be argued that balancing risks and benefits of sharing data, for example, about ethnic minorities, demands that the researcher first considers the perspectives of the groups or persons from whom the data are collected. The Sámi Parliament in Norway can be looked to as an example: The Sámi Parliament has decided that before health research can be carried out among the Sámi people, collective consent must be gained from a representative expert committee. In this way, health research within the Sámi community can be rooted in indigenous peoples' right to self-determination, as well as be carried out with respect for the diversity and uniqueness that characterises Sami culture and Sami communities [36].

Although such collective agents cannot give individual consent on behalf of the data subject, an evaluation of the research project may still help ensure that the potential high risks of the processing are balanced out by the project's possible beneficial effects for the group.

The researcher is ethically obligated to safeguard that the views and wishes of the research subjects or their representatives have been considered before their data can be shared. Ideally, the researcher should evaluate the proportionality between the risks of harm in sharing data vs. the possible gains for the groups or persons included / represented in such data. If the risks of harm exceed the possible benefit of sharing data, the data should not be shared. As a federated system, the conditions and safeguards of the EOSC should have clear and explicit requirements that cover the protection of vulnerable groups, and such requirements should be set for the services that wish to join the EOSC.

2.5 Ethical assessments of existing data

Having ethical guidelines in place before the collection of data will safeguard the possibility of sharing the data at a later point. However, it may also be within the scope of the EOSC to share already collected data, e.g., data that were collected without ethical assessments or before current ethical guidelines were in place. Sharing such existing data may pose ethical challenges.

As already mentioned, it may be necessary to hear the perspectives of the groups or persons from whom data have been collected, and this will also apply to existing data. At the very least, the ethical principles of respect and transparency dictate that the research subject should be informed that their data will be shared. Furthermore, before sharing existing data, an ethical evaluation should be made between possible risks and benefits, both internally for the research community and externally for the research subjects and for society at large. Finally, if existing data should be deemed ethically problematic, for example, because of the way it was collected, a data service should consider whether the data can be shared at all. Arguably, data that have ethical shortcomings may also be scientifically problematic. If such problematic data ends up being shared through a federated service like the EOSC it might challenge the

public's understanding of research, and perhaps of the EOSC itself, as a trustworthy multi-disciplinary environment for scientifically reliable data.

The EOSC should take steps to ensure that existing datasets are evaluated as being scientifically and ethically sound before they are shared. Current ethical guidelines may not straightforwardly be imposed on existing data. However, this report still recommends that ethical considerations should be made ahead of sharing data. The services connected to the EOSC might, for example, put in place requirements for the ethical considerations that should be made by researchers, and demand confirmation that these requirements have in fact been met before data can be shared.

2.6 Input from experts

Selected experts on research ethics have been consulted. The aim of these consultations was to gather input on the ethical issues that can be linked to data sharing, particularly to a data sharing environment such as the EOSC. Contributions were given, either through interviews or in writing, by Arja Kuula-Luumi at the Finnish Social Science Data Archive (hereinafter referred to as Kuula-Luumi), Elizabeth Lea Bishop at GESIS, (hereinafter referred to as Bishop), Paweł Kamocki and Andreas Witt at the CLARIN Legal and Ethical Issues Committee (hereinafter Kamocki and Witt), Luke Drury and Deborah Oughton at ALLEA (hereinafter in this chapter referred to as ALLEA), Julie de Jong at the University of Michigan, and Chair of the ESS ERIC REB (hereinafter referred to as de Jong), an expert in the SSH domain, and by Statistics Norway and The HUNT Study at the Norwegian University of Science and Technology. The contributions provided are presented in the following paragraphs.

Overview of input received

The parties were asked what kind of ethical issues arise when research data from social science and humanities are to be shared through an open science cloud.

ALLEA emphasised that open science does not necessarily mean that everything must be open. A researcher is entitled to impose restrictions when this is justified. Therefore, when data are open to a broad community, the corresponding ethical responsibility is enhanced and even more important. Furthermore, ALLEA stated that different ethical issues can be identified and divided into types according to how the data are collected, the nature/sensitivity of the data, and how the data are to be used. Social science data are often very sensitive data, which can also reflect how the data have been gathered. For instance, gaining consent can be particularly challenging when data have been produced or shared by people publicly on digital platforms.

According to Kamocki, the types of ethical issues that may arise when sharing data depend on the type of data to be shared, as well as on the system of ethical values used to assess the situation. Together with his colleague, Witt, he is currently working on the following categorisation of possible ethical issues:

1) privacy-related issues, e.g., such issues arise when information concerning (also indirectly) an individual is shared and the concerned individual is not given the possibility to effectively keep this information private; 2) property-related issues, e.g., when someone else's intellectual (text, software, databases), but also cultural (traditional dances, stories, rituals, but also endangered languages) property is handled in a disrespectful way, without proper evaluation of harm from the owner's perspective; 3) equality related issues, e.g., when datasets are not well-balanced (e.g., an ethnic, language or age group is over-or under-represented), which may lead to bias and potentially to discrimination against certain groups of subjects/users; 4) transparency-related issues, e.g., when the origin of the data is not revealed or properly described. This is a very broad category of ethical issues, ranging from re-purposing data (i.e., re-using data collected for one project/purpose for another project/purpose, without being transparent about it) to using data from which only limited quality can be expected (e.g., machine-translated) without being transparent about it, all the way to fabricating the data altogether; 5) freedom-related issues, e.g., when the data was collected from participants who cannot fully exercise their freedom of choice, and who are not given the possibility (also technically) to change or withdraw their contribution at a later stage.

Furthermore, Kamocki and Witt state that many of these situations are covered by the legal framework, and illegal actions are always unethical. However, ethics go further than the law, and some legal actions may still be unethical, so simply abiding by the law does not sufficiently guarantee ethicality of our actions.

Similarly to ALLEA, de Jong highlighted that one of the central ethical issues connected to data sharing is informed consent. She stressed that it is important to be aware that the concept of consent may change over time, and that those changes may have an impact on research data. Another issue linked to consent, according to de Jong, is the issue of personal, differential understanding of the consent process across populations. Furthermore, regarding cross-national and cross-cultural populations, de Jong mentioned issues with equity. There may be issues of language, coercion, legal status, and how phrases about open science are understood, and this can impact responses and research across wide populations. One may, for instance, conduct research in a multicultural population where respondents interpret consent differently. Sometimes lack of understanding of the consent process can lead to people electing not to participate in research, which can be problematic for the outcome of the study. Therefore, de Jong stresses that the main issue at hand is how to safeguard the informed consent process, in such a way that each respondent truly understands what the open science cloud is, what its consequences are and what participation in general will involve.

According to Kuula-Luumi, ethical and legal issues are not much different when data are delivered by data repositories using open science cloud services, compared to situations where data are delivered using a data repositor's own platform. The main issue is tackling ethics, data protection/GDPR and other applicable legal issues already in the collection and processing phase.

Kuula-Luumi emphasised the issue of whether the marketplace which is offering data services is responsible for any legal or ethical issues related to data release. ALLEA underlined that even though some ethical responsibility to ensure the quality of data lies with platforms, institutions and data cloud management, the fundamental responsibility must lie with the data provider/the researchers. Similarly, according to an expert in the SHH domain, it is primarily the individual researcher's responsibility to ensure that data collection and sharing is done in accordance with acknowledged norms. However, the expert noted that platforms such as the EOSC will have a certain ethical responsibility, but this is perhaps more an indirect responsibility compared to the researcher's responsibility. This is because a third party may not have the knowledge required to fully understand the details of the data and its context.

Similarly, de Jong stated that the responsibility to act in accordance with ethical norms primarily lies with the researcher. While discussing the possibility of anonymous datasets becoming identifiable over time, de Jong noted that before archiving data in the public, the researcher bears the responsibility to collapse datasets as much as possible and to consider at what level to share variables such as geographic and other identifiable data. Still, de Jong underscored that ideally there should be some sort of partnership between researchers and platforms such as the EOSC. As an example, de Jong mentioned that the archive also bears responsibility in helping the researchers understand what is important when storing and sharing data. Furthermore, she emphasised that it is important that the researchers, when storing data at an open science cloud, think through who might gain access to their data when it is shared, and hence which data to share and how to share them.

Statistics Norway and The HUNT Study was also asked to address which ethical issues can be linked to data sharing, especially to a sharing solution such as the EOSC. In its response, Statistics Norway presumed that the EOSC will facilitate sharing of anonymous or public data, i.e., data that is generally known or processed in such a way that it does not provide information that can be traced back to persons or other statistical units. They stated that sharing data for research purposes means sharing for more insight, transparency and accountability, and that the FAIR principles provide guidelines on what must be facilitated to achieve effective sharing. They also referred to other published principles, such as the European Guidelines for Research Integrity [16]. The HUNT Study's position was that ethical issues may be identified based on facts linked to each project/sharing of data.

Statistics Norway pointed out that when processing data for the purposes of sharing knowledge and providing more insight, inspiration can be found in the UN's 10 general principles for official statistics, which also include ethical issues related to both data collection and processing [37]. They stressed that when assessing ethical issues concerning data sharing, it is important to consider the entire value chain in collection, facilitation, and sharing of data. This is because the premises for the collection and processing of data contain ethical issues that carry over into the actual sharing of the data.

The parties were asked how principles of research ethics can be safeguarded whilst sharing data through an open science cloud.

ALLEA suggested having procedures that are agreed upon prior to data collection, to ensure relevant measures. For instance, measures must be taken to ensure that the research subject know what they are consenting to, not just regarding participation in one specific project but also emphasising that data will be shared with others and what sharing data will imply. Similarly, and as mentioned above, de Jong underscored that the main issue is how to conduct the informed consent process, in such a way that all respondents truly understand what the open science cloud is, what it means, what the consequences are and what participation in general will involve. According to de Jong, doing this contributes to safeguarding the principles of research ethics whilst sharing data through an open science cloud.

ALLEA noted that if research subjects are to be informed that data will be anonymous, considerations must be made as to whether the data can in fact be made anonymous, or whether it will be possible to re-identify research subjects based on for instance scrambling keys, based on advances in artificial intelligence or by combining variables in a dataset. It will also be important to consider who has the power or responsibility to determine how the data are to be processed, e.g., funders of research, the research performing organisation, or in general by applicable statute law. ALLEA noted that it is important to discuss relevant issues between researchers and people participating in research, and to raise awareness of relevant issues, as it can facilitate the safeguarding of principles. One suggestion, mentioned by Kuula-Luumi, is that if an archive does not have self-archiving services, the archive might check ethical and legal issues when considering the data for archiving. Kuula-Luumi stressed that data that are shared must be collected in accordance with ethics (e.g., voluntary participation, exact information on data sharing) and applicable regulations (e.g., in accordance with data protection laws such as the GDPR).

According to the expert in the SHH domain, a case can be made that whoever is funding the study should have a say in the question of future access and re-use of data. However, it was emphasised that a key principle of research ethics is the research subject's autonomy, which is safeguarded through informed consent, and this must be the researcher's responsibility to safeguard. A central issue will always be how present and future use of the data by researchers continues to respect the autonomy of the research subjects. An open cloud service, such as the EOSC, should take great care not to make available or release data which is potentially identifiable without explicit informed consent from participants in research. The practical implications of this suggest that when data is being shared through EOSC it is required that depositing individuals or organisations confirm that participants have been informed of the deposit arrangements, including risks, however remote/low, arising from combining different datasets.

According to Kamocki, how to safeguard principles of research ethics whilst sharing data through an open science cloud depends on how 'research ethics' is defined. However, Kamocki states that no matter which definition is adopted, such principles must be safeguarded through two basic steps: Firstly, researchers should be trained to be made aware of ethical issues related to data sharing in their field of expertise. Moreover, the research community – consisting of such aware researchers – should appropriately sanction unethical practices in data sharing, as well as reward high ethical standards. Secondly, researchers that have been made aware of ethical issues should be given appropriate technical

tools and standards for handling ethical issues, for example related to data attribution and tracing the origin of data.

The parties were asked how ethical uncertainties can be resolved when research data are shared across countries, cultures and legal systems.

All parties seemed to acknowledge that this is in fact an issue and that there are large cultural differences which can be hard to manage within Europe.

However, according to Kuula-Luumi, if data are anonymised, there are no uncertainties as long as data have been originally collected in accordance with ethics and applicable regulations. For instance, if research subjects in a study have been informed that data will be shared in anonymised form through the EOSC for research, teaching and studying, or even shared free for all, this will be ethically sound. It was noted that from an ethical point of view, defining and adhering to the purpose of re-use of the data is essential. The information that research subjects have been given will affect this. When considering personal data, Kuula-Luumi underlined that issues connected to data protection laws, such as the GDPR, must be resolved in the collecting phase, when providing data protection information (for instance in privacy notices) and determining the legal basis (for instance consent), so that data service is available to deliver/transfer data outside EU/EEA according to GDPR Chapter V.

Bishop stated that what is at stake when sharing data is competing ethical duties or rights, such as openness vs. protection/privacy. According to Bishop, the nature of an ethical dilemma is that it cannot ever be resolved, only managed. She stated that it is therefore essential to understand underlying principles, e.g., dignity and privacy, and to assess if and how it is ever morally justifiable to encroach on these rights for other goods, such as openness of data and transparency of scientific methods. De Jong agreed that an ethical dilemma must be managed rather than resolved. She noted that when collecting data, a researcher must, for instance, both acknowledge what is culturally appropriate in one country and acknowledge the privacy of the research subject.

De Jong emphasised that understanding whether informed consent includes sharing of data can also prevent uncertainties about data sharing. In addition, de Jong highlights that those applicable uncertainties might be resolved by having people with experience to consult with, for instance people on an ethics board. However, de Jong acknowledged difficulties connected to such an approach. For instance, the countries within Europe have different ethics boards and guidelines, which can make it hard to find common guidelines/agreements between the countries. Such a risk was also mentioned by the expert in the SHH domain, who stated that if an international project must seek ethical assessment in several countries, there could be a risk of different ethical committees in various countries all reaching different conclusions. De Jong followed up on this concern by noting that differential ethical assessments across multiple countries in a comparative survey may threaten data quality in the comparative context because of potential differences in ethically-informed protocols. According to ALLEA, international

collaboration can be helpful but still be hard to organise. ALLEA stressed that the International Science Council, which covers the humanities as well as the social sciences, and similar organisations, can be relevant to include when managing ethical uncertainties.

To be able to resolve or manage ethical uncertainties, ALLEA highlighted the importance of debating and discussing applicable uncertainties in the open, at the start of a collaboration, e.g., to be clear about what differences may be and how they can be managed. Similarly, Kamocki noted that even though there cannot be one single solution to the issue of sharing data across countries, cultures and legal systems, all solutions should involve transparency and stakeholder dialogue. According to Kamocki, sufficient importance should be attached to subjective evaluations of ethical harm, i.e., no stakeholder's complaint should be overlooked without thorough examination and a carefully performed balance of interest test.

Finally, ALLEA noted that sharing research data across countries and legal systems is usually carried out in a good manner. People are most often happy to contribute to a common good, for example improving healthcare and improving knowledge on social conditions.

Summary of expert views

In summary, the challenges identified by experts were not only pertaining to the sensitivity of the data itself, but also to how it is collected and how it will be used. One fundamental issue is that of informed consent, and the challenges of ensuring that each respondent fully understands what participation entails, what the open science cloud is, and what it means. The prime responsibility to tackle these challenges lies with the researcher, though a federated data sharing system such as the EOSC must state clearly what it expects from data providers, services and researchers who wish to share data.

The experts agreed that the key point is again the informed consent. Prior to data collection there must be agreed upon procedures for data collection, for use and re-use, and for future access to the data. Such matters must be carefully thought through, as this is essential information for the research subjects when they are asked to give an informed consent.

With different ethical committees, cultural norms, and national laws, cross-national sharing of research data might be difficult despite the uniform European framework of GDPR and other applicable law. The respondents underlined the fact that ethical challenges might never be resolved but that they can be managed. To identify and consider potential challenges at the outset of a project is central. Furthermore, seeking advice from internationally oriented research committees could be of great value.

2.7 Implementation of biological data in survey research

Health interview and examination surveys can offer useful information for policy, practice, and research purposes. Technological advancement continues to offer new ways of collecting health data, beyond asking survey questions. This includes the collection of biological samples for laboratory analyses and tracking the behaviour, health, or movements of the participants. Including such data in social survey

databases involves new methods of data collection for survey researchers, as well as new ethical issues that need to be addressed. Consequently, the already applied measures to ensure compliance with ethical standards and data protection requirements must be reconsidered when including such innovative variables in social survey databases.

The Survey of Health, Ageing and Retirement in Europe (SHARE) is an interdisciplinary panel study among the population aged 50 plus in 28 European countries and Israel [38]. SHARE covers a wide variety of topics including economic, social, and health related aspects of life. In addition to the questionnaire, physical performance tests such as grip strength, peak expiratory flow, walking speed, and accelerometer data are now part of the SHARE interview.

The report "D5.3 Data access protocol for accelerometer data, linked to survey data, conforming FAIR principles (Access to biomedical data)" outlines the SHARE accelerometer study and demonstrates how the data access plan conforms to the FAIR principles [8]. This report deals with the challenge of adapting the FAIR principles to the access of biomedical data available for the research community. As an intermediate step to the actual data access, the report provides the data access plan for making accelerometer data available. This report describes the data access plan for accelerometer data collected within the SHARE and aims to inform and guide other researchers, survey methodologists and data archives.

Furthermore, the report "D5.1 Guidelines for ethics considerations in making biomedical survey data FAIR (Access to biomedical data)" provides guidelines for ethical considerations in making biomedical survey data FAIR, considering the specific setting of a survey data collection and the fact that biomedical data are a special category of personal data (data concerning health as defined in the GDPR[3]) that require special protection when made accessible and reusable [7].

As stated by Börsch-Supan et al, the use of biological data in survey research provides new opportunities for the research environment, but also requires new ethical considerations. The principle of 'do not harm' must be extended beyond breach of confidentiality to include physical harm, such as temporary discomfort or pain. Each project should be reviewed by an ethics committee to ensure the practical implementation of the ethics standards and principles, as well as data protection measures.

Reusability, but also open access and data sharing, are valuable requirements for replication and validation of previous research findings, and thereby the promotion of scientific knowledge. For the data to be reusable, both researchers and computer systems must be capable of assessing whether the data is suitable for the attempted research. This sets requirements for both the data and metadata to be sufficiently well-described.

---

[3] See Art. 9 GDPR; Processing of special categories of personal data

3 Legal issues and implications for the EOSC

In the following chapter some selected legal issues are presented and how they can have implications for the EOSC. The identified issues are related to intellectual property rights (hereinafter IPR) and contractual law. A summary of findings on data protection issues from another SSHOC report, "Report on the impact of the GDPR and its implications for EOSC" is also included. At the end of Chapter 3, input on IPR issues from experts are presented.

3.1 Data protection

Research data often contain personal data and storing/sharing these data includes processing of personal data. Within EU/EEA countries, this results in the application of data protection laws, including the GDPR.

The "Report on the impact of the GDPR and its implications for EOSC" is a report on the impact of the GDPR on research and its possible implications for EOSC. It describes and compares national implementation of the GDPR across Europe by examining the national laws of a selection of European countries and by conducting interviews with 14 researchers in three of these countries (Norway, Italy and Denmark) [4]. It also describes implications the GDPR might have for the EOSC.

The purpose of these interviews was to get insight to researcher's experiences on how/if GDPR had affected their work. In sum, most of the researchers asked, identified some changes after the introduction of GDPR. According to the report, it was stressed that projects often are prolonged as it, compared to before GDPR, takes longer time to receive data. In addition, some medias shared less data than before GDPR and that collaboration is harder due to different interpretations of GDPR. Some had experienced that the process for applying for research founds after GDPR had changed, as there was more paperwork and more focus on ethics and data protection.

The report indicates that national legislation supplementing the GDPR is implemented differently throughout Europe. Furthermore, there seems to be different interpretations from judicial experts on how GDPR should be interpreted in some countries, as well as different national supplementary provisions, making harmonisation more challenging [4]. It can also result in making it more challenging to share datasets containing personal data across borders in general. Where the GDPR leaves room for interpretations, it is indicated that ethical considerations can affect the choice of legal basis (in GDPR) [4]. This indicates that GDPR has not fulfilled its purpose, which is harmonising the legal framework as well as legal practice throughout Europe, and that additional work is necessary to improve cross-border data flow within Europe [4, 5].

In an EOSC-Nordic report, "Open Science in the Nordics: Recommendations on Legal Issues" it was also mentioned that different legal experts sometimes can give different interpretations. The report presents some measures to help facilitate cross-border research, reduce subjective interpretation of the

legislation and help organisations to have one common way of sharing data [6]. The report highlights harmonisation of data protection laws throughout Europe by creating Code of Conducts, soft-law tools and "how-to-do" guidelines. Further, the report highlights that EOSC must keep into account that small institutions not necessarily have the resources to hire personnel for legal support, hence providing training/consultancy in a centralised matter would be valuable [6].

The GDPR leaves room for different exceptions when the purpose of processing personal data is research and archiving purposes. For instance, the GDPR art. 5 no. 1 letter b presents an exception for purpose limitation in cases where the data are being processed for research purposes. In addition, in the same article, letter e, an exception from storage limitation is presented. A question can be raised whether these exceptions can be made even though it deviates from information that has been given to research subjects in a research project, and possibly contracts made between the researcher and research subjects as referred to in Chapter 3.3 of this report. This should be elaborated in further work.

3.2 Intellectual property rights ("IPR")

Introduction

Intellectual property rights are rights related to intellectual products, and an intellectual product is the result of creative activity that is not just related to a specific object [39]. This can, for instance, be research data and datasets in general. Sharing of research data on a federated service like the EOSC, can lead to some intellectual property law (hereinafter IPR) issues. Some identified issues will be presented in the following paragraphs.

Individuals, researchers, organisations, businesses, and governments all face common challenges when data are accessed and shared. But as noted by the EU, OECD and others, it is highly desirable that data produced by or used in public research should have as few restrictions as possible when it comes to re-use terms and conditions to make data useful and beneficial for research and other public-interest purposes [40, 41].

The EOSC is one of many initiatives that aims to make data useful across borders and across disciplines; it is recognised by the Council of European Union as a science, research, and innovation "data space", as defined in the European strategy for data (a strategy that aims to create a "single market for data" that will "allow it to flow freely within the EU and across sectors for the benefit of businesses, researchers and public administrations" [40]).

The ambition of the EOSC is to "...provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes". Furthermore, the EOSC aims to enable

seamless access to, FAIR management of, and "…reliable reuse of research data and all other digital objects produced along the research life cycle" [1].

The details of how this "multi-disciplinary environment" will be implemented and how it will facilitate access to and sharing of data is still not particularly clear, but the overarching goal will be achieved by attempting to bring together inter-domain research infrastructures to facilitate greater sharing of resources and data across several domain-specific infrastructures. As such, it will be an interdisciplinary catalogue of data and services with domain-focused sub-portals (or hubs).

In this sense, it is fair to assume that the practical implementation of the EOSC will consist of partly technical aspects (shared ontologies and vocabularies, machine-actionable metadata, etc.) and non-technical aspects (human-curated knowledge hubs, data collections and data catalogues).

IPR and licensing of research data in a cross-national environment

This combination of a multi-disciplinary and cross-national approach (i.e., connecting research data from different jurisdictions), can increase the friction between the desire for more openness and sharing of research data on the one hand, and issues connected to IPR on the other.

IPR in the context of research and research data - means copyright, sui generis database rights, patent, trademark, trade secret, and other similar laws that establish a regime for access, use or reuse of data, metadata, or data products [36].

The differences in national implementation of copyright principles and laws may affect the sharing of data across borders. The same is true for the differences in methods, academic traditions, and views on – and definitions of - data across research subjects and disciplines. Both aspects may create obstacles for the legal interoperability of research data that are shared through the federated services of the EOSC project.

Legal interoperability 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) [42]. In other words: when data are combined from multiple sources, the resulting dataset will incorporate the accumulated restrictions imposed by each source [43].

This implies that the fewer restrictions there are in the source datasets, the fewer restrictions will be contained in the combined or derivative datasets. Which further 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 license [44]), legal interoperability will be maximised [45].

However, deciding on and setting the most appropriate license for a specific set of data presupposes a considerable knowledge and understanding of both the content of the data and of the legal framework

for the platforms where the data will be shared (in addition to knowledge of legal issues and frameworks connected to the sources - or the creation - of the data).

Although the body of laws at the EU level which directly affects legal interoperability within the context of the EOSC is limited (it includes intellectual property - in particular copyright and database rights - in addition to data protection and laws aimed at protecting sensitive or confidential data), the complexities inherent in these bodies of laws has the potential to adversely affect sharing and reuse opportunities [41, 46]. As a result, the complexities and uncertainties of legal issues connected to research data may – due to lack of competence, expertise, and resources – lead to research institutions and researchers being hesitant to make their data available for re-use. Where the expertise is lacking, identifying which data to share, and defining the scope and conditions for access and re-use may be perceived as a major challenge.

Exposing data through a federated service like the EOSC may, in the eyes of the data creators, increase the risk of copyright infringement and/or other improper, illegal, or unethical uses of data. This may possibly lead to reluctance exposing the data, which again can lead to less data to be exposed and shared through the EOSC platform.

In other words: where data can be accessed and is shared, data breaches are more likely to occur. In addition, inappropriate sharing of data can lead to significant costs to the organisation and/or individual researchers, including economic loss and loss of competitiveness and reputation. Also, sharing data too prematurely and/or inappropriately (e.g., with a wrong license) can undermine the ability to obtain IPR protection [46].

It may also be difficult for potential re-users of data to determine when certain data are subject to certain copyrights or other IPR-issues, and when data are open and can be utilised without any restrictions. The use of Creative Commons licenses (or other similar standardised international licenses) may reduce the uncertainty, by clearly stating what can and cannot be done with the licensed material. But the decision on setting the appropriate license is not an easy task. The owner of the data – the licensor – is responsible for deciding the license or the declaration of release to be used on a specific set of data. Before deciding on a license, it is important for the licensor to have clarified whether, and possibly to what extent, the research data are subject to legal rights such as copyright and database protection, and whether there are rights connected to a third party (specifically relevant when parts of the data come from a secondary source). But as mentioned above, in many cases neither the institutions nor the researcher has the legal knowledge and expertise necessary to make such clarifications and decisions.

Deciding on license when sharing research data can have both intended and unintended consequences, both for the licensor and the licensee. This is the main reason why it is important to have a legal understanding of the licenses and its terms of use when applied to research data. A possible consequence of not using a license when sharing research data, is that it may create uncertainty with regards to the legal status of the data. This may reduce the potential reuse of data since the data (re)user

may not be willing to take any legal risk [47]. A plan should be made on how to ensure licensing arrangements are in place, when sharing data through the EOSC.

However, even where individuals and organisations agree on and consent to specific terms/licenses for data sharing and data re-use, including the purposes for which the data should be re-used, there remains a significant level of risk that a third party may intentionally or unintentionally use the data differently.

Data access and sharing are about taking data from one context and transferring it to another context, and some experts have argued that this change of context may make it challenging to ensure that legal expectations connected to the initial usage of data are not undermined when data are being re-used by others [46]. In other words, information derived from data is context dependent and so are thus the risks associated with data re-use [46].

IPR issues in heterogenous data – an example

Another contextual aspect of data sharing is the fact that the understanding of the concept of 'research data' may be different in different contexts. Even though there seem to be a rather clear consensus on the basic elements of the definition of the concept (the EOSC defines research data as: "Data collected or produced in the course of scientific research activities and used as evidence in the research process, or commonly accepted in the research community as necessary to validate research findings and results"), there are several examples where the understanding of the concept is more ambiguous and where the distinction between intellectual property and the collected/captured data is unclear [48]. This may be particularly true in multi-disciplinary research where data comes from many different sources and consist of many different data types.

Within research disciplines that study artistic or cultural expression (for example language, literature, music, dance, theatre etc.), research data is often based on copyrighted material. In these instances, each 'data point' can be an independent intellectual property, protected by copyright. Examples can be linguistic corpora (scientific annotations of text or speech), or a musicology database that contains notes in addition to both audio and video recordings of various performances. In the musicology example, each single performance can have several different authors (composer, lyricist, performer, producer, choreographer, etc.), and it can be different organisations that manage the copyright holders' IPR and financial rights [47]. In cases like this, further sharing, re-use, and 'recalibration' of the research data is complex and difficult without infringing on the original rights.

This is just one example, but the problem the example aims to illustrate may be further enhanced in a multi-disciplinary, federated service environment like the EOSC. The ambitious 'capture-all' approach of EOSC may (unintentionally) create conditions where it is more challenging to protect the distinctive character of specific academic disciplines where the captured and created data are closely interconnected with copyrighted material.

Closely connected to this is the aspect of losing control: lack of control over data is perceived as a major issue for both organisations and individuals [46]. As mentioned earlier, when data is accessed or shared, the data will 'move' outside the context and system of the original data holder/owner and thus potentially move out of their control. This may be counteracted by quality controls and data stewardship or processing arrangements, but the risks of loss of control may nevertheless (be conceived as) stronger when the data are shared across multiple tiers (e.g., an open multi-disciplinary environment like the EOSC) and across multiple jurisdictions [46].

Big, cross-national, multi-disciplinary federated services like the EOSC may seem to be the solution to the 'multidimensional' problem of data sharing, but it is crucial that inherent in these kinds of system there are "flexible data-governance framework that take due account of the different types of data and the different context of their re-use, while doing justice to domain and cultural specificities" [46].

Input from experts

Selected experts on research ethics and IPR have been consulted on what kinds of IPR issues arise when research data from social science and humanities are to be shared through an open science cloud. The aim of these consultations was to gather input on IPR issues that can be linked to data sharing, especially to data sharing solution such as EOSC. The contribution was given by Arja Kuula-Luumi (hereinafter Kuula-Luumi) at the Finnish Social Science Data Archive, Paweł Kamocki at the CLARIN Legal and Ethical Issues Committee (hereinafter Kamocki), an expert in the SSH domain, Alain Strowel, Bertil Emrah Oder and Maciej Maryl at ALLEAs Permanent Working Group on IPR (hereinafter in this Chapter "ALLEA"), and Julie de Jong (hereinafter de Jong) at the University of Michigan and Chair of the ESS ERIC REB. In the following, the contribution provided is presented.

Overview of input received

The parties were asked on what kinds of IPR issues arise when research data from social science and humanities are to be shared through an open science cloud.

First, ALLEA emphasised that it is important to clarify applicable research organisations. Recital 12 of the 2019/790 Directive on copyright states that "*Research organisations across the Union encompass a wide variety of entities the primary goal of which is to conduct scientific research or to do so together with the provision of educational services. The term 'scientific research' within the meaning of this Directive should be understood to cover both the natural sciences and the human sciences. Due to the diversity of such entities, it is important to have a common understanding of research organisations. They should for example cover, in addition to universities or other higher education institutions and their libraries, also entities such as research institutes and hospitals that carry out research. Despite different legal forms and structures, research organisations in the Member States generally have in common that they act either on a not-for-profit basis or in the context of a public-interest mission recognised by the State. Such a public-interest mission could, for*

example, be reflected through public funding or through provisions in national laws or public contracts. Conversely, organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control because of structural situations, such as through their quality of shareholder or member, which could result in preferential access to the results of the research, should not be considered research organisations for the purposes of this Directive".

In addition, ALLEA mentioned that the IPR issues can be considered differently depending on the context of scientific research and the influence of private actors/commercial undertakings. For instance, and in the health sector, where a substantial amount of research is done through private funding and with private actors. Some of these projects might be considered as social research, such as for instance research on mental health and epidemiology.

According to de Jong, researchers have an ethnical responsibility to respondents when using their data, versus the responsibility to the researchers of whose intellectual property it was in the first place. This responsibility must be managed in each project. Similarly, Kuula-Luumi underlined that, as ethics and data protection, IPR issues must be resolved before data are to be shared. It was mentioned that there must be an agreement with external parties concerning the transfer of copyright and ownership rights, which include right to distribute the material. External parties may be individuals such as research subjects, or organisations and societies. Further, according to de Jong, it is important to determine agreements about the length of time between data collection and data archiving, to provide the researcher(s) responsible for data collection the first opportunity for data analyses but also providing others an opportunity for timely data access. She mentioned that in some cases, for instance during a pandemic, it is important that data to do with public health are released, as the world changes quickly and therefore the data can quickly become irrelevant.

ALLEA stressed that there are several applicable IPR issues. Confidentiality issues can arise and trade secrets or know-who are protected under a broad view of IPR. On this topic, they mentioned that more could be done to clarify the framework and provide for clearer exceptions for scientific research, if well defined. In addition, corporate secrecy can be a barrier for scientific research. For instance, the inner working of algorithms such as those used by online platforms, and which arguably are promoting disinformation as can be read about in EDPBs preliminary opinion on data protection and scientific research, page 9 [49]. Further, ALLEA stated that confidentiality issues are often linked to data protection issues. For instance, if a dataset consists of survey material, interviews or an oral history collection, possibly with vulnerable individuals. According to ALLEA, it can be considered that IPR here include all rights that allow to control data, and data protection regulations lies within this category. In addition, they mentioned that further clarifications are needed concerning this topic and referred to in EDPBs preliminary opinion on data protection and scientific research [49].

Similarly, according to Kamocki, for SSH data, the most important intellectual property framework is copyright. He emphasised that several issues must be observed to comply with copyright framework. For instance, if the use is covered by a statutory exception (especially the new TDM exception for research

organisations), or if permission from rightsholders is necessary. Further, Kamocki mentions, if such permission is necessary, relevant issues to address are: how to reach those rightsholders, and can one still use the data under the 'orphan works' framework if they cannot be contacted. Also, if some sort of permission has already been granted in a license, one must assess if that specific use will fit within the permission. In addition, it will be relevant to assess if authorship properly has been attributed. Moreover, Kamocki mentions that the sui generis database right, or (in some countries) the rights of editors of scientific and critical editions, or the rights in unoriginal photographs, should not be ignored, as they also may affect datasets SSH. An expert in the SSH domain underscored the need for contractual issues to be aligned with IPR arrangements.

According to ALLEA, when addressing copyright issues, the focus should be on text and data mining in general, they expressed support to the exception for TDM for scientific organisations. In addition, it was mentioned that the costs of making the data of SSH research available under the FAIR principles should also be considered, as funding for such research can be lower and thus less resources are available to clean-up the data, and it can include complex data as personal data can be mixed with non-personal research data. It was underlined that this will differ from all approaches, and more tailored proposals are needed considering the specificities of various types of SSH research.

Summary of expert views

In summary, the parties consulted acknowledged that there exists IPR issues when research data from social science and humanities are to be shared through an open science cloud, and the underlying premise seems to be that the IPR issues can be considered differently depending on the context of scientific research, the influence of private actors/commercial undertakings and at which step the issue is addressed during the research process. ALLEA mentioned, among others, issues with confidentiality, trade secrets and corporate secrecy. Kuula-Luumi stressed that IPR issues must be resolved before data are to be shared, and de Jong stated that the responsibility must be managed in each project. It was also mentioned that there must be an agreement with external parties concerning the transfer of copyright and ownership rights, which include right to distribute the material.

3.3 Contractual law

Participation in research is voluntary and, when a researcher is in direct contact with a research subject, consent is often gathered from research subjects. This consent can be gathered to document an ethical consent, thereby making sure to document that participation in the project is voluntary. It can also be gathered as part of a legal basis for processing personal data, as consent/explicit consent is one of the applicable legal bases in GDPR article 6 and 9. In other cases, it can be given as part of repealing the duty of confidentiality. Even though some differences, a common identifier for these consents is that it represents an agreement or contract between the research subject and a researcher, on how their role will be performed and how data gathered in the project will be used during the project, and after the project.

Therefore, how data gathered in some projects can be used during the project period, and after, might depend on the content of the information provided to the research subject before the contract was signed. If data is used for different purposes and shared with external parties not included in the information before the parties entered the agreement, it can be considered as a breach of the contractual obligations. The same can be argued if the data are processed for a longer period than agreed between the parties. In the following, the report addresses some implications this might have for EOSC.

First, if research subjects have been informed that all data will be deleted after a project period, it can be argued that the researcher will, based on the content of the agreement, be hindered from storing both personal data and/or an anonymised dataset. This will have implications for the EOSC, as fewer data can be stored in the cloud and shared with others in the SSH Environment.

Second, in some cases research subjects might have been informed that no data will be shared with external parties. It can therefore be argued that the researcher will, based on the content of the agreement, be hindered from storing and/or making data available in a repository or other data services that are connected to the EOSC.

On the other hand, if the agreement did not regulate the possibility of sharing these data with external parties, a question can be raised if it will be considered appropriate to share/store these with the EOSC and further, to share these as part of open access. This can, as above, also mean that fewer data can be stored and through data repositories and other services that are connected to the EOSC and shared with others in the SSH Environment.

If the planned procedure for storing and sharing data given from a participant, deviates from the initial agreement made with the researcher, it can be argued that new information should be presented to the participant. In some cases, it might also be necessary to gather a new/updated consent. The necessity of this should be subject to a concrete assessment performed by the responsible researchers in all cases. If the researcher at the given time is not able to re-contact the participant, this might not be possible and it must be decided whether data can be shared with the EOSC, and others. This responsibility is likely to lay with the researcher, as the other agreement party. However, EOSC should also have a plan on how to verify which data can be indexed, discovered and shared through its cloud, to act in accordance with agreements made with participants.

Further, the contract between a researcher and a research subject can also regulate how long the data are to be stored, and for which purposes. If storing data in a trusted repository (that eventually will be connected to the EOSC) is in accordance with the agreement between a researcher and a participant, the agreement can regulate how long the data can be stored until deleted. This can have implications for the EOSC, as some data may need to be deleted after a time period.

The content of contracts can therefore have implications for the EOSC, as it can affect which data that can be indexed, how long it can be stored, and possibly shared with others, in the cloud. It will be important to determine whose responsibility is it to resolve the issues addressed, and that it will be

helpful to have a mechanism able to verify that legal and ethical issues are resolved, before storing data in the EOSC.

4 Conclusion

Promoting data sharing in the scientific community is important and has become an essential practice, whether between different groups within the same organisation, between partners in larger platform endeavours, or even, as in growing open data movements, with the public. Data sharing helps in the advancement of science in small increments rather than through single studies, and ensures replicability and thereby, helps confirm a study's findings. Data sharing enables new insights from existing data, and lets organisations make full use of this core resource. However, it also introduces new ethical and legal issues. The report identifies some ethical and legal issues with potential implications for the EOSC.

Related to the legal issues, the report finds that the content of contracts between a researcher and a research subject can have implications for the EOSC, as it can affect which data can be stored, how long it can be stored, and possibly shared with others. Further, the report finds that sharing of research data on cloud-based services like the EOSC can lead to some IPR issues. When consulting external parties, the IPR issues can be considered differently depending on the context of scientific research, the influence of private actors/commercial undertakings and must be resolved before data are shared and managed in each project. There should be an agreement with external parties concerning the transfer of copyright and ownership rights, which include right to distribute the material.

Further, related to identifying ethical issues, the report finds that questions can be raised on the risk of the research subjects losing control over their personal data. As the number of research projects increases, keeping track of where, when, and by whom your data is being processed, might seem like an insurmountable task. This applies even if, or perhaps precisely because, research subjects receive information concerning each project that includes their data.

In addition, the report finds, also related to the first point of ethical issues, that there is a risk that data might be used for commercial interests. Unless specified, future re-use of data collected for a certain (wide) purpose might gradually lead to a change in what purposes the data is used. There is not a clear line separating academic research and commercial interests, but it would be in the interest of EOSC to draw a line between accepted purposes for re-use, and the required academic affiliations for researchers seeking to apply for data.

In addition, the report finds that sharing data that is potentially identifiable can increase the risk of harm for the participants. This is perhaps most dire in cases where the research subjects are ethnic minorities. The research subjects from an ethnic minority might be more vulnerable for re-identification, and such re-identification might entail a greater risk to their safety. It can be argued that the ethical issues may arise from sharing these kinds of sensitive data.

It can be of benefit for the EOSC, to have routines that ensures possible ethical and legal issues in each case have been addressed before data are stored in the EOSC.

One issue that must be settled by the EOSC is where the responsibility for securing safe and ethical data collection lies. Experts consulted emphasised the individual researcher's responsibility in this matter. This report therefore recommends that EOSC take great care to formulate and publish precise and consistent requirements for how the data it will accept should be collected and processed by researchers who gather it.

The responsibility to communicate this information to the individual research subjects of a research project must fall to the researcher. Here, the EOSC must inevitably place trust in the researchers' and research organisations' ability to sufficiently inform the research subjects of the future use and re-use of the data. This may no doubt be challenging. As a participant, to fully comprehend the life cycle of research data, from collection, to publishing, and future re-use, may be difficult even if the data is kept solely at one institution. If the data is to be shared and re-used among numerous European institutions, it may seem dazzling. The EOSC should provide well-ordered and searchable information about each project which uses data from the infrastructure.

Another issue raised by experts is the potential problems that can occur when an international project must seek ethical approvals from several national ethical committees. Though different ethical committees must obviously abide by their national rules, it could be a worthwhile task for the EOSC to aid the researchers in order to help them navigate these waters, perhaps in the form of an ethics task force. A Code of Conduct, regulating specific topics and presenting guidelines, could perhaps be considered helpful to resolve such an issue.

5 References

1. EOSC. *The European Open Science Cloud (EOSC) is an environment for hosting and processing research data to support EU science*. 2021 [cited 2021 October]; Available from: https://eosc-portal.eu/about/eosc.

2. EOSC. *What the European Open Science Cloud is*. 2022 [cited 2022 Feb]; Available from: https://www.eosc.eu/about-eosc.

3. UiB. *Tilgjengeliggjøring og deling av forskningsdata. Oppfølging av nasjonal strategi. Anbefalinger fra arbeidsgruppe*. 2017 [cited 2021 October]; Available from: https://www.uib.no/sites/w3.uib.no/files/attachments/4001_tilgjengeliggjoring_-_arbeidsgruppe_0.pdf.

4. Maurizio Sanesi, Lea Sztuk Haahr, Mathilde Steinsvåg Hansen, Ina Nepstad, Belinda Gloppen Helle, Marianne Høgetveit Myhren, Marita Ådnanes Helleland, Tore Andre Kjetland, Ingvild Eide Graff, & Vigdis Kvalheim. (2021). D5.7 Report on the impact of the GDPR and its implications for EOSC (1.0). Zenodo. https://doi.org/10.5281/zenodo.4723645.

5. Ina Nepstad, Inga Brautaset, Mathilde Steinsvåg Hansen, Tore A. K. Fjeldsbø, Siri Tenden, Marita Ådnanes Helleland, Christopher Ongre Autzen, Ingvild Eide Graff, Marianne Høgetveit Myhren, & Vigdis Namtvedt Kvalheim. (2021). 5.8 Draft SSH GDPR Code of Conduct (v1.0). Zenodo. https://doi.org/10.5281/zenodo.5181223.

6. Fisher, Lars, Hammargren, Per-Olov, Iozzi, Maria Fancesca, Kontkanen, Pirjo, Sarti, Jarmo, & Gunnarsdóttir, Birna. (2021). D2.7: Open Science in the Nordics: Recommendations on Legal Issues. Zenodo. https://doi.org/10.5281/zenodo.5713297.

7. Axel Börsch-Supan, Carolina Brändle, Johanna Bristle, Fabio Franzese, Daniel Schmidutz, & Luzia Weiss. (2021). D5.1 Guidelines for ethics considerations in making biomedical survey data FAIR (Access to biomedical data) (v1.0). Zenodo. https://doi.org/10.5281/zenodo.47851988. Franzese, F., *D5.3 Data access protocol for accelerometer data, linked to survey data, conforming FAIR principles (Access to biomedical data)*. 2021.

9. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship.* Sci Data, 2016. **3**: p. 160018.

10. *Danish National Forum for Research Data Management. How To FAIR*. 2021 [cited 2021 November]; Available from: https://howtofair.dk/what-is-fair/#fair-principles.

11. *Go FAIR. FAIR Principles*. 2021 [cited 2021 November]; Available from: https://www.go-fair.org/fair-principles/.

12. Jacobsen, A., et al., *FAIR Principles: Interpretations and Implementation Considerations.* Data Intelligence, 2020. **2**(1-2): p. 10-29.

13. *The Norwegian National Reseach Ethics Committees. The Nuremberg Code*. 2020 [cited 2021 October]; Available from: https://www.forskningsetikk.no/en/resources/the-research-ethics-library/legal-statutes-and-guidelines/the-nuremberg-code/.

14. *The World Medical Association, Inc. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*. 2018 [cited 2021 November]; Available from:

https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.

15. SERISS, *Deliverable Number D6.10. WP6. Synopsis of Policy - Rules for Collecting Biomarkers in Social Surveys*. 2016. p. 1-76.

16. ALLEA. *The European Code of Conduct for Research Integrity Revised Edition*. 2017 [cited 2021 October]; 1-20]. Available from: https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf.

17. Steve Canham, et al., *White Paper 4: Ethics Supporting Document to D3.3 Draft Policy Recommendations*, in *EOSC Pilot*. 2018.

18. Munafo, M.R., et al., *A manifesto for reproducible science.* Nat Hum Behav, 2017. **1**: p. 0021.

19. *Berlin Declaration. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. 2003 [cited 2021 October]; Available from: https://openaccess.mpg.de/Berlin-Declaration.

20. Moonesinghe, R., M.J. Khoury, and A.C.J.W. Janssens, *Most Published Research Findings Are False—But a Little Replication Goes a Long Way.* PLOS medicine, 2007. **4**(28): p. 0218-0221.

21. Baker, M., *1,500 scientists lift the lid on reproducibility.* Nature, 2016. **533**(7604): p. 452-4.

22. Camerer, C.F., et al., *Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015.* Nature Human Behaviour, 2018. **2**(9): p. 637-644.

23. Renkewitz, F. and M. Heene, *The Replication Crisis and Open Science in Psychology.* Zeitschrift für Psychologie, 2019. **277**: p. 233–236.

24. Maxwell, S.E., M.Y. Lau, and G.S. Howard, *Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?* American Psychologist, 2015. **70**(6): p. 487–498.

25. Button, K.S., et al., *Power failure: why small sample size undermines the reliability of neuroscience.* Nature Reviews Neuroscience, 2013. **14**(5): p. 365-376.

26. West, R., *Data and statistical commands should be routinely disclosed in order to promote greater transparency and accountability in clinical and behavioral research.* Journal of Clinical Epidemiology, 2015.

27. Houtkoop, B.L., et al., *Data Sharing in Psychology: A Survey on Barriers and Preconditions.* Advances in Methods and Practices in Psychological Science, 2018. **1**(1): p. 70-85.

28. Iina Kohonen, Arja Kuula-Luumi, and S.-K. Spoof, *The ethical principles of research with human participants and ethical review in the human sciences in Finland. Finnish National Board on Research Integrity TENK guidelines 2019*. 2019.

29. GDPR, *Art. 24. Responsibility of the controller*. 2018.

30. GDPR, *Recital 75. Risks to the Rights and Freedoms of Natural Persons**. 2018.

31. GDPR, *General Data Protection Regulation. Chapter 2 Principles*. 2018.

32. GDPR, *Art. 1. Subject-matter and objectives*. 2018.

33. European Commision Directorate - General for Research & Innovation. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Version 3.0.* 2016. p. 1-12.

34. GDPR, *Recital 33 Consent to Certain Areas of Scientific Research*. 2018.

35. GDPR, *Art. 5. Principles relating to processing of personal data*. 2018.

36. Sametinget. *Etiske retningslinjer for samisk helseforskning*. 2021 [cited 2021 November]; Available from: https://sametinget.no/_f/p1/iab4f8a7f-1c05-4f9f-a5e2-1d4f6196671d/etiske-retningslinjer-for-samisk-helseforskning-ny.pdf.

37.   UNSD. *Fundamental Principles of National Official Statistics. The United Nations Statistics Division.* 2021 [cited 2021 November]; Available from: https://unstats.un.org/fpos/.

38.   SHARE. *SHARE - Survey of Health, Ageing and Retirement in Europe*. 2022 [cited 2022 Feb]; Available from: http://www.share-project.org/home0.html.

39.   *Norwegian Customs. Intellectual Property Rights (IPR)*. 2021 [cited 2021 November]; Available from: https://www.toll.no/en/corporate/intellectual-property-rights-ipr/.

40.   *European Commission. European data strategy - Making the EU a role model for a society empowered by data.* 2020 [cited 20201 November]; Available from: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

41.   OECD, *Data-Driven Innovation*. 2015.

42.   Doldirina, C., et al., *Legal Approaches for Open Access to Research Data.* OSF, 2018.

43.   RDA-CODATA Legal Interoperability Interest Group. *Legal Interoperability of Research Data: Principles and Implementation Guidelines*. 2016   [cited 2021 October]; Available from: http://www.codata.org/uploads/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines_Final2.pdf.

44.   *Creative Commons. CC0 "No Rights Reserved"*. 2021 [cited 2021 October]; Available from: https://creativecommons.org/share-your-work/public-domain/cc0/.

45.   Graber-Soudry, O., et al., *Legal Interoperability and the FAIR Data Principles.* Zendo, 2020: p. 1-77.

46.   OECD, *Enhancing Access to and Sharing of Data*. 2019.

47.   Forskningsrådet, *Hvordan skal vi dele forskningsdata? Utredning og anbefalinger om lisensiering og tilgjengeliggjøring*. 2021. p. 1-67.

48.   EOSC. *EOSC Glossary*. 2021   [cited 2021 November]; Available from: https://eosc-portal.eu/glossary.

49.   The European Data Protection Supervisor (EDPS). *A Preliminary Opinion on data protection and scientific research*. 2020. p. 1-35.