White Paper 4: Ethics

Supporting Document to D3.3 Draft Policy Recommendations

Author(s)	*The ECRIN writing group:* **Steve Canham, Christian Ohmann, Mihaela Matei, Serena Battaglia** *Expert group:* **Dr Sarion Bowers**, *Research Policy Lead, Wellcome Trust Sanger Institute, UK* **Mrs. Prof Dr Christiane Druml**, *Vice-Rector of the University, Chairperson of the Austrian Bioethics Commission, UNESCO Chair of Bioethics at the Medical University of Vienna;* **Prof Dr Mats G. Hansson**, Professor of Biomedical Ethics, Director, Centre for Research Ethics and *Bioethics at Uppsala University, Sweden.* **Dr Michaela Mayrhofer**, *Chief Policy Officer CS ELSI, BBMRI* **Dr iur Fruzsina Molnár-Gábor**, *Heidelberger Akademie der Wissenschaften* **Dr Lorna Ryan**, *Research Manager, Department of Sociology; City University London, UK*
Status	Final
Version	v0.3.2
Date	28/08/2018

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

The European Open Science Cloud for Research pilot project (EOSCpilot) is funded by the European Commission, DG Research & Innovation under contract no. 739563

Versioning and contribution history

Version	Date	Authors	Notes
	Nov 2017 - Jan 2018	ECRIN writing group	Initial work by the ECRIN team (interviews, research of existing initiatives and documents)
	01/02/2018	Expert group and ECRIN writing group	Telephone conference with senior ethics experts
1.1	18/02/2018	ECRIN writing group	First draft
1.1	To 29/03/2018	Expert group	Written feedback by the senior ethical experts
2.0	20/03/2018	ECRIN writing group	Revised document
2.0		Expert group	Further feedback
2.4	09/04/2018	ECRIN writing group	Successive revisions, to final version inserted into D3.3
0.1	25/04/2018	Dale Robertson	White paper format created
0.1	25/04/2018	ECRIN writing group	Transfer of material from D3.3 to White Paper format; version number reset to reflect new document.
0.2	07/05/2018	ECRIN writing group	Addition of Impacts table.
0.3	26/06/2018	Dale Robertson, Steve Canham	Final edits.

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

This paper discusses how an ethical dimension can best be included within the policies, structures and services of the developing European Open Science Cloud (EOSC).

Although the inclusion of ethical principles and policies are seen as of fundamental importance to the EOSC, it is difficult to anticipate all the ethical issues that may emerge as the scientific, technical, social and political landscape evolves. It is therefore seen as crucially important to have governance mechanisms in place that can ensure ethical issues are appropriately dealt with in the future, however and whenever they are presented, as well as identifying and proposing responses to current issues.

An analysis of ethical issues relating to organisational conduct and policies, research conduct, research decision making, the use of data, especially sensitive personal data, and the interaction between science and society, serves to underline the complexity and diversity of potential issues.

Because of this complexity, the paper proposes that ethical issues can be managed at a variety of different 'layers' of involvement and commitment. It proposes that *as a minimum*:

- EOSC as an organisation should commit to act, and be seen to act, in an ethical manner, with policies and processes that reflect that commitment,
- EOSC should have structural mechanisms in place to support research integrity, for instance by establishing metadata systems that ensure accurate provenance data and appropriate acknowledgement of previous work.

Many other issues will need specialist ethical (and often legal) expertise, in time limited working groups, to propose specific policy responses. We strongly support the use of such groups, but we also *strongly recommend*

- the establishment of a coordinating body, **an Ethical and Legal Advisory Board**, to oversee the work of specialist working groups, and to monitor and report on the ethical practice of the organisation.

Such a board would be distinct from the EOSC executive, and be drawn from ethical and legal experts from stakeholder communities.

We believe 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. We believe such input could be introduced as the EOSC itself matures, and the level of available resources becomes clearer.

Key next steps include wide consultation on these proposals, with a variety of stakeholders, plus exploration of how the recommendations above can be integrated into the 'policy supporting services' that now need to be developed.

1. INTRODUCTION

This paper discusses how an ethical dimension can best be included within the policies, structures and services of the developing European Open Science Cloud (EOSC). It is designed to support the policy recommendations found in Deliverable D3.3, by providing a more detailed explanation and justification for the proposals made, linking the discussion as appropriate to the existing literature.

After first clarifying the main drivers and constraints surrounding this work, the paper identifies both ethical areas and examples of specific ethical issues that need to be considered. It then provides some suggestions as to how these might be addressed within a future EOSC structure, in a discussion of policy options, makes some specific proposals, and ends with some suggestions for disseminating and testing those proposals. A summary table describes the possible implications of the recommendations for various stakeholders.

To help give some context to this document, a short annex provides some background on the EOSC and open science. A glossary is also provided, clarifying definitions of personal and sensitive data, as those terms are used within this document.

1.1. Drivers and Constraints

The principal driver for this activity is the need to identify the possible ethical issues within EOSC functions and processes, and ensure that those issues are considered and managed in the most appropriate way, for the benefit of EOSC itself, its scientific user communities, and the wider European public.

Although 'ethics' is rarely explicitly identified as a topic in current EOSC documents, it is implicit in many of the statements being made about the science cloud. For instance, the 'EOSC declaration' (1), includes phrases such as: 'data stewardship', 'duly justified … proportionate limitations' (to access), 'confidentiality', 'minimal and rigorous global standards', and 'trust and increasing mutuality'. Clarifying what these words mean in practice requires an analysis, not just of stakeholder views and the applicable legislation, but also of the ethical concepts underlying each term, and their application to 'open science'.

An initial difficulty, however, is that the term 'open science' is neither precisely defined nor uniformly applied. As Fecher and Friesike point out, 'open science' is an "umbrella term encompassing a multitude of assumptions about the future of knowledge creation and dissemination" (2). They distinguish five distinct 'schools of thought' of open science – which they characterise as democratic, pragmatic, infrastructure, public and measurement aspects. Different scientific disciplines have assimilated these various aspects of open science to different degrees.

For example, in genomics open science has been encouraged and applauded for nearly 20 years (3). In other fields, such as clinical research, 'open science' is a relatively new concept, and the implications of working within that paradigm are still being discussed (e.g. 4). The situation is, however, more complex than these simple statements would imply. Within genomics, the need for controlled access in some situations is also recognised (5), and open science did not, for example, prevent the outbreak of recent 'patent wars' over the CRISPR / Cas9 procedure (6). Conversely, within clinical research some initiatives already share individual participant data from clinical trials (e.g. 7). Being aware of the various practices in different disciplines, as well as some of the current challenges to open science, will be essential to framing any debate on ethics within the EOSC.

In addition, the volume, types and sensitivity of data stored within EOSC are likely to expand rapidly in future years (e.g. with the storage of many more individual genomes, or the greater use of routine clinical data for research purposes), as will the tools available to process that data, particularly using algorithmic approaches or artificial intelligence (AI). The legal and social context of EOSC activity will inevitably evolve – for instance with changing ideas about what is an acceptable level of privacy. New ethical issues and concerns, or new variations of old questions, are therefore likely to emerge as data, technologies, and context develop. Legal and regulatory constraints will necessarily frame the governance and management of the EOSC, but these too are likely to evolve, or be subject to re-interpretation, over time. It is as important, therefore, to propose

structures and processes within EOSC that can handle new ethical challenges, as they arise, as it is to try and provide definitive answers to those that exist now.

Unfortunately, some of the details of EOSC are still rather imprecise and, in particular, there is no clear indication from the developing governance structures about how far ethical and other policy requirements will influence or restrict the management of EOSC processes. There are statements referring to planned 'Principles of Engagement', saying that "all the component systems, service providers and the services they contribute to EOSC should adhere to the regulations and practices established by these principles" (8). These principles are still being defined, however, and currently seem to be largely concerned with technical issues. At the same time we have statements declaring that EOSC needs to be a 'light touch' federation of facilities and services, without any direct control or ownership of data (9). The balance between 'adherence to principles' and a 'light touch' is yet to be determined.

As the following section makes clear, there are also a multitude of potential ethical issues that could be considered. Many of these would more normally be seen as the responsibility of the individual researcher, (or their employer or funder), rather than the concern of EOSC as an infrastructure organisation. This does not prevent EOSC from supporting and promoting ethical behaviours from other actors, but the extent to which the organisation will want to play this role has yet to be decided. The exact scope of 'EOSC ethics', as a subset of all those possible, therefore remains unclear.

Despite these uncertainties, the discussion that follows is based on the conviction that ethical considerations should be an integral part of EOSC decisions and processes from the outset, and not seen as somehow additional or peripheral to the core scientific endeavour. This belief is based both on a moral argument – EOSC has a duty to act in an 'ethical way' and promote ethical behaviour in others – and the more pragmatic reason that this is the only way for the organisation to build and retain the trust it needs, of its users and funders.

The document therefore focuses first on the range and types of ethical issues that will need to be considered by EOSC. Specific issues are given as examples but there is no attempt to rehearse the arguments surrounding them in detail. The document then proposes a mechanism for building policies and structures in EOSC that can ensure ethical considerations are fully but flexibly integrated within the work of the science cloud.

2. METHODOLOGY

The methodology employed by the task group was essentially that of a dialogue between the ECRIN writing team and the group of senior ethical experts who were recruited to provide specialist expertise, as shown by the version and contribution history.

After initial information gathering and discussion, the ECRIN team produced a first draft document and set of proposals. From the outset, we tried to ensure that debates around ethical issues considered the whole of the European Open Science Cloud, its data and its user communities, and was not just concerned with the processing of personal data, important as that is. This approach was reflected in the initial draft.

The expert group provided initial feedback and comment, which allowed the ECRIN writing group to expand their arguments and put them on a firmer basis with respect to the literature. In addition, a summary of an early draft of this document was presented by Michaela Mayrhofer (BBMRI) at a WP3 Policy Workshop during the EOSCpilot plenary meeting. Feedback was collected by discussion and by presenting prepared questions to the audience (see: EOSCpilot Plenary Meeting, Pisa, WP3 Policy Workshop, 8 March 2018, Notes).

The consultation process was repeated, producing a series of iterative improvements, until the current document was created. Because the resulting text had grown, (in this and the other sub-tasks), the decision was taken to move the bulk of the supporting arguments into a separate white paper, leaving the policy recommendations as the main component of the D3.3 deliverable.

3. ETHICAL AREAS AND ISSUES

The European Open Science Cloud is designed to support scientific activity as well as data storage. The services and system it offers will therefore be involved in the initial collecting, processing and analysis of data, as well as storing, sharing, re-analysing, and curating it in the longer term, and the provision of appropriate metadata.

As figure 1 indicates, this means that the possible ethical 'areas' to be considered include both 'Research ethics' and 'Data ethics', though as the figure shows there is substantial overlap between the two. Within both of those two main areas, and their overlap, are the ethical concerns about the use of personal data and sensitive data, and in particular data that is both personal and sensitive (please see the glossary for formal definitions of these terms).

In addition, there is a set of ethical issues that might be called 'organisational ethics', concerned with how EOSC, (or any other organisation), is governed and conforms to ideas of good practice, and a collection of issues that are labelled here as 'science and society' ethics, referring to the ways in which science and the larger society (the press, social media, governments) can influence each other, and how the benefits of that interaction can be maximised. The issues in each of these areas are considered in turn.

3.1. Organisational Ethics

EOSC will be a federation of different organisations, each with their own ethical practices and obligations, reflecting their own regulatory environments and the activities in which they are engaged. But EOSC needs to consider how its own governance, and its constituent organisations when they are acting as or for EOSC, can demonstrate that they are upholding some basic ethical principles. Christopher Bennett's conception of ethical behaviour as actions that "can be defended under scrutiny" (10) point to what those principles could be – i.e. they should ensure that EOSC always acts in ways that it can readily justify, to both its own stakeholders and to wider society.

It is suggested that this means insisting on transparency, with strategy and decisions documented and public. It means honesty, including disclosure on financial issues and data usage, so that there is no suspicion of hiding possible conflicts of interest. If Google (for instance) wished to give the EOSC large sums of money in return for access to some of its data, then the whole exercise needs to be transparent – and not, as was the recent case of Google's interactions with the Royal Free Hospital in London, hidden from public and patients alike (11). It should include fairness, in making resources available to different users, at defensible costs, and it means conforming to modern conceptions of fair and non-discriminatory employment practice.

It should also mean, as discussed below, putting into place systems that support and incentivise the research integrity of individual researchers, and demonstrating a commitment to periodic ethical inspection and oversight by an independent body of experts, acting as an advisory board.

3.2. Research Ethics

Following Drenth (12), we split research ethics into the ethical principles underpinning good practice by the individual researcher, often referred to as *research integrity*, and the wider 'external' aspects of research. Research integrity has been described as concerned with:

- Reliability in ensuring the quality of research (by appropriate design, methodology, analysis and the use of resources).
- Honesty, in developing, reporting and communicating research in a transparent, fair, full way.
- Respect for colleagues, participants, society, ecosystems, heritage and the environment.
- Accountability, for the research from idea to publication, for its management and organisation, and for training and supervision,

while commonly quoted violations of integrity include fabrication and falsification of results, including manipulating processes as well as adding, changing, or suppressing data, and plagiarism (12).

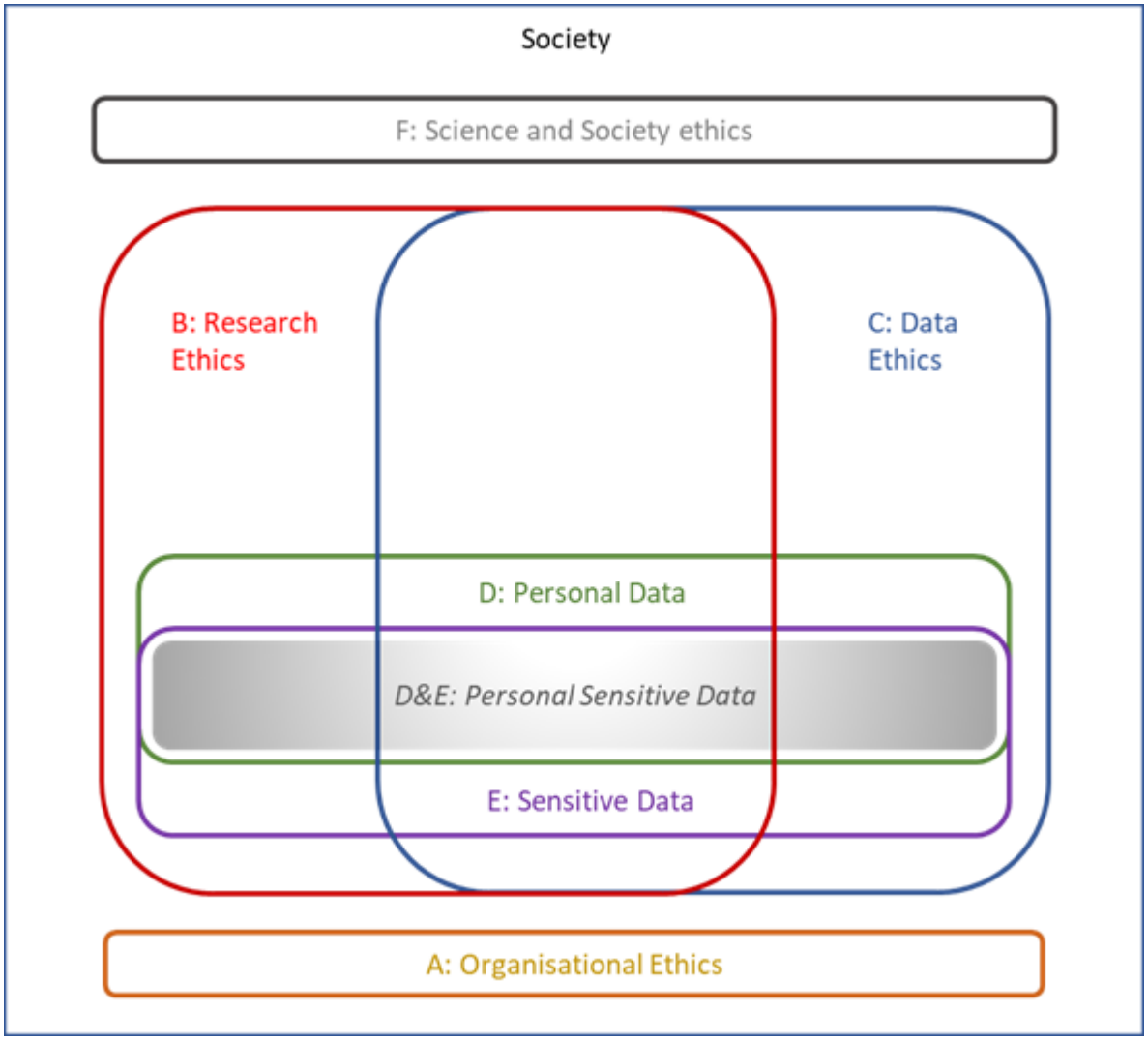

Figure 1 - A Venn diagram representation of the different types of ethical issues that need to be considered

Research integrity is supported by a large number of codes of conduct. Generic codes have been produced by both international organisations (e.g. 13, 14, 15), and national associations, universities and funders (e.g. 16, 17, 18), but more specialist codes, focused on specific fields of research, have also been created. For instance, the European biobank infrastructure BBMRI is currently chairing development of a Code of Conduct for Health Research, with particular reference to compliance with the new GDPR data protection legislation (19), while the RESPECT code of practice was drawn up for socio-economic research, and is structured around the triple requirements of upholding scientific standards, compliance with the law and the avoidance of social and personal harm (20). Codes dealing specifically with cloud services include the Code of Conduct for GPR Compliance from the Cloud Security Alliance (21), focusing mainly on legal requirements, and the Code of Conduct for Cloud Infrastructure Service Providers by CISPE (22) that covers privacy and data protection transparency, assurance and compliance as well as a certification scheme.

There are also structural mechanisms to enhance research integrity, especially in areas seen as being high-risk. The best known is the use of research ethics committees to review proposed clinical research, but review boards, ombudspersons' offices, and research integrity offices are also available in Europe for monitoring research ethics (23).

There would be little point in EOSC trying to add to the various codes of conduct and ethical review mechanisms supporting research integrity. That should not prevent, however, EOSC establishing policies and structures that promote research integrity, in particular by insisting on standards of metadata for stored materials. For instance, to make potential conflicts of interest as clear as possible, metadata should include conflict of interest statements, as found in published papers now, and identify not just the individuals that created the materials but also their organisational affiliations. If papers are retracted for any reason, then associated materials also need to be retracted, or the metadata should clearly indicate when and why a retraction of the associated paper has occurred.

Having clear policies about metadata, especially this core 'provenance metadata', which is common to almost all types of scientific data and data objects, and enforcing those policies to ensure consistency, therefore has an ethical dimension (as well as simply being a more efficient way of organising the metadata). Metadata policies should also include the need for full citation, and thus credit, for the originators of datasets that are re-used by other investigators. Insisting on this, and monitoring compliance, would be one way in which EOSC could encourage researchers to make their data available to others, especially in clinical research and other domains where access to data is traditionally controlled.

The type of topics included within the 'wider conception' of research ethics include

- The justification of the choice of research topic. Is what is being investigated worth knowing?
- The independence, or not, of the research (from clients, interested parties, sponsors)
- The extent of the researcher's responsibility for what is done with his or her results
- The possible need for a 'no go' or 'go slow' decision on certain types of research, because of the possible consequences of that research

These are very much issues for research governance and funders to consider, as well as individual researchers, and many are amenable to enquiry followed by collective decision making. For example the match between research activity and disease burden has been explored as a research topic (e.g. 24, 25), and regulations and constraints on genetic experiments in humans have often come from the research community itself (e.g. 26).

Whether and how EOSC should be involved with these questions requires further debate. It could, for instance, identify and / or support courses for researchers in research ethics, to increase awareness and possible involvement. It could provide data on usage and materials stored that would help to inform the debates on research activity. It could (probably should) monitor and feedback usage data to the generators of materials, so that researchers are kept fully aware of who was using their data and for what projects.

Although EOSC is unlikely to lead these sorts of ethical debates, it does need to consider how it can support them, including providing empirical data to inform discussions and decisions, by monitoring and reporting on the activity within the science cloud.

3.3. Data Ethics

The term 'Data ethics' was coined by Floridi, who defined it as

> "a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes)"… (27)

It is related to the earlier 'computer ethics', and then 'information ethics' or 'digital ethics', but represents a change of emphasis that reflects the growing concern about the pervasiveness of data in our societies. The 2015 report of the European Data Protection Supervisor (28) summarises some of the elements driving this increased role for data. They include

- Growing accumulations and analysis of 'big data'
- The 'internet of things' (beyond ATMs and POS machines to domestic appliances)
- Ambient computing (e.g. voice activated speakers, heating systems)
- Cloud computing (as a key enabling technology)
- Personal data-dependent business models (e.g. Facebook)
- Robots, drones and autonomous vehicles
- 3D printing
- Artificial intelligence

And in the context of EOSC we should add…

- Open Science

These developments are generating rapid change now, but in fact discussion about the possible impact of information systems on society began in the early 1950s with Norbert Wiener (the 'father of cybernetics'), whilst concern over the behaviour of robots started with Isaac Asimov's fictional 'three laws of robotics', in 1942. Both these men were visionaries, however, and it was not until the 1980s that these topics were debated more widely, with specialist journals established in the late 1990s, followed by an explosion of interest from the turn of the century, as everyone has become increasingly aware of the power of data driven systems (29).

Data ethics includes a variety of thematic issues. For instance, making data FAIR (findable, accessible, interoperable, reusable) (30) is a data management principle, and it is one that has been rapidly and widely adopted, including by EOSC (something that will have obvious impact on the need to generate and maintain metadata within the science cloud). But the A in FAIR denotes 'accessible under well-defined conditions'. As Mons et al point out:

> "The FAIR principles, although inspired by Open Science, explicitly and deliberately do not address moral and ethical issues pertaining to the openness of data. … the degree to which any piece of data is available, or even advertised as being available (via its metadata) is entirely at the discretion of the data owner. FAIR only speaks to the need to describe a process – mechanised or manual – for accessing discovered data…" (31)

EOSC, as a 'FAIR based' organisation, will therefore need to clarify its expectations regarding access, of different types of material in different domains, (i.e. establish its own ethical principles around data access) and then develop data management strategies for supporting those mechanisms. For example, it may be that researchers who do not allow open access should always provide a public justification for their decision, indicate if access is possible but on demand (and describe how any demands will be processed), and the length of time controlled access arrangements are likely to remain in place. The details are likely to vary between different scientific domains, but the underlying principles should remain the same across the organisation.

Another issue that EOSC will need to grapple with might be called the 'law of unintended data consequences'. The nature of a large federated set of data generators and data repositories means that data and other materials may be combined in unpredicted ways to generate unforeseen results – indeed that is seen as an intrinsic part of 'open science' and part of the rationale for developing the system. But should EOSC monitor and manage this type of data aggregation, so far as it can, and even restrict it in some cases?

Should, for example, there be controls over opening EOSC based data to commercial companies who can combine it with their own 'big data' to provide better targeted marketing information. The need to protect groups, rather than individuals, from stigmatisation and political or commercial targeting is a topic that is growing in importance. As Floridi puts it:

> "There are very few Moby-Dicks. Most of us are sardines. The individual sardine may believe that the encircling net is trying to catch it. It is not. It is trying to catch the whole shoal. It is therefore the shoal that needs to be protected, if the sardine is to be saved." (32)

Restricting access for commercial use may be reasonable but it could also cut off a valuable source of funds for EOSC – big data companies: Amazon, Microsoft, Apple, Google, Facebook etc., have huge budgets and are a source of data, data processing resources and expertise themselves. Similarly, if a drug company can make its development pipeline more efficient by using chemical data drawn from EOSC, should that process be managed, in part as a financial transaction? This also raises ethical and legal questions around what 'stewardship' and 'ownership' of data mean, and how intellectual property rights can be managed in an era of open science. The unintended consequences of data can even create security and safety issues, as with the recent incident of the layouts of US air force bases being exposed by the publicly shared running maps of exercising personnel (33).

Other issues considered within data ethics include algorithmic processing and artificial intelligence, for instance the constraints and transparency requirements that might need to be applied to both if some form of control is to be retained. Although much of current 'big data' is commercial in nature, generated by point-of-sale (POS) terminals and used for market analytics, governments at various levels are also creators and consumers of such data, and this social, economic and political data is likely to be of interest to some EOSC users.

An obvious question is whether, and under what circumstances, algorithmic processing of this type of big data could occur without explicit consent, in the 'public interest', (assuming other human rights were not violated), by scientists and / or government agencies, and how a public interest justification might be regulated and work in practice.

All of these are fast developing areas and underline the fact that for EOSC the main requirement will be to have a responsive governance structure that can react to rapidly evolving ethical data issues as they arise. It will also be necessary to acknowledge the specific needs of different scientific domains. This is probably most obvious with sensitive personal data, considered below.

3.4. Sensitive Personal Data

Because of its importance, many of the discussions within data ethics have focused on this type of data, often drawing on the longer tradition of medical ethics, to the extent that some have even claimed a new discipline of 'bioinfoethics' (34). In this discipline, the traditional concepts and principles of ethical research involving human subjects (35) are addressed, including:

- beneficence,
- avoidance of therapeutic misconception,
- minimising risk,
- seeking informed consent,
- the right to withdraw consent,

but the questions are about how the approaches of open science may change the interpretation and implementation of these principles, in effect to see if it is possible to "utilise established ethical frameworks as tools for the evaluation of and decision-making about open science" (36), and if not, to investigate how those established frameworks can be extended or modified.

For example, one of the effects of the development of systems for processing 'big data' has been to make the use of routine clinical data more feasible as a source for health research (data that could, at least in theory, be stored in EOSC). This has led to discussions about whether the traditional distinctions between research and routine care, along with the different ethical and regulatory frameworks around each, now need to be re-assessed (37). A specific proposal, making use of the concept of 'easy rescue' – the principle

that persons should benefit others when this can be done at no or minimal risk to themselves – is that explicit consent is not required for the use of such data if it can be categorised as low risk (38). Other questions being examined in this domain include:

- Ethics review: What is needed in ethics review of data research and who can provide it? Is there a role for traditional research ethics committees, taking into consideration the heterogeneity in the different countries and the fact that there has been – up to now – only limited training in this area for the members required?
- Risk-benefit evaluations: Is there a need to develop a specific framework for risk-benefit evaluations of research based on data already collected?
- Consent and the right to withdraw it: Is it possible to have 'broad' consent, e.g. in which it is acknowledged that future use cannot be predicted, but a framework can be agreed under which their data will be shared and used? How practical is the right to withdraw consent once data has been shared with others?

The intention is not to rehearse these debates here, only to give examples of the issues that will need to be considered. In health research in particular, any discussion on ethics is also made more complex by the presence of a variety of legal and regulatory frameworks, (especially for work involving medicines or medical devices). For instance in Europe there is the Clinical Trials Regulation (39) and the GDPR (40), and globally there are ICH GCP guidelines (41) for the conduct of clinical trials. Although these mostly apply to the collection of research data, guidelines have also been developed for ethical principles and governance regarding health databases and biobanks, such as those adopted by the World Medical Association as the Declaration of Taipei (42).

The problem for an organisation like EOSC, spanning many countries and legal jurisdictions, is that slightly different regulatory and legal frameworks may apply in those jurisdictions. For example it has been made clear that, in terms of personal data used in research, national legislatures will have the power to modify the provisions of the General Data Protection Regulation. This means that researchers risk operating unlawfully if they share research data and samples across bor-ders where different laws are in force, without operating due diligence. This is a critical issue and one that many are worried about (43).

Health research data provides perhaps the most obvious but certainly not the only example of a set of domain specific ethical issues. The social sciences and humanities also process data relating to individuals and many of the same concerns exist within that domain (e.g. consent, data anonymisation, the changing legal framework), together with some issues specific to those sciences, e.g. the linkage of administrative and survey data, or the privacy risks associated with linguistic samples (44). Genetics, psychology, and micro-economics are amongst other disciplines that can also generate personal sensitive data, and of course most of us, whether we are aware of it or not, now also leave a detailed digital record of our behaviour as we navigate the web, or generate content on various forms of social media. The latter is mostly held by commercial companies, but as shown by the recent (March 2018) fiasco over Facebook and the use of big data for 'psychographic profiling' in political campaigns, it is not necessarily securely managed. A concern for open science advocates, including the EOSC, is that the device used to collect the Facebook data, by a company calling itself 'Global Science Research', was a personality quiz that claimed to be part of a psychological research project (45). The longer-term impact, on genuine research attempting to collect data in this way, remains to be seen.

The European Data Protection Supervisor (EDPS) has called for a 'big data protection ecosystem' to protect individuals from abuse of their data. This has four main components (adapted from 28):

- Future oriented regulation – building on the GDPR and the proposed revisions to the e-Privacy directive, ensuring competition and consumer rights are not distorted by data abuse, and developing a 'commons' of open data where data like statistics and maps can be available and exchanged with less risk of surveillance.

- Accountable controllers – Controllers with internal policies and control systems that ensure compliance and provide relevant evidence, in particular to independent supervisory authorities.
- Privacy conscious engineering – Including research and development into methods for ensuring accurate audits and for determining the compliance of controllers and processors, such as by 'tagging' units of personal data with 'metadata' describing data protection requirements. The implementation of algorithms that conceal identities at the same time as harnessing the predictive power of the data.
- Empowered individuals – promoting a 'prosumer' environment, where individuals are not passive objects but enjoy rights and responsibilities in terms of data creation and control; clarifying the nature and bounds of consent; developing the concept of ownership and control over personal data and its practical implications.

These are all, clearly, important issues that have direct relevance to the European Open Science Cloud and any sensitive personal data held and processed within it. Indeed it is suggested that the EOSC will be in a unique position, in both policy and engineering terms, to develop these ideas further and promote their practical implementation.

3.5. Science and Society Ethics

There is a group of questions that are specifically about the interface between scientific activity and the wider society. These include issues about how scientific projects and results are reported in the press and media, including social media, and thus how the general public view and 'consume' science, and how scientific results are fed into government policy, how government policy influences scientific activity, and how governments collect their own data.

For example, most academic analysis is presented in statistical terms, but, as shown by the success of lotteries, casinos, and the betting industry, relatively few people can or want to apply statistical reasoning in their own lives. They are helped in ignoring statistical caveats by much science reporting, which notoriously simplifies research results down to a few simple headlines or graphs, designed for impact rather than accuracy. Similarly, the questions of internal and external validity of the data and its analysis, although they may be present in the original report, are often removed in the popular presentation of the data.

That raises the question of whether EOSC should initiate and / or support initiatives to improve the quality of scientific reporting, and especially of the data it manages and the scientific papers generated from it. This could include educational programmes for journalists as well as members of the public, the use of web-based information to clarify or qualify what appears to be incorrect or over simplified information, and active liaison with the popular press, including support for researchers writing press releases. Where misleading statements do appear to be made they could be actively opposed using more accurate information.

Communication with the public is not a trivial issue. In the UK the Care.Data initiative, an attempt to centralise the health and social care data of the whole population (and make some of it available for research), degenerated into an expensive shambles and was abandoned, because of poor presentation of the programme's aims and objectives to both the general public and their doctors, the general practitioners (46). It also failed because the opt-out offered was not a true opt-out (you could have your identifiers removed but not your data). As a consequence, major damage was done to public trust and considerable work will be needed to restore confidence in similar initiatives in the future, for instance by using frameworks like the UK's National Data Guardian (NDG). The NDG was introduced in 2014, and "advises and challenges the health and care system to help ensure that citizens' confidential information is safeguarded securely and used properly" (47).

Communication with governments is perhaps even more critical, especially at a time when some governments seem ready to define their own 'truths' to meet their political ends. Partly in response to this, the Brussels declaration of 2017 (48) proposed a twenty-point set of ethics and principles to inform work at

the boundary between science, society and policy. It was offered as “an attempt to provide guidelines for incorporating scientific progress into the policy-making that affects all areas of our lives.”

Conversely, some governments are seeing the potential of big data analysis for informing their policies, and are encouraging its use. Perhaps because of past failures, the UK government has recently published the Data Science Ethical Framework, providing a set of principles for setting up data collection systems by government departments (49).

How EOSC could or should insert itself into the interface between governments and scientific data is an interesting open question, but one that the organisation’s governance needs to at least consider. Within the H2020 programme, the issue is already being tackled, in part, by the Responsible Research and Innovation (RRI) programme, “an approach that anticipates and assesses potential implications and societal expectations with regard to research and innovation” (50). RRI promotes the idea that all the actors involved (researchers, citizens, policy makers, business, charities, etc.) should try to work together during the research and innovation process, to better align both the process and its outcomes with the needs and expectations of society. It may be that the list of actors should be extended to also include the EOSC.

4. POLICY RECOMMENDATIONS

Given the wide range of possible ethical issues that could potentially be considered by or within EOSC, and the constraints, described earlier, caused by current uncertainties over exactly how EOSC will develop, policy proposals are presented here using a 'layered approach' (see figure 2). The initial 'layer' of policies are seen as fundamental and inevitable. Above them are sets of policies that almost certainly will be useful and should be implemented, and then come policies that would be 'nice to have' but are probably not essential, or that could be developed in later years.

Layer 0

These are seen as fundamental ethical, or ethics related, requirements that EOSC must implement. They fall into two groups:

- Organisational policies that demonstrate that EOSC itself, as an organisation, both recognises and implements the principles of ethically sound practice, to try and ensure its own actions are always defensible, and to maintain the trust of its users and the wider community. This would include a commitment to transparency, of financial dealings as well as of decision making, demonstrating independence from commercial interests or pressure groups, building in appropriate appeals mechanisms for some types of decisions, fairness in the management and allocation of services, and selection and treatment of staff according to modern best practices.
- Data management practices that support and enable research integrity. These include the clear identification of provenance of all materials, including company affiliations where applicable, and the clear signalling of problems identified with data or other materials, including retracted material. It also includes developing mechanisms to ensure that re-used material is properly cited, so that credit can be given to the original data generators. It implies a co-ordinated and comprehensive approach to metadata management within the science cloud.

Layer 1

This layer consists of a variety of theme or discipline specific, time limited, expert task groups, created to consider specific issues and responses – examples include the use of AI in analysing data, the use of consent for sharing personal data, the accessibility of potentially dangerous microbiological data, or intellectual property rights with reference to the pharmaceutical industry. A task group might be concerned solely with an ethical issue, or debate that issue within a wider discussion, for instance about the community's requirements for services in a particular discipline.

There may be issues that can be identified, and groups established, as EOSC begins, as with the current data protection subtask. Another example might be a group looking at general metadata management, which as discussed above has a strong ethical component. In general, however, these groups are likely to be reactive, in the sense that an ethical issue or problem will arise and be recognised as requiring a policy decision, and a task group will then be formed to consider it and report back. This approach fits in well with the proposed EOSC governance model, with the groups being a natural part of the diverse 'stakeholder forum' that will make up the organisation's 'steering layer' (51).

In most cases legal expertise will be required as well as subject specialist knowledge and ethical input, as applicable legislation will always need to be considered when proposing policies and processes. In fact some groups may be largely 'legal', and some largely 'ethical', but most will require a mix of the two types of input.

Layer 2

As a natural extension of Layer 1, this layer introduces central co-ordination of the specific task groups described above, by establishing a standing, general EOSC ethics advisory board. As above, it probably makes more sense to make this a combined ethics and legal board, so an EOSC Ethics and Legal Advisory Board or ELAB. Such a committee should be independent of the EOSC executive, but it forms a natural part of the envisaged 'strategic layer', in effect setting and co-ordinating the ethical agenda for the organisation.

Establishing this group would also make an important public statement about the importance of ethical and legal considerations to EOSC.

Figure 2 - A summary of the different levels of commitment to ethical (and legal) issues possible within EOSC

An ELAB can provide several key functions:

- It can coordinate the work of the specific task groups described in layer 1 by anticipating as well as identifying issues and then seeking input from the relevant specialists. Rather than appearing reactive, EOSC becomes an organisation that is actively trying to advance open science, and support scientific communities, by clarification of the best and most ethical practices within the current legislative framework.
- It can provide the focal point that any specialist ethics / legal task group can report to, freeing such groups from any pressures (real or perceived) from the EOSC executive.
- It can provide a periodic report (perhaps every one or two years) to the EOSC strategy forums about the ethical and legal issues facing the organisation and the current 'performance' of EOSC's executive in ethical terms. A public version of that report should be published.
- When necessary, it can coordinate internal processes in which problems can be shared confidentially, discussed and acted upon. The risk is that if all problems must be discussed publicly there is the risk of a defensive culture developing, potentially leading to 'cover-ups', so having this option available is a useful additional safeguard.
- It can identify and coordinate possible initiatives, such as those listed in the 'higher' layers described below, to help EOSC take the lead in selected ethical and legal issues surrounding science and its interactions with society.

Developing an ethical / legal infrastructure up to this level is proposed as **the minimum required of EOSC** if its commitment to ethics is to be taken seriously. It provides a mechanism for managing the ethical (and legal) aspects of the science cloud, and for ensuring that the organisation is periodically 'inspected', with the

feedback made public, and it also gives the organisation an effective mechanism for dealing with new ethical and legal problems as they arise. In that sense establishing an ELAB is like setting up a Security Management System within an ISO 27001 certification process. The key requirement of ISO 27001 is not specific security measures but a mechanism for keeping security under review, and thus identifying the measures required on an ongoing basis. The EOSC ELAB would have the same function – keeping the organisation informed of the issues that need to be addressed and overseeing the process by which they are tackled and implemented.

Layer 3

Once the basic ethical and legal infrastructure of layer 2 is in place it becomes possible to build on that. One way is by providing training and training materials for research staff in relevant ethical and legal issues, raising awareness and helping to promote research ethics in the more general sense. The focus here is on scientific practitioners and organisations, and other activities could therefore include promoting debates on ethical / legal issues within scientific meetings, collecting the views of researchers about these issues, and helping to organise the production of ethical / legal guidelines.

Layer 4

This final layer considers how EOSC could play a part and influence the wider debates concerned with the interface between science and society. This could include expanding the training programs described above, but this time providing input and materials to non-scientists, especially journalists. It could include examining how social media could be used to actively fight misleading interpretations, or even denial, of scientific data and results. It could also include working with government agencies to see how policies can best be informed by the data being generated and stored within EOSC. At this time these activities are hard to define exactly, but they represent an important opportunity for EOSC, to raise its own profile, to advance the cause of science and scientists, and to increase the 'return-on-investment' provided by EOSC at a societal level.

5. CONCLUSIONS AND NEXT STEPS

This paper has attempted to show that the promotion of ethical behaviour within the EOSC needs to consider a very wide range of issues. In particular, it has tried to emphasise that 'ethics' covers a much wider area than the management of sensitive personal data, extremely important though that is, and that debates about ethics need to include all scientific disciplines, and not just those that collect and process personal data. It has also tried to stress that awareness of ethical issues, and appropriate policy responses to them, need to be integrated into the 'core' of EOSC and the way it is managed from the outset, and not seen as a peripheral topic, or as somehow a supplement to the main scientific activity.

At the same time, there is a recognition that it is difficult to predict the precise nature and priority of the ethical challenges that the organisation will face in the future. At the moment, it is also difficult to judge the level of financial and practical support that EOSC will be willing and / or able to provide for promoting ethical behaviour. Because of that we have proposed a highly flexible 'layered' approach, that stresses the need for responsive systems, targeted expert input and periodic review.

To be clear, the 'layers' here represent different levels of commitment to the active management of ethical issues in scientific research. The twin components of level 0 are seen as essential and inevitable. The activities and structures described as levels 1 and 2 are seen as reactive and proactive mechanisms, respectively, for the EOSC to manage specific ethical challenges. Level 3 takes this further by proposing that the science cloud works with scientific communities to promote understanding and involvement in ethical issues, whilst level 4 has EOSC taking a societal perspective, and working with governments and public opinion to promote the ethical understanding and application of science.

The clear recommendation is that the proposals in layers 0, 1 and 2 should be implemented within the EOSC from the outset, if the organisation is going to have a credible and effective ethics strategy. The activity of levels 3 and 4 are certainly desirable, but could be introduced as the EOSC itself matures, and the level of available resources becomes clearer.

To support this recommendation, we suggest that the next steps should include:

- Discussion and refinement of these proposals, including these 'next steps', with at least four different groups:
    - Experts in the field of ethics, particularly in the areas discussed in this document. (Such a group could possibly form the core of an ELAB). A one-day face-to-face workshop with senior ethical experts and the EOSCpilot task force on ethics has been proposed.
    - Representatives of ministries and other funders of the EOSC.
    - Science demonstrators or alternative case studies dealing with ethical issues. This will help to identify the particular issues raised by the types of activity envisaged in the EOSC.
    - Senior managers in Research Policy Organisations and Research Infrastructures

  The results of these discussions should feed, so far as time allows, into the final deliverables of WP3, and refine and support the initial proposals made here.
- An exploration of how an Ethical and Legal Advisory Board (or equivalent body) might be established and run, e.g. terms of reference, size and membership, accountability and reporting, with a deliverable of specific proposals.
- Establishment of a small expert task group to examine how EOSC's own procedures and policies, and those of organisations working as EOSC service providers, can be developed to ensure that they incorporate good ethical practice, and how that practice can be monitored. This fits with the development of WT 3.2 'Policy supporting services' (e.g. inclusion of conformance to ethical standards / codes of conducts within components of the policy toolkit).

- Establishment of a small expert task group to look at metadata requirements and management in EOSC, especially in terms of generic ‘provenance’ and ‘discovery’ metadata (rather than detailed descriptive metadata, which will usually be domain specific), to ensure that the policies developed can support research integrity and promote the proper reward and recognition of data generators. Again this feeds into the development of WT 3.2 ‘Policy supporting services’.
- The expansion of the current task group looking at data protection issues to include an appraisal of general ethical issues in the area of sensitive personal data. In effect this would be an amalgamation of work around managing such data, dealing with both the legal and ethical aspects, as the current split between two task groups is artificial. At this stage this might not need to produce much more than a list of the ethical issues to be tackled, but a preliminary ‘road map’ in this area would provide important clarification about future tasks.

6. IMPLICATIONS OF RECOMMENDATIONS

The table presented on the following pages summarises possible implications of the recommendations for each of four key stakeholder groups

- EOSC Governance structures and the Rules of Participation
- Funders and Ministries
- Research Producing Organisations
- Research Infrastructures

Some of these implications are requirements – because some stakeholders will be involved in funding and organising the recommendations, others are impacts, and describe the beneficial consequences of implementing the recommendations.

Draft Policy Recommendation	EOSC Governance/Rules of Participation	Funders and Ministries	Research Producing Organisations	Research Infrastructures
L0A: EOSC and its constituent organisations demonstrate ethical practices (transparency, independence, fair decision making etc.)	(a) The details of the 'ethical practices' need to be defined and agreed. (b) A commitment to them needs to be built into the Rules of Participation. (c) EOSC should then be more defensible, against potential legal action or social criticism.	(d) The process of defining and agreeing 'ethical practices' (see a) needs to be funded. (e) Assurance is available that EOSC activities are justifiable and defensible. (f) A clear set of policies is available that EOSC can be monitored against – see L2B.	(g) A demonstrable commitment to the defined ethical practices will be required. (h) Benefit from a more consistent, clearer set of organisational policies and guidelines. (i) Benefit from a more settled and professional EOSC (core) organisation, with greater staff satisfaction and stability.	(j) A demonstrable commitment to the defined ethical practices will be required. (k) Benefit from a more consistent, clearer set of organisational policies and guidelines. (l) Benefit from a more settled and professional EOSC (core) organisation, with greater staff satisfaction and stability.
L0B: Metadata is managed and monitored to support research integrity, (provenance, credit, status etc.)	(a) Consistent policies for provenance and discovery metadata, that support research integrity, need to be defined. (b) A commitment to them needs to be built into the Rules of Participation. (c) Ongoing monitoring and support of metadata application (including this type of metadata) needs to be developed.	(d) The process of defining and agreeing the required metadata rules (see a) needs to be funded. (e) Assurance is available that research integrity is being actively supported within EOSC, thereby enhancing the quality of the available resources.	(f) Demonstrable, consistent application of the provenance and discovery metadata will be required. (g) Training and tools to support correct metadata application required, (and will need funding). (h) Research outputs and their provenance are more accurately described. (i) Individual researchers enjoy more accurate, fairer recognition (e.g. academic credit apportioned accurately).	(j) Support of consistent application of the provenance and discovery metadata will be required. (k) Tools to support correct metadata application will be required, will need development and funding.

Draft Policy Recommendation	EOSC Governance/Rules of Participation	Funders and Ministries	Research Producing Organisations	Research Infrastructures
L1: Establish a variety of theme or discipline specific, time limited, expert task groups, created to consider specific issues and responses	(a) A framework needs to be developed for establishing such groups, their terms of reference and reporting, and their funding.	(b) Funds (relatively limited) need to be set aside for the support of these groups. (c) Funders would have the ability to request work on specific topics, if they had concerns about them.	(d) Problems are examined by the most appropriate experts with solutions arising within the relevant communities. (e) Problems are examined in response to perceived need, with flexibility about where effort is applied.	(f) Problems are examined by the most appropriate experts with solutions arising within the relevant communities. (g) Problems are examined in response to perceived need with flexibility about where effort is applied.
L2A: Co-ordinate specific groups using an EOSC Ethics and Legal Advisory Board (ELAB) – identifying issues, establish groups, etc.	(a) As L1, but the framework now co-ordinated by the Ethics and Legal Advisory Board (ELAB). (b) The terms of reference, reporting lines, membership selection (etc.) of the ELAB itself need to be agreed. (c) Rules of Participation need to include recognition of ELAB and a commitment to co-operate with it when necessary.	(d) As L1, but funds now also needed to pay for the ELAB standing group, though still relatively limited. (e) A 'go-to' group* is available with which funders can discuss ethical / legal issues of concern and co-ordinate activity. (f) Confidence that scientific communities are themselves addressing relevant ethical issues in a timely and pro-active way.	(g) As L1, but potential problems can now be identified and addressed by relevant research communities in a pro-active fashion. (h) A 'go-to' group* is available within the EOSC structure with which these issues can be discussed.	(i) As L1, but potential problems can now be identified and addressed by relevant RIs in a pro-active fashion. (j) A 'go-to' group* is available within the EOSC structure with which these issues can be discussed.

Draft Policy Recommendation	EOSC Governance/Rules of Participation	Funders and Ministries	Research Producing Organisations	Research Infrastructures
L2B: Periodic review of EOSC activity and related contextual issues by ELAB, feeding into executive structures	(a) Mechanism available for feedback to executive on ethical and legal issues and compliance with relevant policies. (b) Details of process (frequency, terms of reference, etc.) would need to be agreed.	(c) As L2A, but additional funding required for the periodic review activity (d) Increased confidence that the management of EOSC and scientific communities are monitoring themselves and holding themselves responsible.	(e) Provides greater assurance to researching organisations of appropriate ethical behaviour within EOSC. (f) Improves opportunity to identify new issues that need to be examined.	(g) Provides greater assurance to research infrastructures, of appropriate ethical behaviour within EOSC. (h) Improves opportunity to identify new issues that need to be examined.
L3: Providing training and training materials for research staff in relevant ethical and legal issues related to EOSC	(a) Some oversight required to record and monitor activity in this area (b) Co-ordination of occasional impact and cost / benefit studies on this type of activity	(c) Funding (relatively modest) required these training activities (d) Improved self-governance of scientific activity, with less governmental input	(e) Research staff become better prepared to identify and manage ethical issues, pro-actively.	(f) Research infrastructure staff become better prepared to identify and manage ethical issues, pro-actively.
L4A: Providing training and training materials for civil servants, journalists and others involved in interpreting scientific results	(a) As L3 (b) As L3	(c) Funding (relatively modest) required these training activities. (d) More extensive and considered application of scientific evidence within governmental planning policies.	(e) More accurate reporting of scientific investigations to the public. (f) Better liaison between government departments and research communities. (g) Greater recognition and status for EOSC scientific communities.	(h) Greater recognition and status for EOSC service providers.

Draft Policy Recommendation	EOSC Governance/Rules of Participation	Funders and Ministries	Research Producing Organisations	Research Infrastructures
L4B: Participating in debates over scientific data and results to try to ensure accurate interpretation (formal and informal media)	(a) As L3 (b) As L3	(c) Funding (relatively modest) required on ad hoc basis. (d) Better understanding and interpretation of scientific data (and its limitations) amongst both politicians and the public.	(e) Increased input from scientists and scientific data in public debates and decision making. (f) Increased appreciation of the importance of scientific evidence, increased role and influence for scientists and scientific groups.	

*i.e. A recognised group, permanently available within the EOSC structure, with which these issues can be discussed and to which proposals can be made.

ANNEX A. BACKGROUND ON EOSC AND OPEN SCIENCE

The European Open Science Cloud (EOSC) will be one of the most relevant instruments for meeting the Open Science challenge. Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. Open Science has an impact on the entire research cycle, from the inception of research to its publication, and on how this cycle is organised. Each step in the scientific process is linked to ongoing changes brought about by Open Science (8). Open Science is expected to deal with scientific knowledge of all kinds (e.g. data, code, online software tools, questions, ideas, and speculations) and is expected to support the as early as is practical release of this knowledge (8). EOSC is one of the five lines of policy actions to support the development of Open Science in Europe (52).

So far it is not totally clear what EOSC will be if implemented and functional. The EOSC first High Level Expert Group (HLEG) report advises on what the EOSC system should/could offer. It should be a federated, globally accessible environment where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, enable trusted access to services, systems and the re-use of shared scientific data across disciplinary, social and geographical borders and include the required human expertise, resources, standards, best practices as well as the underpinning technical infrastructures (53).

The EOSC Declaration further expands the EOSC requirements. The EOSC will be developed as a data infrastructure commons serving the needs of scientists. It will federate existing resources across national data centres, European e-infrastructures and research infrastructures and be based on local-to-central subsidiarity. It refers to a continuous dialogue to build trust and agreements among funders, users and service providers (1). Within the EOSCpilot project other documents have been developed, helping to understand the future architecture, governance and functionality of the EOSC (8, 54, 55). For further discussion of ethical issues, the following points are of major relevance (8):

1. EOSC will spread across disciplinary, social and geographical borders
2. EOSC will cover any types of research artefact (e.g. datasets, papers, methods, workflows)
3. EOSC is planned to have light-weighted international guidance and governance and a large degree of freedom regarding practical implementation
4. EOSC will be a system of systems with several actors and diverse roles (e.g. EOSC suppliers, EOSC service providers, EOSC system owners and managers) with the necessity to leverage, harmonize and federate multiple existing systems and solutions provided by various providers.

REFERENCES

(1) EOSC declaration. European Open Science Cloud, New Research and Innovation Opportunities. Brussels, 26 October 2017. Available at https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf

(2) Fecher B., Friesike S. Open Science: One Term, Five Schools of Thought. In: Bartling S., Friesike S. (eds) Opening Science. Springer, Cham. 2014. DOI: https://doi.org/10.1007/978-3-319-00026-8_2. Available at https://link.springer.com/chapter/10.1007%2F978-3-319-00026-8_2

(3) Joint statement by President Clinton and Prime Minister Tony Blair of the UK. The White House. March 14, 2000. Available at https://clintonwhitehouse4.archives.gov/WH/EOP/OSTP/html/00314.html

(4) Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open 2017;7:e018647. doi:10.1136/ bmjopen-2017-018647. Available at http://bmjopen.bmj.com/content/bmjopen/7/12/e018647.full.pdf

(5) Joly Y, Dove E, Knoppers B, et al. Data Sharing in the Post-GenomicWorld: The Experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office(DACO). 2012. PLoS Comput Biol 8(7): e1002549. doi:10.1371/journal.pcbi.1002549. Available at http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002549

(6) Finnie I and Williamson C. CRISPR patent wars. Haseltine Lake. 17 July 2017. Available at https://www.haseltinelake.com/media-centre/blog/2017/july/crispr-patent-wars

(7) Clinical Study Data Request. Available at https://clinicalstudydatarequest.com/

(8) Candela et al. EOSC Pilot D5.1 Initial EOSC service structure, v1.0, 2018.

(9) Casey K. Nine things we now know about the European open science cloud. Horizon, the European Research and Innovation Magazine. 30 November 2017. Available at https://horizon-magazine.eu/article/nine-things-we-now-know-about-european-open-science-cloud_en.html

(10) Bennett C. What is this thing called ethics? Routledge, 2nd Edition, 2015 (page *xiv*).

(11) Hodson H. Revealed: Google AI has access to huge haul of NHS patient data. New Scientist. 29 April 2016. At https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/

(12) Drenth, P. Responsible conduct in research. Science and Engineering Ethics 12(1):13-21. February 2006. DOI: 10.1007/PL00022265. Abstract available at https://link.springer.com/article/10.1007/PL00022265

(13) All European Academies. The European Code of Conduct for Research Integrity. 2017. At https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

(14) The Singapore Statement on Research Integrity, 2nd World Conference on Research Integrity, 21-24 July 2010, Singapore, At http://www.singaporestatement.org/statement.html

(15) The Montreal Statement on Research Integrity in Cross-Boundary Research Collaborations. 3rd World Conference on Research Integrity, 5-8 May 2013, Montréal. At http://www.researchintegrity.org/Statements/Montreal%20Statement%20English.pdf

(16) Irish Universities Association. National policy statement on Ensuring Research Integrity in Ireland. 2014. Available at http://www.iua.ie/wp-content/uploads/2014/06/National-Policy-Statement-on-Ensuring-Research-Integrity-in-Ireland-2014.pdf

(17) University of Hamburg. Richtlinie zur Sicherung guter wissenschaftlicher Praxis an der Fakultät für Erziehungswissenschaft der Universität Hamburg. 2014. Available in German at https://www.ew.uni-hamburg.de/studium/files/richtlinie-zur-sicherung-guter-wissenschaftlicher-praxis.pdf and in English at https://www.fid.uni-hamburg.de/satzung-gute-wissenschaftliche-praxis-en.pdf.

(18) The Wellcome Trust. Guidelines on Good Research Practice. Available at https://wellcome.ac.uk/funding/managing-grant/guidelines-good-research-practice

(19) BBMRI. Code of conduct for health research (in preparation 2018). At http://code-of-conduct-for-health-research.eu/

(20) The Respect Project. Respect code of practice for socio-economic research. 2004. Available at http://www.respectproject.org/code/respect_code.pdf

(21) CSA (Cloud Security Alliance). Code of conduct for GDPR compliance. Available through https://gdpr.cloudsecurityalliance.org/resource/csa-code-of-conduct-for-gdpr-compliance/

(22) CISPE (Cloud Infrastructure Service Providers). Code of conduct for cloud infrastructure providers. 2016. Available through https://cispe.cloud/code-of-conduct/

(23) European Network of Research Ethics and Research Integrity (ENERI). H2020 project. At http://eneri.eu/

(24) Leonelli S. Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems. Phil. Trans. R. Soc. A 2016 374 20160122; DOI: 10.1098/rsta.2016.0122. Published 14 November 2016 . At http://rsta.royalsocietypublishing.org/content/374/2083/20160122

(25) NIH. Report on NIH Funding vs. Global Burden of Disease. Research Portfolio Online Reporting Tools. 2017. At https://report.nih.gov/info_disease_burden.aspx

(26) Joseph P, Caldwell P, Barnes E, Craig, J. Disease burden-research match? Registered trials in child health from low- and middle-income and high-income countries. Journal of Paediatrics and Child Health 53(7). April 2017. DOI: 10.1111/jpc.13537.

(27) Lanphier E, Urnov, F, Haecker S, et al. Don't edit the human germ line. Nature 519, 410–411 (26 March 2015). doi:10.1038/519410a. At https://www.nature.com/news/don-t-edit-the-human-germ-line-1.17111

(28) Floridi L, Taddeo M. What is data ethics?. Published 14 November 2016. DOI: 10.1098/rsta.2016.0360. At http://rsta.royalsocietypublishing.org/content/374/2083/20160360

(29) European Data Protection Supervisor. Opinion 4/2015 – Towards a new digital ethics, Data, dignity and technology. 2015.

(30) Bynum T. Computer and Information Ethics, *The Stanford Encyclopaedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), At https://plato.stanford.edu/archives/win2016/entries/ethics-computer

(31) The FAIR data principles. FORCE 11. At https://www.force11.org/group/fairgroup/fairprinciples

(32) Mons, B et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use, vol. 37, no. 1, pp. 49-56, 2017. 7 March 2017. DOI: 10.3233/ISU-170824. At https://via.hypothes.is/https://content.iospress.com/articles/information-services-and-use/isu824

(33) Floridi, L . Open Data, Data Protection, and Group Privacy. Philosophy & Technology, March 2014, Volume 27, Issue 1, pp 1–3 . At https://link.springer.com/article/10.1007%2Fs13347-014-0157-8

(34) Hern A. Fitness tracking app Strava gives away location of secret US army bases. The Guardian. 28 January 2018. At https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

(35) Carbo T, Smith M. Global information ethics: Intercultural perspectives on past and future research. Perspectives on Global Information Ethics. 15 April 2008. DOI: 10.1002/asi.20851. At http://onlinelibrary.wiley.com/doi/10.1002/asi.20851/full

(36) World Health Organization. Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants, 2011

(37) Lahti L, da Silva F, Laine M, et al. Alchemy & algorithms: perspectives on the philosophy and history of open science. Research Ideas and Outcomes. 2017. 3: e13593.

(38) Kass N, Faden R, Goodman S, et al. The research-treatment distinction: a problematic approach for determining which activities should have ethical oversight. Hastings Center Report Spec. No.S4–S15. 2013. http://dx.doi.org/10.1002/hast.133.

(39) Porsdam Mann S, Savulescu J, Sahakian B. Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue. Phil. Trans. R. Soc. A 2016 374 20160130; DOI: 10.1098/rsta.2016.0130. Published 14 November 2016. At http://rsta.royalsocietypublishing.org/content/374/2083/20160130

(40) Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use; At https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf

(41) General Data Protection Regulation. At http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679

(42) ICH Good Clinical Practice E6(R2) guidelines. At http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf

(43) World Medical Association. Declaration of Taipei on ethical considerations regarding health databases and biobanks, 2016. At https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

(44) European group on ethics in science and technologies: the governance of large research and medical databases in clinical and research multi-centre trials. From the 4th meeting of the EC International Dialogue on Bioethics, Copenhagen, Denmark, 19 June 2012. At https://publications.europa.eu/en/publication-detail/-/publication/a4611683-048c-4160-9f21-d526b128710e/language-en

(45) Schmidutz D, Ryan L, Gjesdal A & De Smedt K. DASISH (Data Service Infrastructure for the Social Sciences and Humanities), Work Package 6: 'Legal and Ethical Issues'. 2013

(46) Bell K. Here's why the Cambridge Analytica controversy is such a big deal. Mashable UK. Available at https://mashable.com/2018/03/18/what-is-cambridge-analytica/#cuck9jQNQGqm

(47) Presser L, Hruskova M, Rowbottom H, Kancir J. Care.data and access to UK health records: patient privacy and public trust. Journal of Technology Science. Published 2015-08-11. At https://techscience.org/a/2015081103/

(48) GOV.UK: National Data Guardian (see: https://www.gov.uk/government/organisations/national-data-guardian)

(49) Ethics & Principles for Science & Society Policy-Making. The Brussels Declaration. At http://www.sci-com.eu/main/docs/Brussels-Declaration.pdf?58b6e4b4

(50) Data Science ethical framework. UK Cabinet Office. Published 19/05/2016. At https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf

(51) European Commission: Responsible Research and Innovation (see: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation)

(52) Hienola et al. EOSC Pilot. D2.2 Draft governance framework for the European Open Science Cloud, v1.1. 2017

(53) European Commission's Directorate-General for Research & Innovation (RTD): Open innovation, open science, open to the world – vision to Europe, https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe

(54) EU Commission EOSC High Level Expert Group. Realising the European Open Science Cloud. European Union, 2016 https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

(55) Tsiavos et al.: D3.1 Policy Landscape Review, 7 January 2018

GLOSSARY

Term	Explanation
Personal data	Means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. N.B. The GDPR does not apply to anonymous data or personal data of deceased persons. *From Definitions, General Data Protection Regulation (EU) 016/679 (GDPR), Recital 27.*
Sensitive Data	Four types of sensitive data can be distinguished: personal information, business information, classified information and 'sensitive issue' data:
Sensitive personal information	Data that can be traced back to an individual and that, if disclosed could result in harm to that person. The following data is considered sensitive personal data and subject to specific processing conditions: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data processed solely to identify a human being; health-related data and data concerning a person's sex life or sexual orientation. These categories of personal data are subject to additional protection under the GDPR. *See Article 4(13), (14) and (15) and Article 9 and Recitals (51) to (56) of the GDPR.*
Sensitive business information	includes anything that poses a financial or existential risk to a company if discovered by a competitor or the general public. Such information includes trade secrets, acquisition plans, financial data and supplier and customer information.
Sensitive classified information	is restricted according to a designated level of sensitivity (for example, restricted, confidential, secret and top secret), usually by a government body. Examples could be plans to build a bomb or a highly pathogenic virus. Classified information is data that must be protected from unauthorized access to safeguard the safety or security of an individual, group or organization. *Adapted from http://whatis.techtarget.com/definition/sensitive-information*
Sensitive issue data	Includes – for example – sociological data relating to different social or ethnic groups, e.g. their occupations, income levels, criminality; biological data such as

	species decline, or the impact of genetically modified crops; psychological data, e.g. levels of mental illnesses, measured stress levels, or social psychology experiments on conformity; environmental data, e.g. relating to climate measurements or marine pollution; even geological data, e.g. the presence of frackable shale, radioactive gases in granite, or rare mineral deposits. It is not expected or suggested that this type of data have restricted access. But it is data that is often selectively reported, simplified, distorted, etc., to make a political point. The question is whether there is an ethical duty to a) think carefully about how such data should be presented, and b) monitor the (mis)use of such data, and try to counteract that misuse when it is considered important to do so. Whether and how EOSC should tackle this issue is an open question, but it is a question that should be debated.