journal of Geoethics and social Geosciences

Journal of Geoethics and Social Geosciences, 02, Special Issue, 2025; doi:10.13127/jgsg-64

The ethical dimension of sharing solid Earth Science data

Massimo Cocco[1,], Rossana Paciello[1,2], Daniele Bailo[1,2], Mario Locati[3], Federica Tanlongo[2], Luca Schirru[4,5,6], Maria Incoronata Fredella[2], Daniela Mercurio[2] and Carmela Freda[2]*

[1]Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

[2]EPOS ERIC, European Plate Observing System European Research Infrastructure Consortium

[3]Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy

[4]Centre for IT and IP Law KU Leuven, Belgium

[5]Brazilian Copyright Institute, Brazil

[6]National Institute of Citizen Science, Brazil

Article history: received November 5, 2024; accepted January 30, 2025; published February 21, 2025

Abstract

Open Science is the paradigm driving the sharing of research data worldwide. It includes the ambition to make FAIR (Findable, Accessible, Interoperable and Re-usable) data sharing the default. FAIR guiding principles for research data have been recently proposed to scientific communities as the new horizon for sharing data. The FAIR principles create the conditions to foster data sharing and improve data stewardship, provided that several legislative, organizational, and ethical issues are addressed. In this paper, we aim to discuss the ethical dimension of sharing solid Earth science data. Earth scientists have a long-lasting tradition in data acquisition, quality control, and standardization, being the key actors in feeding and implementing metadata and services for qualification, storage, and accessibility. Pan-European Research infrastructures like EPOS (European Plate Observing System) involve scientific communities and research organizations federating facilities and resources to ensure data management and interoperability through e-science innovation. After introducing the ethical issues associated with the protection of personal data, intellectual property rights, and data misuse, we will focus on the impartiality for public good. This opens a *new horizon to the ethical dimension of open access to research data, going well beyond research integrity. This assumes an outstanding relevance when referring to solid Earth science data since they also concern natural and anthropogenic hazards and risk communication relying on sharing scientific information with different stakeholders. Although we present a specific perspective for solid Earth science, we believe that the addressed ethical dimension is relevant for environmental science in general.*

Keywords: Open Science; Research Data Management; Research integrity; Impartiality for public good

1. Introduction

Open science is a paradigm fostering innovation in science and the adoption of novel approaches to scientific research and communication. It started as a new way of looking at science and its dissemination, becoming nowadays a target regulated by principles and rules adopted by European and national authorities, and research organizations worldwide [European Commission, 2016a]. When applied to research data, Open Science relies on unified access to heterogeneous data sources from openly accessible providers [Vicente-Saez et al., 2018; Paic, 2021; UNESCO, 2022] and technical and engineering solutions to steward data through open data systems. To this goal, FAIR (Findable, Accessible, Interoperable, and Re-usable) guiding principles have been proposed as the basic standard for scientific research [Wilkinson et al., 2016]. They respond to the increasing demand for more efficient data sharing and interoperability, and they play a relevant role for data infrastructures willing to enable a more effective and open research environment. Data FAIRness is therefore considered a relevant goal for research infrastructures in different scientific domains and at a global level. The adoption of FAIR data principles requires viable practices and sustainable solutions. The sustainability of data sharing encompasses various dimensions, including governance, financial, technical, legal, and ethical aspects. Therefore, sustainable data sharing is one of today's key challenges for research organizations and pan-European research infrastructures [Saleh Contell et al., 2022; Bailo et al., 2020, 2022, 2023].

Open access to scientific data relies on the implementation of the Research Data Management (RDM) process during the whole data lifecycle, from data creation to publication and archiving [Sinaeepourfard et al., 2015]. RDM requires the engagement of different actors within the research communities, such as data providers in charge of collection, quality-control, policy definition and curation of research data and metadata, as well as the adoption of technological solutions to ensure access, interoperability, and (re-)use, as exemplified in Figure 1.

Figure 1. *Data lifecycle and Research Data Management (RDM) processes. The figure illustrates the core stages of RDM lifecycle highlighting the key actions. It relies on concepts derived from several established data lifecycle models [Sinaeepourfard, 2015]. It encompasses processes such as data generation, quality assurance, standardization, harmonization, integration, storage, processing, security, distribution, and long-term maintenance. Each stage is critical in ensuring that data is Findable, Accessible, Interoperable, and Reusable (FAIR), addressing the ethical, technical, and organizational challenges associated with data stewardship.*

This is nowadays represented by the FAIR data management paradigm. RDM also requires the governance of the whole process and suitable approaches to tackle the sustainability challenge. In Europe, research infrastructures coordinated by ESFRI (European Strategic Forum on Research Infrastructures) have tackled this challenge [ESFRI, 2017] and are collaborating to design and implement shared solutions to address the sustainable FAIR data management [Petzold et al., 2019].

Research Infrastructures engage scientists, research organizations, and national authorities (i.e., ministries and funding agencies) committed to supporting their mission of providing access to data and facilities, therefore fostering Open Science and the generation of scientific products and services. Open Science and RDM require dedicated human, technological and financial resources, as well as skills and the sharing of solutions and best practices: hence, the key role played by international research infrastructures, which are essential elements to govern and harmonize the RDM process.

Solid Earth science refers to the study of planet Earth and aims at understanding the structure and dynamics of the Earth's solid surface (both on land and offshore) and its interior. It brings together many diverse disciplines (e.g., geophysics, geology, geodesy, and geochemistry). The relevance of solid Earth science extends beyond academic research, influencing human aspirations to make Earth a safe and habitable planet. Geo-processes are independent of national frontiers and discipline boundaries, consequently, scientific investigations into these processes benefit tremendously from seamless, transnational integration of geographically distributed measurements through observing, experimental, and modeling systems. This also holds for the access to standardized and quality-controlled data from the different scientific domains and disciplines of solid Earth science. Progress in the understanding of the physical processes controlling earthquakes, volcanic eruptions, and tsunamis, as well as those driving tectonics and Earth surface dynamics requires a long-term plan to facilitate the integrated use of data, models, and facilities from distributed research infrastructures [Cocco et al., 2022; Atakan et al., 2022; Bailo et al., 2023, 2024].

The European Plate Observing System (EPOS[1]) represents this long-term plan for the integration of research infrastructures for solid Earth science in Europe. EPOS brings together European research infrastructures and their associated data and services together with the scientific expertise into its integrated open platform. By improving and facilitating the integration, access, use, and re-use of data, data products, services, and facilities, EPOS developed a sustainable, multidisciplinary research platform to provide coordinated access to harmonized quality-controlled data from diverse Earth science disciplines together with tools for their use in analysis and modeling enabling a step change in multidisciplinary scientific research. The engaged diverse thematic communities have experiences and skills on data acquisition, quality control, and standardization of data and metadata, implementing metadata and services for qualification, storage, and accessibility.

Data sharing and stewardship rely on addressing legislative, organizational, and

[1] https://www.epos-eu.org/ (accessed 28 January 2025).

ethical issues. This is particularly relevant for pan-European research infrastructures like EPOS, since these issues belong to governance and legal dimensions of their sustainable operation spanning different countries and legislations. Dealing with solid Earth science data and services, the ethical dimension for EPOS is particularly relevant, and it includes both the broader ethical issues of Open Science and those of risk communication related to natural and anthropogenic hazards. Contextualizing the Open Science principle "as open as possible, as closed as necessary" strongly depends on the scientific field, and this is particularly true for solid Earth science data, since these research products are used for evidence-based decision-making and risk communication. In this perspective, it is important to distinguish the access to research data from the sharing of scientific information and knowledge, as this distinction has implications for the ethical dimension of Open Science.

This paper aims at discussing the ethical dimension of sharing solid Earth science data, providing a context for the Open Science principle "as open as possible, as closed as necessary" and the CARE[2] Principles for Indigenous Data Governance [Carrol et al., 2020]. We first briefly discuss some well-recognized ethical issues related to data sharing, namely the protection of personal data, Intellectual Property Rights (IPR), and the misuse and abuse of research data. The latter issue is becoming increasingly important due to the evolving international political contexts and the new developments of Artificial Intelligence (AI) [Cleverley, 2024]. However, discussing the new ethical dimension of Open Science due to the evolution of AI is beyond the goals of the paper. We will focus on the ethical dimension of public research, namely impartiality for the public good and its applications to open access to the research data routinely used for the surveillance of territories from natural and anthropogenic hazards (such as earthquakes, tsunami, volcanic eruptions). This paper aims to share the experiences matured in designing, implementing, and operating the EPOS Research Infrastructure (EPOS RI) and its platform to access data and services, hoping that they might be of interest to other research infrastructures and organizations addressing the ethical issues of sharing research data.

2. EPOS, a Research Infrastructure for solid Earth science

The EPOS RI has been designed (2007-2010) with the ambition to address data interoperability for the whole solid Earth sciences. During the Preparatory Phase (2010-2014), the EPOS architecture (Figure 2) was designed to enable the RI to operate as a distributed research infrastructure.

[2] CARE is an acronym for Collective benefit, Authority to control, Responsibility, and Ethics.

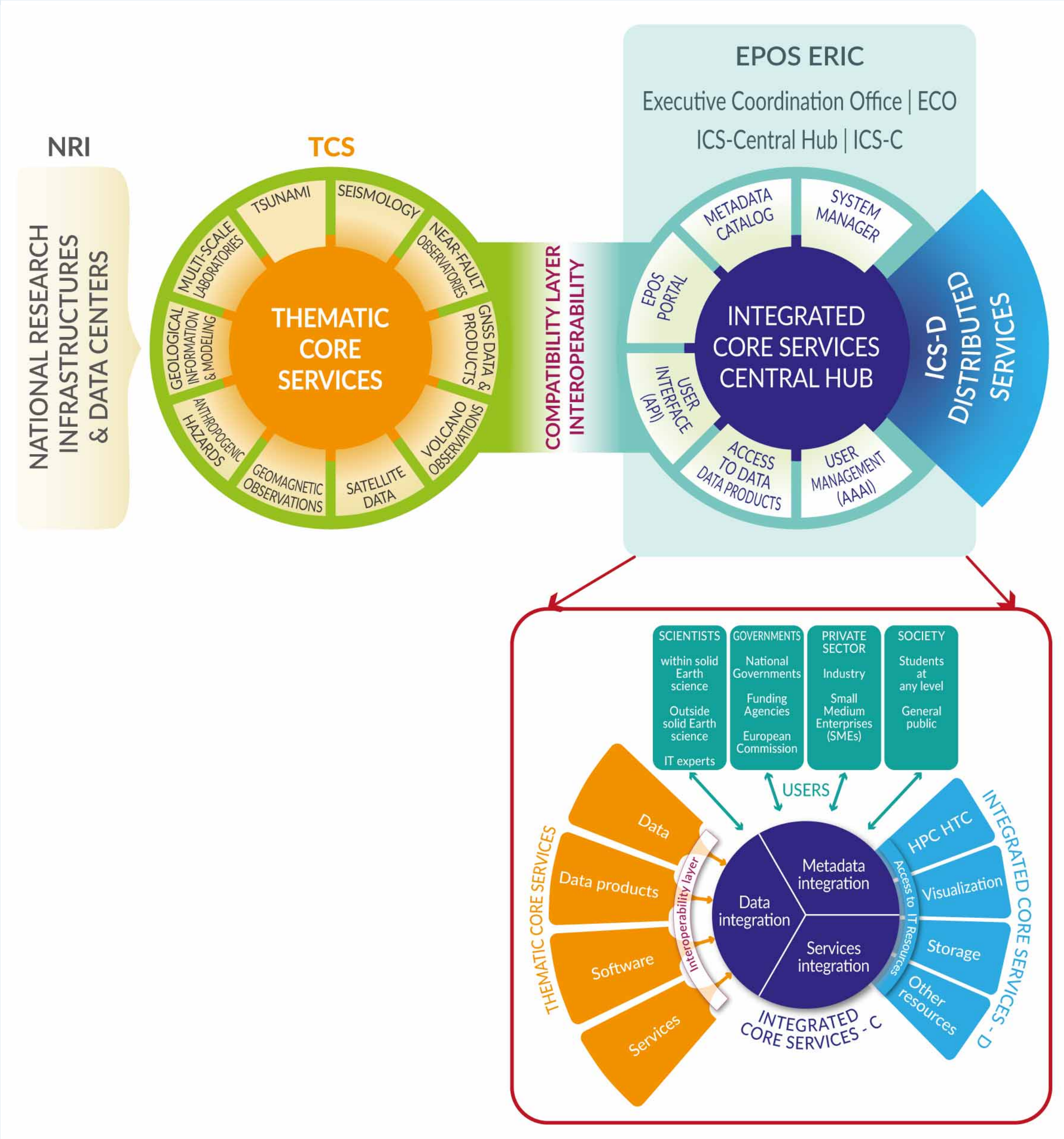

Figure 2. *The EPOS architecture consists of four key elements designed for enabling: i) data generation from National Research Infrastructures (NRIs); ii) data qualification and management by thematic communities (i.e., the Thematic Core Services, TCS); iii) data integration and provision through the Integrated Core Services (Central Hub, ICS-C and Distributed, ICS-D); iv) sustainable operation and governance of the pan-European research infrastructure by EPOS ERIC (through the Executive Coordination Office). The Integrated Core Services are further described in the box, showing the integration of data, metadata and services to enable data interoperability and use. The different user categories are also shown in the box.*

The decision to build and operate a single research infrastructure for the whole solid Earth science was motivated by the awareness of the added value of data integration, and it has been strategic for the design and the implementation of the EPOS architecture during the Implementation Phase (2015-2019).
The distinctive feature of the architecture design relies on assembling key elements belonging to four complementary categories, mapping the data generation, qualification, integration, and provision: namely, National Research Infrastructures (NRIs), Thematic Core Services (TCS), Integrated Core Services (Central Hub, ICS-C and Distributed, ICS-D), EPOS ERIC Executive Coordination Office (EPOS ERIC ECO) [see Cocco et al., 2022; Saleh Contell et al., 2022; Bailo et al., 2020, 2022, 2023; Atakan et al., 2022; Cocco and Montone, 2022 for further details]. For the goals of this paper, it is relevant to discuss a few key features of the EPOS architecture because of their implications on the governance of the RDM process, ethical issues, and Open Science:

i. *Centrality of data and service interoperability*. This has been identified as the founding element of functional architecture since EPOS' conception and well before the creation of the FAIR acronym. Then, after the release of the FAIR principles, the approach undertaken for the construction of the EPOS Platform, which focuses on the management of data and metadata as well as access and usage of services, has been tailored to the FAIR principles (Figure 1) [see Bailo et al., 2020] to highlight its coherency with the FAIR data management. This emphasizes that EPOS has been designed to enable findability, access, interoperability, and use of research data and corroborates its key role in Open Science and sustainable open access.
ii. *Federated approach* to integrate different scientific communities at a pan-European level. This has implications for the engagement of many diverse data and service providers distributed among different European (and Mediterranean) countries and contributing to the RDM governance while addressing the sustainability of open access and Open Science with numerous European research organizations [Bailo et al. 2024]. Indeed, a common challenge of distributed RIs is the balance between the need to acknowledge data traceability and accountability of data providers and the necessary adoption of rules for protecting personal data. The pan-European dimension of EPOS, resulting in 180 research organizations from 26 countries representing 10 disciplinary communities, gives an idea of the effectiveness of its federated approach enabled by the operative governance of EPOS ERIC, the European Research Infrastructure Consortium acting as the legal entity operating and coordinating EPOS.
iii. *Involvement of national research organizations,* which are engaged in EPOS

with the dual role of data creators at the national level (NRIs) and the provision of human resources and skills to the thematic communities (i.e., Thematic Core Services, TCS). National (public) research organizations play a fundamental role in Europe in supporting open access to scientific data and governing research data management, as highlighted also in EPOS [Atakan et al., 2022]. They are responsible for creating and maintaining the proper working environment for the RDM governance, employing the required professional profiles to operate Open Science and FAIR data management (e.g., data stewards, data curators, data managers) [Locati et al., 2024]. They are committed to responding to the laws and rules enacted by national authorities to support Open Science and open access to scientific data. Putting the emphasis back on solid Earth science, research organizations often use research data for the monitoring and surveillance of national territory for Civil Protection agencies and national authorities. In some cases, they are legally involved in the provision of authoritative expertise and knowledge for decision-making through formal commitments and protocols. This role is also associated with the communication of scientific information concerning natural and anthropogenic phenomena and related geo-hazards and risks to society. Providing access to research data used for the surveillance of national territory and authoritative knowledge for decision-making further clarifies the ethical dimension of data provision for solid Earth science.

2.1 EPOS Data key concepts

So far, we have referred generically to research data. Specifically for solid Earth science, it is necessary to adopt a specific taxonomy to distinguish different categories of scientific data. For this purpose, EPOS has adopted an effective data taxonomy based on the following levels [Cocco et al., 2022; Bailo et al., 2020, 2023]:

- Level 0: Raw or basic data acquired by observing systems and instruments representing the output of the sensors or actuators used to measure specific physical quantities.
- Level 1: Data products generated by automated procedures applied to basic (level 0) data.
- Level 2: Data products generated and published by scientists through their research activity. Scientists' intervention is required to analyze and add values to Level 0 or Level 1 data.
- Level 3: Integrated data products coming from complex analyses and

community-shared products, which require collaborative processes and shared commitments.
- Level 4: Software, services, and IT tools.

This distinction is essential to effectively address the governance and legal issues, with particular attention to the ethical aspects associated with Open Science and RDM. It is evident that, while basic data (level 0 in data taxonomy) are usable and used essentially by scientists, research products with high taxonomy level (level 2 and 3) are also used for decision making and risk communication. This explains the different ethical implications of providing open access to raw data (such as a seismogram) and scientific products (i.e., a seismic hazard map, a tsunami inundation map, or a volcanic eruption scenario). Moreover, access to services and IT tools to analyze and interpret high-level taxonomy scientific products might also have ethical implications because their use requires specific expertise, which might restrict the fully open access to all users (for instance, services for seismic hazard assessment at national or regional scale).
Providing access to data, scientific products, and information to different stakeholders requires effective solutions and harmonization with national and international priorities and strategies [Atakan et al., 2022; Martì et al., 2022]. The data and the services accessible through the EPOS data portal [Bailo et al., 2023] concern geo-hazards related to natural (i.e., earthquakes, volcanic eruptions and unrest episodes, tsunami, surface geology) and anthropogenic (induced seismicity) phenomena, but EPOS is not directly involved in the provision of expertise for decision-making. It is evident that increasing the level of data taxonomy increases the impact on decision-making, risk communication, and the safe exploitation of geo-resources. This is a further element to consider when discussing the ethical dimension of sharing solid Earth science data by pan-European research infrastructures or national research organizations.

2.2 The EPOS Data Policy

EPOS relies on data from distributed national data infrastructures coordinated by Thematic Core Services (TCS) through Service Providers (SP), which operate under collaboration agreements with EPOS ERIC. The EPOS Data Policy[3] promotes open, free, and easy access to data, fostering innovation, collaboration, and efficiency, encouraging a culture of openness in data sharing, especially for research funded

[3] https://www.epos-eu.org/data-policy-2018 (accessed 28 January 2025).

by public sources, while also ensuring compliance with European and national regulations (e.g., Public Sector Information, High-Value Datasets) like the Aarhus Convention[4], INSPIRE Directive[5] and Data Governance Act[6].

Data provision is entirely carried out by the National Research Infrastructures (NRIs) according to their mission and internal rules. The data ownership remains with the data providers, which are responsible for the continued management and curation of the data and metadata, as well as their quality and accuracy. The metadata are provided in accordance with the shared standards and practice of the scientific community (TCS). Metadata also includes information about authorship and provenance, always provided in the EPOS platform, thus ensuring proper recognition, transparency, and accountability.

All data providers in EPOS are scientific organizations of proven expertise and authoritativeness, and their participation is validated by a community of peers (the TCS) to ensure the relevance and accuracy of the data that are then channeled into the EPOS infrastructure. This is also the reason why individual researchers are not allowed to use the Research Infrastructure as a repository for their research, as this would eliminate validation by the TCS and national research infrastructures, which is essential to ensure quality and accountability. The EPOS Data policy also foresees that, in addition to the peer validation carried out within the TCS, further quality control can be performed by EPOS itself through external feedback mechanisms.

Metadata is always made freely available, while for data EPOS adopt the “as open as possible, as closed as necessary” policy. Reasonable restrictions (e.g., embargo periods, data access only for certain roles/user profiles, restricted data, no derivative licenses to avoid critical data being altered) are admissible and can be put in place when necessary. Ethical concerns, as well as data protection and IPR-related issues are the main reasons for the motivated restrictions on data sharing.

3. The evolving ethical dimension of open access to research data

The ethical dimension of open access to research data is constantly evolving because of scientific, technological, and legislative advancements at different political scales. Here, we focus on solid Earth science data, and we use the EPOS

[4] https://unece.org/environment-policy/public-participation/aarhus-convention/text (accessed 28 January 2025).
[5] https://knowledge-base.inspire.ec.europa.eu/index_en (accessed 28 January 2025).
[6] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act): https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng (accessed 28 January 2025).

federated approach to integrate research data, scientific products, software, and services as a representative example to discuss the ethical issues of data sharing. EPOS is an operational research infrastructure with data and services already accessible through its portal (see footnote 1). We believe that it is a useful example to address emergent ethical issues in Earth sciences, also representing research organizations committed to collecting, managing, and sharing research data. In this framework, we refer to research data in a more specific way than the broader definition provided by the European Commission (Directive (EU) 2019/1024[7]). There are numerous definitions of research data [CODATA, 2023; Locati, 2024], and discussing this issue is beyond the goal of this paper. Research data can be defined as the evidence that underpins the answer to research questions and can be used to validate findings regardless of their form. These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modeling, interview, or other methods, or information derived from existing evidence [UK Research and Innovation, 2016]. As anticipated above in presenting the EPOS data taxonomy, we refer to research data essentially as digital objects created and used as primary sources to support scientific and technological research and used as evidence in the research process, following commonly accepted standards in the research community, necessary to validate research outcomes.

3.1 Protection of Personal Data

The protection of personal data is the first legal and ethical issue addressed by international research infrastructures and national research organizations. The main challenge for scientific research is to use and share the data while protecting personal privacy [EUI, 2022]. For this reason, it is useful to start from this issue to (partially) describe the evolving landscape of ethics in sharing research data. There is broad legislation at the European level for the protection of natural persons about the processing of personal data and the free movement of such data, such as the Regulation EU 2016/679 (the so-called GDPR: General Data Protection Regulation)[8] and the e-Privacy Directive[9] [Schirru et al, 2023]. The GDPR "*lays down rules relating*

[7] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information: https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng (accessed 28 January 2025).

[8] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation): https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed 28 January 2025).

to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data" (art. 1(1)). The rules and mechanisms to ensure strong protection of personal data and counteract their unlawful and purposeless processing are monitored by Supervisory Authorities, who also provide dedicated guidelines. The GDPR is a regulation, which means it is directly applicable in any country of the European Economic Area (EEA) without the need for implementing the national legislation. Personal data means any information relating to an identified or identifiable natural person (art. 4(1)). GDPR warrants rights to data providers, but also imposes obligations on data controllers and data processors, for example, determining how to process personal data. However, according to GDPR, Member States can enact more stringent restrictions as in the case of certain categories of data such as health and genetic data, national security and defense, prevention of criminal actions and terrorism (among several others). It is evident how this legal framework has ethical implications on open access to research data and affects the solutions to ensure accountability of data providers and traceability of shared research data and products.

The adoption of GDPR involves responsibility and effectiveness in organizing effective data protection measures by research organizations and international research infrastructures. It also requires compliance with the accountability principle by providing evidence that significant measures to protect personal data have been undertaken. The meaning of accountability for a distributed research infrastructure is twofold: on one hand, it implies putting in place reliable and effective measures and records to demonstrate compliance with data protection rules; on the other hand, it means ensuring recognition, visibility, and traceability of data providers.

While most data within the EPOS Delivery Framework do not qualify as personal data, there are specific instances where the management of personal data is necessary. This includes the contact details of individuals responsible for data management, dataset curation, and service provision, as well as dataset authors. Access to this information is essential for infrastructure managers and end users, enabling communication regarding data queries, access to additional datasets, and proper attribution when datasets or services are reused. The collection, storage, management, and deletion of personal data within EPOS comply with the GDPR and other relevant legal provisions. While the data handled is not sensitive (on "sensitive data", see, e.g., GDPR's Recommendations 10 and 51), and personal

[9] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). This act has been changed. Current consolidated version: 19 December 2009: https://eur-lex.europa.eu/legal-content/EN/AUTO/?uri=CELEX:02002L0058-20091219 (accessed 28 January 2025).

information, when collected (e.g., in surveys or contact details), is kept only for as long as needed before being anonymized and aggregated, the complexity of EPOS's multi-component environment requires attention to data handling processes. For example, in the case of contact information for data providers, while this information is provided to the Service Provider as part of their agreement with the data provider, there is no explicit contract between the data provider and EPOS ERIC. This raises questions about EPOS ERIC's entitlement to process such data. Given that the EPOS Delivery Framework operates as a single entity under the governance agreements between its components, careful consideration is necessary to ensure GDPR compliance when personal data is passed between different EPOS entities. To mitigate such complexities, clear data processing agreements should be established when relevant, and continuous monitoring is essential to ensure compliance across the distributed network. Moreover, the persons whose data are being processed should be properly informed of the existence of such agreements. This approach underscores the importance of GDPR adherence not only as a legal obligation but also as a best practice in managing personal data responsibly within collaborative scientific infrastructures.

3.2 Intellectual Property Rights

The second ethical issue in providing access to research data is the protection of Intellectual Property Rights (IPR). In addition to GDPR, research data can be protected by IPR, which could create a scenario where some uses - though not all (e.g., limitations and exceptions[10] for using public domain resources) - require prior and explicit authorization from the rightsholder. This is particularly demanding when integrating research data from different and heterogeneous data providers hosted in different countries and working in distinct research organizations, which can seriously hinder the participation of some players, especially from private sector.

From a legal point of view, intellectual property is composed of industrial property (e.g., patents, trademarks, geographical indications) and copyright (e.g., literary, artistic, and scientific works) [WIPO, 2020]. When the research data is a copyrighted work and is in the public domain, from a pure copyright perspective, they are characterized by the lack of exclusive economic rights, which means that everyone could use them. Data providers may associate a license to the data to address this

[10] See, e.g., Quintais [2017, p.203]: "Limitations are essential tools to balance copyright exclusivity with the public interest and fundamental rights. They enable the promotion of access to, and dissemination of, culture, education and knowledge. In the online environment, they provide a necessary counterweight to the expansion of technically defined exclusive rights to digital activities outside the commercial core of copyright."

issue and clarify the terms of use. In the framework of EPOS, affixing a license to datasets and other scientific products has indeed been an obligation for data providers since 2021, as a part of the EPOS Data policy, with the aim of protecting the rights of the data providers, streamline the reuse of the shared scientific assets, and clarify possible limitations. Most of research data in solid Earth science do not have patents or copyrights but IPR issues may still apply since usually research organizations govern and finance data acquisition and management by means of well-structured original and non-original databases that may be covered by copyright (original databases) or the sui generis rights (non-original databases) in the European Union. Licensing of data requires harmonization and coordination, which is part of the RDM governance done by thematic communities in EPOS. A license is commonly used for granting permission to access, re-use and redistribute research data with few or no restrictions, depending on the chosen type. Creative Commons (CC) open licenses are the most widely adopted licensing framework[11] to deal with copyrighted content and, in some cases, with non-original databases covered by Sui Generis Database Right (such as CC BY 4.0). Free-Libre Open Source (FLOSS) licenses like GNU GPL and MIT licenses are commonly adopted for software. It is evident that this legal framework regulates the implementation of open access, including restrictions (open as possible, restricted as necessary) as well as the re-use of research data, and it is part of RDM. Several licenses for instance prevent the commercial use of research data and this has ethical implications for solid Earth science data, as we will discuss in the following sections. To facilitate broad data dissemination and proper attribution, the EPOS Data Policy requires data and service providers to affix their scientific products with open licenses (e.g., Creative Commons Attribution or open-source licenses such as MIT or GPL, depending on the nature of the scientific product). As for data access, restrictions are allowed when motivated, in case there is a need to protect third-party IPRs or to make sure that the scientific product is not modified. The latter is typically implemented through the application of non-derivative licenses (e.g., CC BY-ND). Ethical concerns regarding the improper use of a scientific product altered by a non-authoritative source may justify the application of this reasonable restriction. All data and service providers are required by the terms of the EPOS Data Policy and the EPOS Data Portal Terms and Conditions[12] to make sure to protect third-party rights in case they are distributing data and scientific products created by others and to put in place

[11] 2011/833/EU: Commission Decision of 12 December 2011 on the reuse of Commission documents: https://eur-lex.europa.eu/eli/dec/2011/833/oj (accessed 28 January 2025); Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use: https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj (accessed 28 January 2025).
[12] https://www.epos-eu.org/sites/default/files/Terms_and_Conditions.pdf (accessed 28 January 2025).

relevant measures for data protection and integrity.
Research organizations play a relevant role in contributing to RDM with particular attention to legal and ethical aspects through their data policies. Institutional Data Policies should offer researchers a consistent and standardized framework for managing, storing, and sharing data, ensuring the protection of individual rights, the confidentiality of sensitive information, and the integrity of the entire research process. These policies simplify the often complex legal and technical requirements set by various external organizations into clear, practical guidelines, making compliance more straightforward and helping researchers navigate the broader regulatory landscape. They also promote collaborations by encouraging open data practices and sharing research findings, benefiting the broader scientific community. In embodying ethical principles such as transparency, fairness, and respect for privacy, these policies are crucial. An institutional data policy can significantly enhance research data management capabilities by providing a structured framework that supports human, technological, and financial resources. Additionally, EPOS ERIC's commitment to interoperability, standardization, transparency, and accountability in data management provides both an incentive and a model for other organizations involved as data providers.

3.3 Data misuse

The first two ethical issues discussed above, the protection of personal data and Intellectual Property Rights, are well-known and regulated by a quite complex but known legal framework. Research infrastructures and research organizations have adopted dedicated measures and rules to address them. A further ethical issue related to open access to data, which has a more heterogeneous context, is the misuse of research data. Concerns about data misuse can lead to restrictions to the open access to research data and the conditions of their reuse, and even arrive as far as totally impeding the open sharing of data. Misuse of research data can take different forms and degrees of severity, spanning from *bona fide* scientific and methodological mistakes, unauthorized reuse, and intentional misrepresentation, including malevolent use. Pasquetto et al. [2024] discussed the implications of underscoring the complexity of defining misuse, considering different epistemological perspectives and the evolving nature of scientific methodologies. These authors found quite different ways in which we can interpret research data misuse, including misuse as analytical errors, misinterpretation, misrepresentation, reputational harm, privacy and geo-privacy violations, exploitation, and uncritical reuse of biased and offensive data. Pasquetto et al. [2024] conclude that misuse of open research data cannot be completely avoided, but this should not be used as a reason for not sharing data or to avoid taking the necessary actions to mitigate the risk caused by data misuse. The complexity of misuse definition should suggest avoiding the adoption of different terms, such as abuse of research data, although this distinction exists and was initially adopted in EPOS to identify and emphasize malevolent use. There is a continuum between misuse and abuse and in some cases, the boundaries between the two can be blurred. Furthermore, a key challenge is constituted by the fact that, while research infrastructures become more open (as the EPOS mission states) and non-experts increasingly participate in the process of using research data, most of the existing practical solutions and guidelines [e.g., APEGGA, 2010] for mitigating misuse in data sharing and reuse are designed for expert data practitioners and are not easily usable. This does not prevent the involuntary misuse of data or, in any case, limits the effectiveness of mitigation actions, requiring specific and demanding contingency actions.

EPOS ERIC has chosen to separate its responsibility for data access from its use. Specifically, EPOS informs users of the portal that by agreeing to the EPOS Data Portal Terms and Conditions - which outline permitted usage - they release EPOS ERIC from any liabilities connected to any subsequent use of the scientific data and products made available through the EPOS infrastructure. Moreover, EPOS ERIC is committed to maximize data accessibility, balancing with its responsibility to foster ethical behavior and a culture of integrity within its community. For these reasons, the EPOS Ethical Guidelines [Di Capua and Peppoloni, 2022] are dedicated to share the principles and provide indications concerning the respect of IPR, personal data protection, accountability of data providers, licensing, and data traceability. Strategies and solutions are still under discussion to prevent the misuse or inaccurate application of provided data and products, which could impact public safety, economic assets, or social structures.

The misuse of research data has direct ethical implications on research integrity and the governance of RDM. The concept of research integrity refers to a set of ethical standards that serve as the foundation for the execution of research activities. In other words, research integrity consists of core principles and ethical values determining the way to conduct and evaluate scientific research, which concerns not solely deontological obligations and professional standards adopted by individual researchers, but also the role of public institutions in promoting and performing scientific research. Safeguarding research integrity is extremely important to preserve the credibility and amplify the influence of scientific research in society, while also preventing and dealing with instances of scientific misconduct. Moreover, conducting research with public funding entails both individual and collective responsibilities, that is, as individual researchers and as members of a community of experts, which concern both the exploitation of the scientific results to ensure a science for society and the adoption of efficient, transparent, and informed policies for the governance of public scientific research. Therefore, promoting research integrity is key to guaranteeing the quality of the research and enhancing the reputation and the public image of science, contributing to progress, innovation, and science for the benefit of society. For the goals of this paper, we focus on three key implications of research integrity in providing open access to research data: i) the role of public research organizations in open data sharing, ii) the role of scientists in providing expertise and knowledge as individuals and members of a scientific community, iii) the role of research organizations in governing the RDM process. We will further discuss these implications in the following sections.

3.4 Artificial Intelligence (AI)

Although a comprehensive discussion of the ethical implications of AI on RDM is beyond the scope of this paper, it is important to highlight three key challenges that arise from the increasing use of AI and Machine Learning (ML) algorithms in research data management.

First, the preparation of datasets for AI applications introduces the risk of embedding biases during the data curation and pre-processing stages and increases the number of "customized" input data products adapted to AI applications whose quality might not be easily verified. These biases, whether intentional or unintentional, can influence the outcomes of ML models by shaping the patterns that the algorithms are designed to detect [Masuda et al., 2024; Mavrogiorgos et al., 2024]. The key issue here is the quality assessment of input data impacting the reliability of AI results.

Second, AI (in particular ML algorithms) has the potential to generate a substantial volume of new data products, further complicating the already complex task of managing, standardizing, and archiving research data. The sheer volume of AI-generated data will challenge the sustainability of RDM, especially in ensuring that these data products meet quality and interoperability standards. Both challenges highlight the need for careful governance and oversight in AI-driven research, as these technologies will also affect the credibility and authoritativeness of scientists, particularly in the realms of risk communication and decision-making. This underscores the growing ethical implications associated with the use of AI in scientific research and data management.

The third challenge, known as Explainable AI (XAI), addresses the "black-box" problem in machine learning, aiming to make AI decisions and processes more transparent and understandable [Dwivedi et al., 2023]. As AI models grow in complexity, particularly deep learning models, understanding the reasoning behind their predictions becomes challenging, even for developers. This opacity can affect the reliability and trustworthiness of AI-generated data products, particularly in research contexts where reproducibility and transparency are critical. XAI approaches, such as interpretable models or post-hoc explanations, allow researchers and stakeholders to assess the validity and robustness of AI outputs, enhancing the overall credibility of scientific research data. Implementing the ability to explain and interpret can also align AI systems with ethical standards and regulatory requirements [see Cleverley, 2024; and UNESCO Recommendations[13]], such as the GDPR, which calls for transparency in automated decision-making processes.

4. An ethical dimension of public research: impartiality for public good

Solid Earth science data represent the knowledge base to unravel the physical and chemical processes that control earthquakes, tsunamis, volcanic eruptions, unrest episodes, and Earth surface dynamics, offering answers about how to maintain Earth as a safe, prosperous, and habitable planet for human beings. The key challenge is to make the enormous wealth of available solid Earth data openly accessible to promote cross- and multi-disciplinary science for society. EPOS succeeded in harmonizing, integrating, and providing access to multidisciplinary solid Earth data through its data portal. The engaged public research organizations contribute to this innovation, which is a driver to foster excellence in science and science for society. They are committed to monitoring areas prone to geo-hazards for the surveillance of national territory, including anthropogenic hazards and the safe exploitation of geo-resources. As public research organizations, they are mandated to ensure open and FAIR access to their data and products, while also being committed to geo-hazards assessment and risk mitigation activities. This double mission defines the ethical implications, which also concern the EPOS research infrastructure. To uphold integrity and impartiality, EPOS and the involved public research organizations must remain independent of economic interests, political pressures, and prevailing public opinions. The primary aim is to advance scientific knowledge and foster technological progress for the public good. This brings an additional ethical dimension to sharing solid Earth science data:

[13] Recommendation on the Ethics of Artificial Intelligence (2023): https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (accessed 28 January 2025).

impartiality for the public good[14]. This requires transparency, authoritativeness, and an independent, third-party status to protect the common good and underscore the value of science for society. Such impartiality also influences interactions with private sector, especially when public research organizations operate surveillance services of the impact of exploitation activities. Impartiality is deeply intertwined with research integrity, intended as treating scientific knowledge as a public good, free from misconduct that could either contribute to spreading false results or cast doubts about the soundness of scientific outcomes. Upholding this integrity in research is essential to maintaining scientific excellence and public trust.

Addressing the ethical issue of impartiality for public good requires balancing among three commitments characterizing the mission of the EPOS RI and the research organizations that are part of it: i) support to Open Science and open access to research data, ii) contributions to geo-hazard assessment, risk mitigation, and communication, iii) the adoption of transparent interactions with the private sector for the safe exploitation of geo-resources. This is illustrated in Figure 3, describing the ethical issue related to impartiality, transparency, and authoritativeness in the framework of EPOS. Addressing the ethical issue of impartiality for public good means positioning research organizations at the intersection (i.e., the center of mass) of these three core commitments. This approach upholds scientific integrity, emphasizes the role of scientists, and reinforces their authority in sharing knowledge and fostering public trust in science. While applicable to many other disciplines, it is especially important in environmental sciences, including solid Earth sciences, because of their societal impacts. When personal, professional, or institutional interests - whether consciously or unconsciously - clash with the designated responsibilities and roles of involved scientists, the impartiality of the individual or the research organization is compromised. This undermines their ability to make decisions free from the influence of conflicting interests and restricts their freedom of action. This anticipates an ethical dilemma between the responsibilities of individuals (i.e., a scientist involved in his/her own research) and the responsibilities (and legal exposition) of the experts (a panel or a community) formally involved in decision-making processes for hazard assessment and risk mitigation.

Addressing impartiality for public good in compliance with the ethical principles of transparency, integrity, openness, and authoritativeness is a serious commitment for research organizations. Integrity in research is the incorporation of principles

[14] Since this article engages with concepts of data and knowledge, it is essential to clarify that 'public good' here does not follow the definition typically used in fields such as economics, which characterizes it as a good that "has two critical properties, non-rivalrous consumption—the consumption of one individual does not detract from that of another—and non-excludability—it is difficult if not impossible to exclude an individual from enjoying the good" [Stiglitz, 1999].

of honesty and respect for ethical standards and norms throughout all stages of the research endeavor, encompassing study design, data collection, analysis and interpretations, reporting to different stakeholders, and publishing. The graphical representation shown in Figure 3 exemplifies the need to find an effective equilibrium among the three formal commitments of contributing to Open Science, hazard assessment/risk mitigation, and safe exploitation of geo-resources.

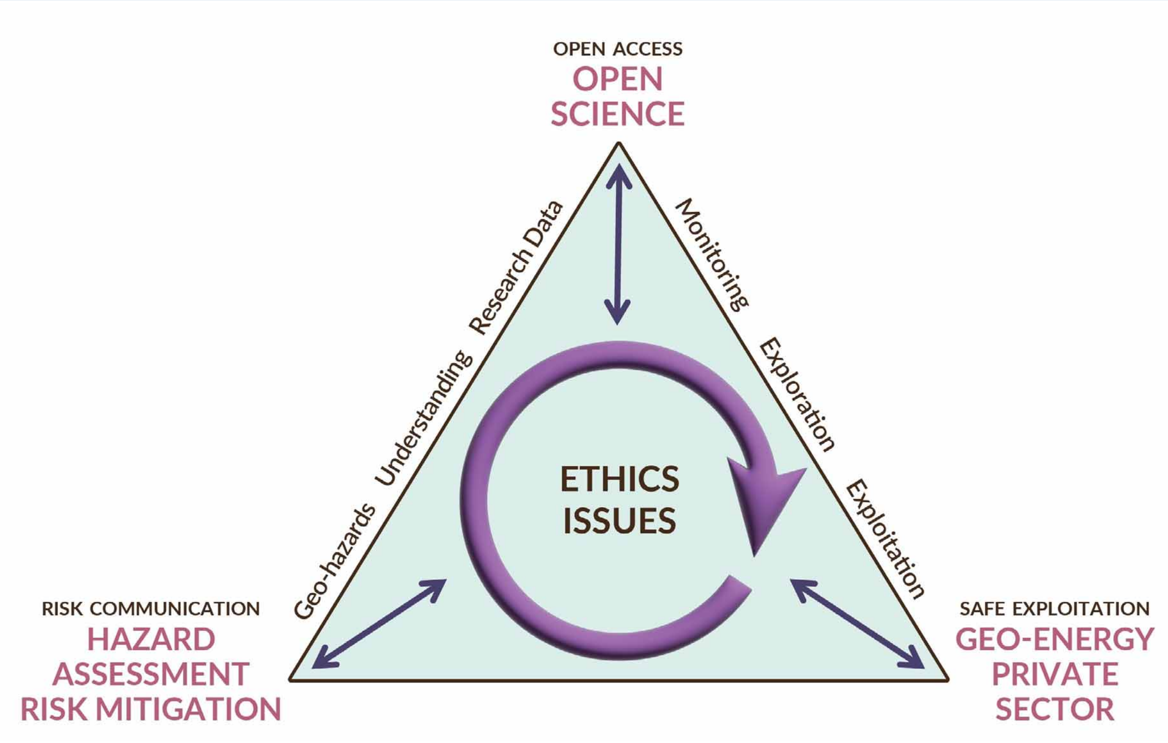

Figure 3. *Sketch representing the EPOS approach to ethics and impartiality for public good. EPOS RI is called to balance among three commitments characterizing its mission: namely, i) support to Open Science and open access to research data through effective RDM and Open Science Commons; ii) contributions to geo-hazard assessment and risk mitigation, which might also include risk communication to different stakeholders and society at large; iii) applying scientific research to the safe exploitation of geo-resources, which implies the adoption of transparent interactions with the private sector for the assessment of anthropogenic hazard.*

The equilibrium between these elements (i.e., the center of mass) depends on the weight assigned to each commitment, and this defines the ethical dimension. Impartiality for public good is not solely an ethical principle, it represents an essential value to gain authoritativeness and preserve the trust of society in science and research organizations. The integrity of science and the role of scientists are enhanced through their image in society: the more this image will account for the real benefit that science can produce for society, the more society will be able to attribute to science its fundamental function.

Because of the commitments of research organizations providing services and expert opinions for decision-making concerning geo-hazards, EPOS ERIC must maintain an external (scientific) advisory role in the decision-making process, based on the data, products, and services at its disposal, and must never assume decision-making tasks [Di Capua and Peppoloni, 2022]. The research organizations play a dual role as both scientific data providers and expert advisors to decision makers in the framework of formalized protocols (and legal frameworks) with national and regional authorities. Moreover, they are also involved in the communication to society of geo-hazards and associated risks directly contributing to disaster risk management. In this framework, research organizations and EPOS must balance between two ethical aspects: the necessity to inform society and the commitment to respect the existing protocols and obligations for contributing to decision-making. This can represent an ethical dilemma between the freedom of individual scientists to inform society about the societal impact of their own investigations (ideally following the ethical guidelines of various scientific societies and unions) and the collective responsibility of a community of experts to provide expert opinions for decision-making within official legal protocols and frameworks. This dilemma is even more relevant and impactful nowadays because of the legal exposition of the scientists involved in expert panels for decision-making. The events related to prosecutions involving scientists engaged in providing expert opinions to support decision-making and emergency management (such as those of the trials that followed the 2009 L'Aquila earthquake in Italy and the 2010 Maule earthquake in Chile) confirm how today the boundary between science, communication and criminal law is blurred and influenced by factors that are difficult to evaluate for the purposes of defining the legal exposure of experts involved in decision-making [Cocco et al., 2015; Stucchi et al., 2016].

5. Access to research products and scientific information

Open Science extends beyond data access, fostering collaboration and knowledge sharing through digital tools [European Commission, 2016b; Vicente-Saez et al., 2018]. In the framework of Open Science, it is relevant to discuss the sustainability and the ethical implications of providing access to scientific information and knowledge, beyond open access to research data for scientists. This is particularly important for pan-European Research Infrastructures. Although the value of public scientific research is often assessed in terms of technological development and economic growth, public scientific research has a societal value which deserves greater recognition. Re-usable (FAIR) scientific products can generate important benefits for society, the environment, and the economy, in particular, because of their suitability for the creation of value-added services for many potential beneficiaries and new applications for assessing natural and anthropogenic hazards and mitigating the associated risks. The value of these scientific products is also measured by the sharing of the associated scientific information through tailored communication that considers the literacy of stakeholders (i.e., policymakers, decision-makers, society at large). This is a challenging task.

To this end, it is essential to distinguish access to research data from access to scientific information and knowledge [Stewart and Lewis, 2017; Ye and Ma, 2023]. Open Science and open access to research data require data management and the governance of this process, involving the whole data lifecycle. In this framework, the sharing of scientific information concerns the communication of the contents, the use, and the application of research data to progress in science and innovation for a science for society. Making scientific information available to stakeholders and society at large is an essential step to improve the understanding of scientific methods, scientific products, and their use and applications to evidence-based decision-making. This is a key contribution to improving the literacy of citizens and stakeholders on the use of scientific data. The sharing of scientific information involves structured communication, starting from the well-calibrated dissemination of the contents and the usage of research data in scientific investigations up to bidirectional communication of research data interpretations and findings to different stakeholders. While this can be seen as an ethical obligation for a public research organization and an international research infrastructure, it is further increasing the sustainability challenge. Research infrastructures like EPOS do not have the human and technological resources and the required skills to address this task in an ethically coherent way. It would require engaging professional science communicators capable of portraying scientific methodologies and the continuous monitoring of the usage of shared research data. However, establishing an open and inclusive dialogue with the public requires that the topics are accessible (i.e., use of lay language) to ensure a basic understanding of the research approach and its integrity. This requires a shift from basic communication to increased audience engagement, where citizens' knowledge (i.e., literacy) is seen as a societal asset and potential for innovation. This is not currently affordable for pan-European research infrastructures that struggle to operate sustainable access to research data and FAIR data management. Research organizations in solid Earth science provide services for communicating the added value of their research outputs and products, but science communication in modern society is extremely demanding, complicating the challenge of sustainability of Open Science. This makes the ethical commitment to share both research data and scientific information not affordable with sustainable practices. Moreover, if the sharing of research data among scientists requires taking into consideration expected ethical standards for research, sharing and communicating scientific information to different stakeholders, not limited to scientists, makes the ethical framework even more complex and less standardized.

The sharing of knowledge for decision-making and related communication moves the target even further the current approaches to sharing research data and the envisioned approaches to sharing scientific information. Evidence-based decision-making involves experts providing expertise to interpret scientific evidence and outcomes as well as their knowledge to feed decision and policy making. This is especially evident in disaster risk management, where decisions are taken based on timely and reliable scientific knowledge. Exploiting the impact of Open Science and Open Access to research data for evidence-based decision-making requires a full understanding of this complex and multi-process framework as well as awareness of the necessary resources, roles, and responsibilities [Peppoloni et al., 2023]. EPOS is not formally and directly involved in providing expert opinions for decision-making, but the innovation in sharing new research products and their applications to geo-hazards and related risks will be used by the committed national research organizations when providing their formal contributions to evidence-based decision-making. Here we are considering the role and responsibilities of national research organizations, but we also need to mention individual scientists (experts) who use the EPOS data products to provide their expert opinions to public and private decision-makers. The problem for a research infrastructure like EPOS is that these benefits for society are intangible and difficult to measure for the assessment of socio-economic impact.

Finally, ethics also involves "*the study of the general nature of morals and of the specific moral judgments or choices to be made by a person*" [Burns, 2024]. This statement situates ethics as a matter of individual choice, but of course, the choices we make as individuals have broad impacts on the community, we are part of and serve. That is, ethics is often practically framed as the result of individual choices and actions, but ethics also encompasses the implicit and explicit values of an institution, community of practice, or even group of researchers [Weber and Locke, 2022; Di Capua and Peppoloni, 2022]. The experiences discussed in this article for sharing solid Earth science data exemplify the complexity of the ethical framework of Open Science and might suggest considering experienced practices when formulating and implementing general principles for Open Science.

6. Conclusive remarks

The ethical dimension of sharing and providing access to solid Earth science data is investigated in this article through three reference frameworks, which have different implications for Open Science.

i. Responsibilities in addressing well-known ethical issues related to the protection of personal data, intellectual property rights and misuse of research data. While the former are well defined through laws (i.e., GDPR), rules, and viable solutions (i.e., data and metadata licensing), the latter relies on the definition of misuse of research data and has implications on the adoption of the principle “open as possible, restricted as necessary”.
ii. Implementation of the general principle for public research organizations, including research infrastructures, to ensure research integrity and impartiality for the public good. This is particularly important for the sharing of research data from solid Earth science related to geo-hazards and risk mitigation.
iii. Ethical implications of using research data and scientific information for decision-making, and in particular, the role of experts (i.e., scientists) committed to providing expert opinions, interpretations, and knowledge within official frameworks (defined by policymakers) aimed at making decisions affecting the life, security, and risks of citizens.

While open access to research data, products, and services for science is the key mission of pan-European research infrastructures like EPOS, the sharing of scientific information moves the mission objectives and the target well beyond, involving communication to society through suitable languages and tools suited to the literacy of stakeholders. This makes the challenge of long-term sustainability even more difficult and demanding. Moreover, the sharing of knowledge for decision-making relies on formal codified procedures proposed by decision-makers and shared with engaged scientists (i.e., the experts) from committed research organizations, further characterizing the ethical dimension of Open Science.
In this context, research infrastructures like EPOS are tasked with responding to the need of the scientific community to bridge the ever-widening gap between scientific rigor and societal behavior and expectations, which have become increasingly significant. Although EPOS functions as a unified entity, it embodies diverse communities that bring together individuals from various countries, academic disciplines, national research institutions, universities, and other organizations. An important factor to address is therefore the organizational culture within this multifaceted entity, which necessitates fostering ethical behavior as a

key element in effectively overseeing the entire organization and its strategic and operational risks. EPOS's dedication to social responsibility is vital to its mission, given that it relies on public investments [OECD, 2017]. As a result, ethical considerations are crucial to the EPOS framework, significantly influencing interactions with National Research Infrastructures, Thematic Core Services, and the management of Integrated Core Services and the EPOS ERIC Executive Coordination Office. To address this challenge and ensure that the Consortium's decisions and activities are rooted in integrity and ethical values, EPOS established a dedicated Ethics Board within the EPOS ERIC governance framework. The EPOS Ethics Board's mandate explicitly addresses aspects such as data misuse and abuse, personal data protection, impartiality concerning public good, and communication and societal impacts. Moreover, recognizing the link between science and ethics, EPOS ERIC has established a dedicated Working Group on Ethics in 2024 with the mandate of defining the EPOS Ethical Guidelines and related documents, raising awareness and identifying ethical issues to be reported and discussed with the Ethics Board. The EPOS ERIC perspective emphasizes the necessity of establishing, improving, and upholding an ethical framework while considering the requirements of researchers and operators within the EPOS Research Infrastructure, particularly those concerning Open Science.

Discussing the ethical dimension of sharing solid Earth science data requires contextualizing the Open Science principle "as open as possible, as closed as necessary". Recently, the Council of the European Union recommended the Member States of the European Union and the Commission to enhance the level of security related to research activities due to the growing international tensions and the increasing geopolitical relevance of research and innovation[15]. The Council emphasized that "*with academic freedom comes academic responsibility*" warning against the undesirable transfer of critical knowledge and technology to those third countries that may impact the security of the Union and its Member States, or for purposes that are in violation of Union values and fundamental rights. This gives a new context to the principle "open as possible, restricted as necessary", in particular to the interpretation of the term "restricted" as well as to the motivation of the term "necessary". This can also broaden the definition of research data misuse and implicate scientists involved in data sharing in national security matters. This is particularly relevant for Earth sciences. For instance, the sharing of satellite images used for different scientific purposes (from measuring surface deformation to monitoring land usage of critical areas) might put national security at risk by identifying possible targets for terrorist actions. The ethical dilemma here is the

[15] Council Recommendation of 23 May 2024 on enhancing research security: https://eur-lex.europa.eu/eli/C/2024/3510/oj (accessed 28 January 2025).

involvement of scientists in the management of national security, thus placing a limit on the sharing of research data for matters not strictly related to science. The current European Union legislation sets these boundaries relying on different elements comprising well-recognized and managed issues (such as protection of personal data, protection of commercially sensitive data, protection of IPR) and more general politically-driven issues (such as public safety, defense or national security, sensitive information relating to the protection of critical infrastructures). Issues related to the health and safety of citizens (mentioned by EU regulations) are related to the geo-hazards assessment and risk mitigation, but solid Earth scientists are addressing these issues in the framework of Open Science (following the vision "as open as possible"). There are also increasing discussions on the negative impact of the use of Open Data exploited by machine learning algorithms for malicious purposes and the European Commission is investigating the issue[16]. The floating boundary between "possibly open" and "necessarily restricted" creates difficulties in preserving the sharing of data used for the mitigation of natural and anthropogenic risks and the services associated with emergency management. It is necessary to affirm in this context the need to keep the data open by balancing the *possibility* in terms of awareness, preparedness and resilience of the population to manage the impact of scientific information. In other words, enlarging the open access to research data by strengthening the literacy of citizens by enabling access to scientific information. Although this strengthens the added value of public research for society also in ethical terms, we must be aware that the sustainability of these practices is not achievable even in the short and medium term.

Acknowledgements. This research was partially supported by the Skills4EOSC project, funded by the European Commission and UK Research and Innovation (UKRI) under UE Grant (101058527) and UKRI's (10040140). Multiple authors are or have been members of the Skills4EOSC team, and knowledge developed within the framework of the project has significantly contributed to the insights presented in this article. We are grateful for the support provided by the Skills4EOSC project, the coordinators, and its members. EPOS ERIC also supported the research on these topics since the EPOS implementation phase. We are grateful to the colleagues working in the EPOS Thematic communities (TCS) as well as in the Service Coordination Board. Exchanges of ideas with the past and current EPOS Ethics boards have

[16] Commission Recommendation (EU) 2023/2113 of 3 October 2023 on critical technology areas for the EU's economic security for further risk assessment with Member States: https://eur-lex.europa.eu/eli/reco/2023/2113/oj (accessed 28 January 2025); Council Recommendation of 23 May 2024 on enhancing research security: https://eur-lex.europa.eu/eli/C/2024/3510/oj (accessed 28 January 2025); The European Cyber Resilience Act (CRA), 2023: https://www.european-cyber-resilience-act.com/ (accessed 28 January 2025).

been extremely useful and inspiring for this study. Finally, we thank Barbara Angioni for drafting the figures and for graphic assistance.

References

APEGGA, (2010). *Guideline for Ethical Use of Geophysical Data*. The Association of Professional Engineers, Geologists and Geophysicists of Alberta. Retrieved from: https://www.apega.ca/docs/default-source/pdfs/standards-guidelines/geophysical-data.pdf (accessed 28 January 2025).

Atakan K., Cocco M., Orlecka-Sikora B., Pijnenburg R., Michalek J. et al., (2022). *National EPOS initiatives and participation to the EPOS integration plan*. Annals of Geophysics, 65(2). https://doi.org/10.4401/ag-8758

Bailo D., Paciello R., Sbarra M., Rabissoni R., Vinciarelli V. et al., (2020). *Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructure*s. Frontiers in Earth Science, 8. https://doi.org/10.3389/feart.2020.00003

Bailo D., Jeffery,K.G., Atakan K., Trani L., Paciello R. et al., (2022). *Data integration and FAIR data management in Solid Earth Science*. Annals of Geophysics, 65(2). https://doi.org/10.4401/ag-8742

Bailo D., Paciello R., Michalek J., Cocco M., Freda C. et al., (2023). *The EPOS multi-disciplinary Data Portal for integrated access to solid Earth science datasets*. Scientific Data, 10:784. https://doi.org/10.1038/s41597-023-02697-9

Bailo D., Paciello R., Michalek J., Mercurio D., Sangianantoni A. et al., (2024). *Data Management in Distributed, Federated Research Infrastructures: The Case of EPOS*. Data Science Journal, 23. https://doi.org/10.5334/dsj-2024-005

Burns S.A., (2024). *Evolutionary Pragmatism, A Discourse on a Modern Philosophy For the 21st Century.* Different Kinds of Ethics. Retrieved from: https://evolutionary-pragmatism.com/pg0405/ (accessed 28 January 2025).

Carroll S.R., Garba I., Figueroa-Rodríguez O.L., Holbrook J., Lovett R. et al., (2020). *The CARE Principles for Indigenous Data Governance*. Data Science Journal, 19. https://doi.org/10.5334/dsj-2020-043

Cleverley P.H., (2024). *Ethical recommendations for Artificial Intelligence technology in the Geological Sciences - with a focus on Language Models*. Journal of Geoethics and Social Geosciences, 1(2). https://doi.org/10.13127/jgsg-63

Cocco M., Cultrera G., Amato A., Braun T., Cerase A. et al., (2015). *The L'Aquila trial*. In Peppoloni, S., and Di Capua, G., (eds.), Geoethics: the Role and Responsibility of Geoscientists, Geological Society, London, Special Publications, 419, pp. 43-55. https://doi.org/10.1144/SP419.13

Cocco M., Freda C., Atakan K., Bailo D., Saleh Contell K. et al., (2022). *The EPOS*

Research Infrastructure: a federated approach to integrate solid Earth science data and services. Annals of Geophysics, 65(2). https://doi.org/10.4401/ag-8756

Cocco M., and Montone P., (2022). *EPOS a Research Infrastructure in solid Earth: open science and innovation*. Preface to the Special Issue dedicated to EPOS. Annals of Geophysics, 65(2). https://doi.org/10.4401/ag-8835

CODATA, (2024). *CODATA RDM Terminology (2023 version): overview*. RDM Terminology Working Group (2024). Zenodo. https://doi.org/10.5281/zenodo.10626170

Di Capua G., and Peppoloni S., (2022). *D6.2 – Ethical Guidelines for EPOS Services to society*. EPOS SP – Grant Agreement no. 871121. Retrieved from: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ef3d11b8&appId=PPGMS (accessed 28 January 2025).

Dwivedi R., Dave D., Naik H., Singhal S., Omer R. et al., (2023). *Explainable AI (XAI): Core Ideas, Techniques, and Solutions*. ACM Computing Surveys, 55(9). https://doi.org/10.1145/3561048

ESFRI, (2017). *Long-Term Sustainability of Research Infrastructures*. European Strategy Forum on Research Infrastructures Long-Term Sustainability Working Group. ESFRI Scripta, Vol. 2. Retrieved from: https://www.esfri.eu/sites/default/files/ESFRI_SCRIPTA_SINGLE_PAGE_19102017_0.pdf (accessed 28 January 2025).

EUI, (2022). *Guide on Good Data Protection Practice in Research*. European University Institute. Retrieved from: https://www.eui.eu/documents/servicesadmin/deanofstudies/researchethics/guide-data-protection-research.pdf (accessed 28 January 2025).

European Commission, (2016a). *Background Note on Open Access to Scientific Publications and Open Research Data*. EU Directorate-General For Research and Innovation, https://research-and-innovation.ec.europa.eu/document/download/4bd9ef8e-0101-457d-9fc5-1096c4e8f6f0_en?filename=ec_rtd_background-note-open-access.pdf (accessed 28 January 2025).

European Commission, (2016b). *Open innovation, open science, open to the world – a vision for Europe*. EU Directorate-General for Research and Innovation. https://doi.org/10.2777/061652

Locati M., (2024). *Definitions of terms extracted from data-related European Union laws, version 2 [Dataset]*. Zenodo. https://doi.org/10.5281/ZENODO.12800223

Locati M., Tanlongo F., Evangelinou B., Dotta G., Cocco M. et al., (2024). *Guidelines and best practices for Honest Brokers: Harnessing Open Science to support Evidence-Based Decision Making*. EC Project Skills4EOSC. Zenodo. https://doi.org/10.5281/ZENODO.12758995

Marti M., Haslinger F., Peppoloni S., Di Capua G., Glaves H. et al., (2022). *Addressing the challenges of making data, products, and services accessible: an EPOS* *perspective*. Annals of Geophysics, 65(2). https://doi.org/10.4401/ag-8746

Masuda A., (2024). *Dirty data: an opportunity for cleaning up bias in AI*. The Current, Science + Technology, University of California, Santa Barbara. Retrieved from: https://news.ucsb.edu/2024/021521/dirty-data-opportunity-cleaning-bias-ai (accessed 28 January 2025).

Mavrogiorgos K., Kiourtis A., Mavrogiorgou A., Menychtas A., and Kyriazis D., (2024). *Bias in Machine Learning: A Literature Review*. Applied Sciences, 14(19). https://doi.org/10.3390/app14198860

OECD, (2017). *Strengthening the effectiveness and sustainability of international research infrastructures*. OECD Science, Technology and Industry Policy Papers, No. 48, OECD Publishing, Paris. https://doi.org/10.1787/fa11a0e0-en

Paic A., (2021). *Open science - Enabling discovery in the digital age*. OECD Going Digital Toolkit Notes, No. 13, OECD Publishing, Paris. https://doi.org/10.1787/81a9dcf0-en

Pasquetto I.V., Cullen Z., Thomer A., and Wofford M., (2024). *What is research data "misuse"? And how can it be prevented or mitigated?* Journal of the Association for Information Science and Technology, 75(12), 1413-1429. https://doi.org/10.1002/asi.24944

Peppoloni S., Woo G., Martí J., and Di Capua G., (2023). *Geoethics: The Missing Piece in the Separation of Responsibility Between Volcanologists and Decision-Makers*. In Malheiro, A., Fernandes, F., and Chaminé, H.I., (Eds.), Advances in Natural Hazards and Volcanic Risks: Shaping a Sustainable Future. NATHAZ 2022. Advances in Science, Technology & Innovation. Springer, Cham, pp. 19-23. https://doi.org/10.1007/978-3-031-25042-2_4

Petzold A., Asmi A., Vermeulen A., Pappalardo G., Bailo D. et al., (2019). *ENVRI-FAIR - Interoperable Environmental FAIR Data and Services for Society, Innovation and Research*. 15th International Conference on eScience (eScience), San Diego, CA, USA, 2019, pp. 277-280. https://doi.org/10.1109/eScience.2019.00038

Quintais J.P., (2017). *Rethinking Normal Exploitation: Enabling Online Limitations in EU Copyright Law*. AMI - tijdschrift v oor auteurs-, media- en informatierecht 2017/6, 197-205. Retrieved from: https://ssrn.com/abstract=3106729 (accessed 28 January 2025).

Saleh Contell K., Karlzén K., Cocco M., Pedersen H.A., Atakan K. et al., (2022). *Long-term sustainability of a distributed RI: the EPOS case*. Annals of Geophysics, 65(2). https://doi.org/10.4401/ag-8786

Sinaeepourfard A., Masip-Bruin X., Garcia J., and Marín-Tordera E., (2015). *A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges*. Technical Report (UPC-DAC-RR-2015-18). Retrieved from: https://www.ac.upc.edu/app/research-reports/html/RR/2015/18.pdf (accessed 28 January 2025).

Schirru L., Colcelli V., Brizioli S., Karabuga,E., Fernandes E. et al., (2023). *D3.6 -*

Coordinated set of guides, fact-sheets and FAQs on ELSI aspects for Civil Servants and Policy Makers. EU project Skills4EOSC deliverable. Zenodo. https://doi.org/10.5281/zenodo.7991444

Stewart I.S., and Lewis D., (2017). *Communicating contested geoscience to the public: Moving from 'matters of fact' to 'matters of concern'*. Earth-Science Reviews, 174, 122-133. https://doi.org/10.1016/j.earscirev.2017.09.003

Stiglitz J.E., (1999). *Knowledge as a Global Public Good*. In Kaul, I., Grunberg, I., and Stern, M., (Eds.), Global Public Goods: International Cooperation in the 21st Century, Oxford University Press, New York, pp. 308-325. https://doi.org/10.1093/0195130529.003.0015

Stucchi M., Pinho R., and Cocco M., (2016). *After the L'Aquila Trial*. Seismological Research Letters, 87(3), 591-596. https://doi.org/10.1785/0220150261

UK Research and Innovation, (2016). *Concordat on Open Research Data*. Retrieved from: https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf (accessed 28 January 2025).

UNESCO, (2022). *Understanding open science*. UNESCO open science toolkit. https://doi.org/10.54677/UTCD9302

Vicente-Saez R., and Martinez-Fuentes C., (2018). *Open Science now: A systematic literature review for an integrated definition*. Journal of Business Research, 88, 428-436. https://doi.org/10.1016/j.jbusres.2017.12.043

Weber N., and Locke B., (2022). *Ethics of Open Data*. Computers and Society, arXiv:2205.10402v1. https://doi.org/10.48550/ARXIV.2205.10402

Wilkinson M.D., Dumontier M., Aalbersberg Ij.J., Appleton G., Axton M. et al., (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 3. https://doi.org/10.1038/sdata.2016.18

WIPO, (2020). *What is intellectual property*. World Intellectual Property Organization. Retrieved from: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_450_2020.pdf (accessed 28 January 2025).

Ye F.Y., and Ma F.-C., (2023). *An essay on the differences and linkages between data science and information science*. Data and Information Management, 7(1). https://doi.org/10.1016/j.dim.2023.100032

*Corresponding author: **Massimo Cocco**

e-mail: massimo.cocco@ingv.it