D13.3 Digital Rights Management State of the Art

Design Considerations for an EOSC Rights Vocabulary

30/09/2025

Abstract

This report offers a detailed examination of the current landscape of Rights Markup Languages (RMLs) and their significance for the European Open Science Cloud (EOSC). It reviews current practices in representative repositories and data infrastructures, and makes recommendations for the development of an EOSC Rights Vocabulary, addressing gaps that were identified in the analysis.

Funded by the European Union

EOSC Beyond receives funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101131875

Document Description

D13.3 Digital Rights Management State of the Art			
WP13			
Due date	30/09/2025	**Actual delivery date:**	29/09/2025
Nature of document	Report	**Dissemination level**	Public
Document Status	Under EC review	**Version**	1.0
Lead Partner	OpenAIRE		
Authors	M. Katsanakis (OpenAIRE), Wim Hugo (DANS), Deborah Thorpe (DANS)		
Reviewers	Germán Moltó (UPV), Montserrat González (EGI), Marjan Grootveld (DANS)		
Approved by	TCB		
Document link	https://documents.egi.eu/document/4129		
DOI	https://zenodo.org/records/17225379		
Keyword	Rights management, GDPR, ethics, copyright, access		

Revision History

Issue	Date	Description	Author/Reviewer
V 0.1	27/01/2025	Confirm Table of Contents	W Hugo, M Katsamakis
V 0.2	16/06/2025	Initial draft	M Katsamakis, P Tsiavos
V 0.3	30/06/2025	Elaboration and scope completion	M Katsamakis
V 0.4	08/07/2025	Case study additions, comments	D Thorpe
V 0.5	28/07/2025	Small amendments, additions, and comment resolution	D Thorpe, M Katsamakis, W Hugo
V 0.6	20/08/2025	Case study additions, high-level considerations, trustworthy repository, TRUST, and CARE-related aspects	D Thorpe, W Hugo
V 0.7	27/08/2025	Final WP review, editorial tasks, document structuring	W Hugo, D Thorpe, M Katsamakis
V 0.8	01/09/2025	Transfer to deliverable template	W Hugo
V 0.9	03/09/2025	Release for internal review	W Hugo
V.0.10	19/09/2025	Completion of internal review	G Moltó (UPV), M González (EGI) M Grootveld (DANS)
V 1.11	24/09/2025	Corrections	D Thorpe, M Katsamakis, W Hugo
V 1.0		**Final**	

Table of Content

List of Tables

Executive Summary

This report offers a detailed examination of the current landscape of Rights Markup Languages (RMLs) and their significance for the European Open Science Cloud (EOSC). It highlights and assesses key standards such as ODRL, ccREL, MPEG-21, and schema.org, evaluating how well they can represent and enforce rights, restrictions, and permissions in formats that machines can understand.

Additionally, the study investigates the management of rights in prominent EOSC-aligned repositories and aggregators, including Zenodo, DANS, the European Genome–Phenome Archive (EGA), Europeana, and the OpenAIRE Graph. This study highlights critical gaps that still exist in the legal, ethical, and technical aspects of rights governance, gaps that often block automation and reduce interoperability. These insights led us to formulate a hands-on solution for classifying and encoding rights relevant to EOSC, which encompasses key categories such as intellectual property, personal data, and ethical concerns.

The concluding chapters consolidate the findings and provide practical recommendations to support EOSC in establishing a robust, machine-actionable rights framework. This involves working with existing infrastructures, developing tailored RML extensions for the EOSC, and outlining a clear strategy for integrating these improvements into metadata services and the EOSC Interoperability Framework (IF). The work will be followed up in the EOSC Beyond project with development of a candidate RML Vocabulary for EOSC, with a target date of September 2026, and pilot adoption of the vocabulary by selected participating institutions.

It is our objective with the EOSC Rights Vocabulary to not only harmonise and standardise the terms used for description of rights provisions—permissions, obligations, and restrictions—but to also allow the flexible implementation of rights definition in data infrastructures and repositories. This includes definition of new licenses and consistent implementation at granularities ranging from individual files or objects to any level of composite sets. It also includes unifying practices that assign individual provisions to an object with the more common practice of assigning a licence to an object.

Glossary of Rights-Related Terms

Term	Definition
CARE Principles	A set of ethical standards for data governance, especially for Indigenous and marginalised communities: Collective Benefit, Authority to Control, Responsibility, Ethics
ccREL	Creative Commons Rights Expression Language. A REL for descriptive metadata to be appended to media that is licensed under any of the Creative Commons licenses.
DCAT	Data Catalog Vocabulary. A RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web
DRM	Digital Rights Management
DUO	Data Use Ontology: an ontology which represent data use conditions
Embargo	A time-based restriction that delays open access to a digital resource, often used to respect publishing policies or protect sensitive data during initial periods
EOSC	European Open Science Cloud. Objective to provide researchers and innovators in Europe with an open and trusted multi-disciplinary environment where they can publish, find and reuse data, tools and services for research and innovation
EU	European Union
FAIR Principles	Guidelines to ensure that digital objects are Findable, Accessible, Interoperable, and Reusable—supporting transparency and data reusability
GA4GH	Global Alliance for Genomics and Health, develops standards for responsibly collecting, storing, analysing, and sharing genomic data
GDPR	General Data Protection Regulation. A legal framework of seven principles established across the EU to protect personal data
ISO	The International Organization for Standardization
JSON-LD	A method of encoding linked data using JSON (JavaScript Object Notation)
Licence Interoperability	The capacity of different licences to be combined or reused together without legal conflict—crucial when mixing datasets or software with varying rights

Machine-Readable Rights	Rights information structured in a standardised format (e.g., JSON-LD, RDF) so that software systems can interpret, enforce, or reason over them automatically
MPEG-21 REL	This standard, from the Moving Picture Experts Group, aims at defining an open framework for multimedia applications
Obligation	A condition requiring users to perform specific actions, such as attributing the creator or sharing derivatives under the same licence
ODRL	Open Digital Rights Language. Provides flexible and interoperable mechanisms to support transparent and innovative use of digital content in publishing, distribution, and consumption of digital media across all sectors and communities.
Prohibition	A rule in a rights statement that explicitly forbids certain actions, such as commercial use or data redistribution
RDF	Resource Description Framework, a standard model for data interchange on the Web.
REL	Rights Expression Languages
SEO	Search Engine Optimisation
TRUST	The 5 aspects the authors argue should be covered to make a repository trustworthy: Transparency, Responsibility, User focus, Sustainability, Technology
Use Case	A real-world scenario that illustrates how rights metadata is applied in practice to solve a specific challenge or meet a stakeholder need
XML	Extensible Markup Language a markup language and file format for storing, transmitting, and reconstructing data
W3C	The World Wide Web Consortium: develops standards and guidelines to help everyone build a web based on the principles of accessibility, internationalization, privacy and security.

1. Introduction

1.1. Open Science and the Problem of Rights

Open Science is reshaping research practices by promoting broad sharing of data, software, and publications. Under initiatives like the European Open Science Cloud (EOSC), openness is now an expectation and even a mandate – guided by the principle of making outputs "as open as possible, as closed as necessary"[1]. In practice, this means that while researchers are encouraged to openly share their results, there are legitimate reasons (commercialisation, privacy, subject protection, security) to impose certain access conditions. Balancing openness with necessary restrictions has elevated the importance of rights management in the Open Science ecosystem [32].

Traditionally, Digital Rights Management (DRM) in content distribution is focused on locking down assets via technical measures (encryption, access control), treating usage as something to tightly restrict. In contrast, Open Science and the FAIR principles [10] demand a more nuanced approach: instead of simply preventing access, the goal is to clearly specify how and under what conditions reuse is permitted. For example, a dataset might be freely accessible for academic research but not for commercial use, or software code may be reused under the condition that derivative works carry the same license. Rather than secretive DRM, the emphasis is on transparent, machine-readable documentation of permissions and obligations – embodying a shift toward openness with accountability. What should be avoided is the common practice of non-transparent, arbitrary access conditions.

The challenge is that research outputs carry a mosaic of rights and obligations. A single dataset or service can be subject to multiple legal and ethical frameworks, creating a complex web of conditions. These include:

- Intellectual property rights like copyright or database rights, which dictate how others may copy, modify or redistribute content;
- License terms chosen by creators or publishers (e.g. a Creative Commons [60] or open-source software license) that grant specific reuse permissions;
- Privacy and data protection laws (in the EU, the GDPR [19] and Digital Services Act [40] are the main legal considerations), when personal or sensitive data are involved, imposing obligations such as obtaining consent or restricting access to authorised persons; and
- Ethical or community guidelines, such as the CARE principles for Indigenous data [9], which call for respecting the authority and interests of Indigenous communities in how their data and knowledge is applied and used.
- The expectations listed by FAIR and TRUST [11] in general, and elaborated for example by trustworthy repository certification authorities [23] in some cases, deal with a variety of considerations – the need for transparent and standardised licences and access conditions, verification of ethics clearance, maintenance of subject rights and confidentiality of sensitive data, frequent re-appraisal, etc.

[1] rea.ec.europa.eu

- There is an emerging emphasis on data sovereignty [33] that is also in conflict with the objectives of Open Science, since it may lead to situations where reuse of resources is limited, for example, on the basis of location or nationality.

In specialised domains, additional constraints apply – for example, genomic data archives such as The **European Genome–phenome Archive** (EGA) require researchers to obtain approval from Data Access Committees before accessing sensitive human data[2].

This multiplicity of rights regimes can lead to ambiguity: what is allowed in one context may be forbidden or unclear in another. Ensuring that all these conditions are honored when data is shared or combined is a non-trivial problem. A recent small scale survey study found that data access conditions are often not made clear to potential reusers of data, and also revealed that many data depositors have not, themselves, fully defined or understood the conditions for data access (and also how multiple conditions might interrelate) before depositing their data [85].

A key issue is the fragmentation and inconsistency of how rights information is managed across the research landscape. Different repositories and platforms document usage rights in different ways. One repository might attach a standard license URL (using a schema.org or DataCite field) indicating, for example, *CC BY 4.0*, while another uses a custom text field "Terms of Use" (which may also indicate a licence), and a third one might label the resource with a rights category such as *"restricted access"*. These descriptions are not automatically equivalent or even compatible. For instance, Europe's cultural heritage platform ***Europeana*** uses a limited set of standardised rights statements (from **RightsStatements.org** [18]) such as "In Copyright" or "No Known Copyright" to uniformly flag the reuse status of digital objects. This provides clarity within Europeana's portal, but those broad labels lack the nuance to convey specific license conditions.

Conversely, a dataset in an institutional repository might include a detailed natural-language disclaimer or a unique user agreement. When such content is aggregated in the EOSC or similar cross-platform environments, the semantic misalignment of rights metadata becomes apparent: consumers of the metadata (both humans and machines) cannot easily interpret and reconcile the different expressions of essentially the same rights information [5]. The result is that researchers and EOSC providers face uncertainty and additional manual effort to clarify what they are allowed to do with a given resource.

This lack of clarity and consistency undermines one of the pillars of Open Science: trust. Researchers need confidence that when they share data, it will be reused in accordance with their conditions, and likewise, when they reuse others' data, they will not unwittingly violate licenses or privacy laws. Without a clear, standardised way to communicate rights, users may err on the side of caution and under-utilise valuable resources, or worse, reuse data in unlawful or unethical ways due to misunderstandings. The aforementioned small scale study [85] found that ambiguity about access can introduce inequality and prejudice as data depositors may make decisions about reuse based on ill-defined or ad-hoc conditions such as: "is the potential reuser one of my colleagues?" or "will the reuser give me co-authorship?"

[2] elixir-europe.org

The EOSC vision of a federated ecosystem hinges on legal and ethical interoperability alongside technical interoperability. Every digital object in EOSC should ideally carry *rights metadata* that travels with it – a machine-interpretable record of what's permitted and required. Indeed, the EOSC Strategic Implementation Plan explicitly identifies the need for common rules defining the rights and obligations of participants in the cloud to enable seamless transactions[3]. In summary, as Open Science becomes the norm, solving "the problem of rights" – how to effectively manage and communicate usage rights and restrictions – has become an essential task.

While this deliverable (D13.3) focuses on the state-of-the-art assessment and the foundational analysis required for building a rights' vocabulary, extensive case studies and implementation examples are intentionally not included at this stage. These will be addressed in detail in the upcoming Deliverable D14.3, which will focus on the final vocabulary specification, practical adoption scenarios, and integration into EOSC services.

1.2. Rights Metadata as Infrastructure

Addressing the above challenges requires treating rights information as a first-class component of research infrastructure. In the same way that metadata about provenance, authorship, or file format is handled in a structured, standard manner, rights metadata should be part of the core architecture of Open Science platforms. Rather than ad-hoc text statements or buried terms of use, rights metadata must be embedded into repositories, catalogs, and workflows in a consistent format that both humans and machines can understand and act upon. For the EOSC, this is not just a feature but a necessity: without a shared approach to rights metadata, the "cloud" cannot function as a seamless environment for sharing and reusing resources under well-defined conditions. Recognising this, stakeholders have begun to define rights metadata frameworks as integral to interoperability. For example, the EOSC Interoperability Framework [66] foresees a legal layer where usage policies and licenses are expressed in standard ways across services[4]. In practical terms, this means a researcher looking for data in an EOSC portal should be able to filter or auto-detect datasets that are reusable for their purpose (e.g. "show me data I can use commercially") based on the attached rights metadata, and if they integrate data from multiple sources, the system should flag any licence conflicts automatically[5].

Fortunately, there is a foundation to build on. Over the past decade, several Rights Expression Languages (RELs) and metadata vocabularies have been developed to formalise how usage rights are described [74]. One prominent example is the W3C's Open Digital Rights Language (ODRL) [4], [76], a policy expression language which provides a flexible and interoperable model (with an associated vocabulary) for representing permissions, prohibitions and obligations on digital content. As a W3C Recommendation, ODRL defines a standard RDF-based syntax for policies, meaning that an ODRL policy can be attached to a digital object (via a link or embedded metadata) and interpreted unambiguously by any tool

[3] esfri.eu

[4] https://zenodo.org/records/11104269

[5] Services to assist with this are emerging, for example from DALICC.

that understands the standard. ODRL's expressiveness allows it to cover scenarios ranging from simple license tags to complex conditional statements (e.g. "dataset X may be used for research purposes if attribution is given and if the user is affiliated with an academic institution").

Another initiative is Creative Commons' ccREL (Creative Commons Rights Expression Language) [58], [61], which is a lightweight approach to embed license information in web resources. ccREL uses RDFa annotations in HTML, essentially attaching machine-readable references (such as a CC BY licence or CC0 rights waver URL) to web pages or metadata records[6]. This was among the first widely adopted methods for machine-readable licenses, and indeed Creative Commons licenses are designed to be understood on three levels: the legal text, a human-readable summary, and a machine-readable code or URL. Beyond these, metadata standards used in scholarly communication have gradually incorporated rights elements. The DataCite schema for dataset metadata [64], for instance, includes fields for license and rights, and repository platforms using schema.org can specify a license or copyright attribute for a dataset [82]. OpenAIRE's guidelines similarly mandate that repositories expose an access rights field (open vs. restricted, etc.) and a license URI for each record [79], [80]. All these efforts indicate a recognition that rights metadata must be standardised to be effective.

Treating rights metadata as infrastructure is more than just having a license field in a database – it requires a rich, interoperable, and actionable metadata layer. Current practices, while a starting point, often fall short of true machine-actionability and cross-platform consistency. Many repositories still use free-text descriptions for usage rights or simply provide a link to a human-readable license, which software agents cannot fully interpret. Even the common practice of indicating a Creative Commons license in metadata – which greatly helps discovery – does not alone enable complex automated reasoning (for example, determining if two datasets with different licenses can be combined legally requires understanding the license terms, not just matching URLs). Moreover, different communities have developed their own conventions. As one study notes, even the basic labels for access level vary: one repository's "Safeguarded" is another's "Controlled", and what one calls "embargoed" another might call "restricted" [7]. Without alignment, such categories are *"of little practical use beyond simple discovery"* because their precise meanings do not translate from one repository to another. In other words, if each service invents its own way to categorise and enforce rights, the burden falls on human intermediaries to broker trust – negating the efficiencies that EOSC aspires to.

To elevate rights metadata to true infrastructure status, two improvements are critical: standardisation and automation. First, adopting or converging on common vocabularies and schemas ensures that a "restricted access" tag or an "attribution required" condition is uniformly understood everywhere. Community profiles of ODRL or similar frameworks can define a core set of terms relevant to Open Science (for example, a standard way to express "non-commercial use only" or "data subject consent required"). If EOSC services agree on such standards, the rights metadata from one repository can flow into another without

[6] ccREL at W3C

[7] Advancing access interoperability with ODRL

losing meaning. Second, the metadata must be actionable by machines – meaning that software should be able to not just read, but also to enforce or react to the rights information.

Recent work in the FAIR and Open Science community has highlighted this need. For instance, in the context of automating FAIR data practices, commentators have noted that while we have made strides in machine-actionable *discovery* metadata, the *access and usage* conditions remain largely human-governed and need to be brought into the automation fold [91]. One use case demonstrated adding ODRL policies to dataset landing pages to serve as a "machine-actionable corollary" to traditional human-only licenses or data use agreements. By encoding the rules (who can do what with the data) in a structured form, one can envision EOSC-wide services that automatically reason over rights – for example, a portal that filters out datasets you are not authorised to see, or a workflow engine that combines only license-compatible resources and alerts when a combination would violate a term.

It is worth noting that this push for structured rights metadata aligns closely with broader principles of Open Science. The FAIR principles (Findable, Accessible, Interoperable, Reusable) explicitly include requirements for clear licensing: to be Reusable, data and metadata should be published with a transparent usage license that is understandable to both humans and machines. In parallel, the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics) for Indigenous and sensitive data emphasise the importance of context and control in data reuse [75]. By integrating rights and ethical considerations into the metadata fabric, we support *Responsibility* and *Ethics* – for example, by indicating that certain data must always remain accessible to certain communities, can *only* be accessed by members of a particular community, or that any reuse requires notifying the data source. In short, rights metadata as infrastructure means building a layer into the EOSC and similar ecosystems that consistently carries access conditions, licensing, attribution requirements, and ethical considerations alongside the data. This layer should leverage existing standards (W3C, Creative Commons, etc.) but also extend them where needed for the academic context, ultimately enabling *legal interoperability*. When done right, it will enhance transparency, trust, and efficiency: researchers will more easily find content they can use, and providers will have confidence that usage rules travel with the content. Recent recommendations indeed call for using standards such as DCAT (for dataset descriptions) combined with controlled vocabularies for licenses and policies (e.g. Creative Commons for licenses, ODRL for detailed conditions, Data Privacy Vocabulary [3] for GDPR-related constraints) to achieve such interoperability [92]. The task now is to implement these ideas in a coherent way for the EOSC.

1.3. Scope and Contribution of this Work

The work presented in this report responds directly to the challenges outlined above. Within EOSC Beyond, the broader Task 13.3 covers the design and validation of a Rights Markup Language (RML). The specific scope of this deliverable (D13.3) is more focused: to lay the state-of-the-art foundation and present the rationale for an EOSC-tailored RML. While the title of this deliverable refers to a "state of the art," the scope is intentionally broader: it surveys

existing approaches but also derives design requirements and provides a rationale for the EOSC Rights Markup Language (RML). In essence, we seek to provide the missing piece of EOSC's metadata infrastructure: a common vocabulary and schema to express all relevant rights information about research outputs in a machine-readable, interoperable form.This deliverable therefore represents the first step: a state-of-the-art review, the identification of key requirements, and initial recommendations. The subsequent steps of design, validation, and pilot implementation will be addressed in later stages of the Task. In essence, we seek to provide the missing piece of EOSC's metadata infrastructure: a common vocabulary and schema to express all relevant rights information about research outputs in a machine-readable, interoperable form. This entails building upon the best of existing standards while addressing their gaps in the EOSC context. For instance, our RML design takes inspiration from W3C ODRL's rich policy expression capabilities[8] and the simplicity of ccREL for embedding license info[9], and it aligns with widely used schemas (DataCite[10], schema.org[11]) to ensure integration with current repository metadata. By reusing and extending standard REL concepts, the EOSC RML can represent familiar constructs like Creative Commons licenses or standard access levels, so that it remains compatible with what data providers already publish. At the same time, the model is being developed to cover *additional requirements* critical in Open Science: not only copyright licenses, but also conditions stemming from privacy law, ethical consent, or funder-imposed mandates. A key contribution of this work is a systematic mapping of those requirements – from GDPR restrictions on personal data to the CARE Principles for ethical data sharing – and an incorporation of them into the vocabulary design (for example, the ability to indicate "data use for health research only" or "consent documentation required for access"). By doing so, we aim for a **comprehensive rights description framework** that goes beyond what general-purpose RELs currently offer, while remaining implementable in technology and policy.

Concretely, this work's objectives, carried out within EOSC Beyond Task 13.3, include: (a) analysing existing practices in open data and open-source platforms to understand how rights are currently managed; (b) formally defining the RML – an RDF-based schema and set of controlled terms – that can express permissions, obligations, prohibitions, and conditions relevant to EOSC; and (c) ensuring that this RML supports interoperability and automation across services. Interoperability here means two things: legally, that the RML can accurately capture standard licenses and usage policies (so that, for example, a Creative Commons license or a well-known data usage agreement can be represented without loss of meaning), and technically, that different EOSC services can exchange and understand RML-encoded rights without custom integration. To that end, we emphasise alignment with existing licensing frameworks: the RML will include terms or mappings for all Creative Commons license elements, common open-source software licenses, and emerging frameworks like RightsStatements.org used by digital libraries. This alignment is crucial so that adopting the

[8] https://en.wikipedia.org/wiki/ODRL#

[9] https://www.w3.org/submissions/2008/02/#

[10] https://schema.datacite.org/meta/kernel-4.6/

[11] https://schema.org

EOSC RML does not mean reinventing licensing, but rather building a translation layer that makes diverse licenses and policies machine-comparable. In terms of automation, our work facilitates what might be called policy-aware services. For example, if one EOSC portal harvests records from another, the presence of RML rights metadata should allow automated checks like “only show this dataset to authenticated users from certain sectors” or “warn if attempting to combine data with license A with data under license B, as there may be a conflict.” This kind of capability is largely missing today due to the lack of structured rights metadata.

In the report, we survey the landscape of rights expression solutions (what exists, what works or not, in Section 2) and derive from that survey a set of design requirements for a rights vocabulary that fits EOSC’s needs (discussed in Section 3). The report consolidates knowledge from diverse sources – from W3C standards[12] and Creative Commons documentation, to policy documents like the EOSC Strategic Implementation Plan and real-world platform practices (e.g. how Zenodo, Dataverse, and OpenAIRE Graph handle licensing, how the EGA handles controlled access, how Europeana implements rights statements). By grounding our work in existing practice, we ensure that the proposed RML is not built in a vacuum but rather *fills the gaps* identified in current approaches. For instance, we highlight that license interoperability is a practical concern: integrating data from multiple repositories often raises the question of whether the licenses are compatible, a check that ideally should be automatable. Our RML framework is being designed with this in mind, so that one can programmatically determine license compatibility or propagation requirements (e.g. share-alike clauses) when combining resources. Another important aspect is the inclusion of usage analytics and compliance use cases – by making rights metadata uniform, it becomes easier to monitor how data is used and ensure it aligns with the given permissions (for example, by logging access under certain conditions or feeding into notification systems if misuse is detected).

Finally, the work developed in this Activity of EOSC Beyond is not purely theoretical: a pilot implementation and validation phase accompanies the RML design. As part of the broader EOSC initiative, we collaborate with a few key platforms (for example, the Zenodo repository operated by CERN, and the OpenAIRE Graph) to test embedding and exchanging the new rights metadata. These pilots allow us to verify in real scenarios that curators can apply the RML to their datasets, that the information can be harvested and parsed by EOSC services, and that it indeed enables smarter behaviors (like filtering or automated access control decisions). Feedback from such implementations will inform refinements to the vocabulary. By the end of the project, the aim is to deliver not only the RML specification but also guidelines and examples showing how repositories and data providers can adopt it in practice (for instance, how to migrate from a simple license field to a full policy description). In summary, the contribution of this work is a coherent approach to digital rights in Open Science: from understanding the landscape, to designing a solution, to demonstrating its value. If successful, it will help bridge the legal, ethical, and technical aspects of Open Science, providing a layer of trust and clarity that supports researchers in sharing and reusing data under well-defined conditions.

[12] https://en.wikipedia.org/wiki/ODRL#

1.4. Structure of the Report

The remainder of this report is structured as follows:

- Chapter 2 – Rights Markup Languages: This chapter reviews the state of the art in rights expression languages and rights metadata practices. We examine existing standards (such as W3C ODRL, MPEG-21 REL, Creative Commons' ccREL, and RightsStatements.org) and assess their features and limitations in the context of Open Science. We also survey how rights metadata is currently handled in various platforms and domains – including academic data repositories (e.g. Zenodo and Dataverse), aggregators (e.g. OpenAIRE Graph), cultural heritage infrastructures (e.g. Europeana), and sensitive data repositories (e.g. ELIXIR/EGA) – highlighting common patterns, gaps, and challenges. This analysis sets the stage by identifying what problems remain unsolved by today's solutions (for example, lack of machine-actionability, fragmentation of standards, limited support for complex constraints).
- Chapter 3 – Rights in EOSC: Scope and Classification: Based on the gaps identified in Chapter 2, this chapter discusses the requirements and design considerations for a domain-specific RML for EOSC. We outline the core elements of the proposed vocabulary and model, explaining how it builds on existing frameworks while introducing necessary extensions for the EOSC context (such as handling GDPR-related restrictions or community-specific usage conditions). We also address issues of interoperability and governance – for instance, how the EOSC RML can align with the FAIR principles and be adopted across heterogeneous services. The chapter may include preliminary examples or use cases illustrating how an EOSC policy would be encoded (e.g. a sample ODRL profile for EOSC). By the end of this section, the reader should understand the scope of the solution and how it meets the identified needs.
- Chapter 4 - Current Practices in EOSC-aligned Platforms. This chapter provides an overview of current practices in selected EOSC-aligned platforms. These include Zenodo, DANS / Dataverse NL, the European Genome Phenome Archive (EGA) and OpenAIRE Graph. These were selected based on their relevance, diversity of use cases, and the maturity of their metadata practices,
- Chapter 5 - Gaps and Limitations in Current RMLs and EOSC Practice. This chapter outlines key shortcomings that must be addressed to achieve a robust and comprehensive rights management framework in EOSC.
- Chapter 6 – Synthesis: Findings and Recommendations: The final chapter summarises the key findings and contributions of the work. We reflect on the implications of having a unified rights vocabulary in EOSC – in terms of benefits for researchers (easier reuse, clarity of conditions) and for service providers (reduced legal uncertainty, enhanced automation). This chapter also outlines future steps, including plans for implementation and integration into the EOSC ecosystem, and recommendations for stakeholders (such as metadata providers, repository managers, and policy makers) to support the uptake of the EOSC RML. Any limitations or open issues are discussed here, along with how they might be addressed through further research or community efforts.

Throughout the report, we provide documented references to relevant standards, platforms, and literature – from W3C specifications and Creative Commons documentation, to EOSC policy papers and principles like FAIR and CARE – to ensure that our arguments and proposals are well-grounded in current knowledge and practice. By structuring the report in this way, we move from the general context and challenges (Chapter 2) to the specific solution and its justification (Chapter 3), and finally to a broader perspective on implementation (Chapter 4), thereby providing a comprehensive narrative of why a digital Rights Markup Language is both necessary and advantageous for Open Science, and how this work contributes to that goal.

2. Rights Markup Languages: State of the Art

2.1. Conceptual Background

The development and implementation of Rights Markup Languages (RMLs) is grounded in the need for structured, standardised, and machine-actionable ways of expressing rights, obligations, and permissions attached to digital resources. These languages have emerged as key instruments for automating rights enforcement, facilitating compliance with legal and ethical requirements, and promoting interoperability in data ecosystems.

Rights Expression Languages (RELs) are formal models used to declare the conditions under which digital content may be accessed, reused, or redistributed. An REL typically defines a set of actors (e.g. licensors, licencees), actions (e.g. use, modify, distribute), constraints (e.g. time, purpose), and duties (e.g. attribution, share-alike). Most RELs are based on formal ontologies and use machine-readable formats such as RDF, XML, or JSON-LD. A Rights Markup Language (RML) refers specifically to the structured implementation of these expression models within metadata schemas, web standards, and repositories.

The role of RMLs in Open Science is to replace unclear or informal usage terms with well-defined legal language that machines can understand. Take for example a dataset shared under a CC BY 4.0 licence — an RML can clearly state that attribution is required for reuse and also address, separately from the CC licence (which only covers copyright issues), any ethical or regional rules. When rights are described explicitly, it becomes easier to build trust, ensure transparency, and meet legal requirements across distributed systems.

In addition, RMLs make it possible for different platforms to interpret rights information in the same way, which is crucial for semantic interoperability. This becomes especially important in large, interconnected environments such as the European Open Science Cloud (EOSC), where research materials flow between platforms, domains, and legal systems. Without a shared way to describe rights, automated processes such as data harvesting, reuse tracking, or access control can easily fail or create legal confusion.

A further contribution of RMLs is the capacity to model granular rights – i.e. rights associated not just with an entire dataset, but with specific components such as a single table, a software module, or an individual image. This granularity is essential in modern research, where digital objects are composite, dynamic, and recontextualised across use cases.

Finally, RMLs enable the articulation of ethical and normative considerations that extend beyond legal licence terms. This includes sensitivity to Indigenous data governance (e.g., CARE Principles [75], and data sovereignty [33], as well as community-based access restrictions. Many of these considerations are not captured by traditional licences, necessitating expressive and extensible rights models that can accommodate non-traditional constraints. For example, RMLs could support terms like "research use only," "community consent required," or "use restricted to non-profit purposes in country X," which

are vital for protecting sensitive data and aligning with principles such as FAIR [10], TRUST [11], and CARE [9].

In summary, Rights Markup Languages are not merely technical artifacts – they are essential infrastructure for expressing, interpreting, and enforcing rights in digital ecosystems. RMLs form a cornerstone of trustworthy and scalable Open Science architectures such as EOSC, by providing the means to formally encode usage conditions and thereby enabling automated, reliable compliance with both legal and ethical requirements.

2.2. Major Standards and Initiatives

In the last twenty years, a range of standards and initiatives has emerged to support the structured, machine-readable expression of rights information. These approaches differ in their purpose, level of detail, syntax, and areas of application. Some of the most impactful ones have influenced how metadata and policy frameworks are designed in fields like research, culture, and digital publishing. These major standards and initiatives are outlined below.

2.2.1. Open Digital Rights Language (ODRL)

ODRL is a World Wide Web Consortium (W3C) Recommendation (since 2018) [4] and is arguably the most widely adopted REL. It provides a flexible and extensible model for expressing permissions, prohibitions, and obligations structured around policies, assets, and parties. ODRL's information model defines core concepts such as Policy, Asset, Party, Action, and Constraint, allowing fine-grained specification of usage rules. Its RDF-based implementation supports integration with Semantic Web technologies, and ODRL is designed to be adaptable via profiles to specific community or domain needs. ODRL is used in a variety of domains, including digital publishing, research data management, and media services. In the context of EOSC, ODRL offers a solid foundation for defining machine-readable usage conditions and policy rules (e.g. encoding licence terms that can be automatically enforced).

2.2.2. Creative Commons Rights Expression Language (ccREL)

ccREL is a lightweight Rights Expression Language developed by Creative Commons in 2008 [61]. It uses RDFa (an HTML-embedded RDF syntax) to describe licence information in web pages and other digital works. ccREL is tightly coupled with the standard set of Creative Commons licences and focuses on simplicity, discoverability, and combined human/machine readability. For example, a content creator can embed RDFa in a webpage to indicate that the content is licensed under a particular Creative Commons licence, and this information is both human-visible (via icons or text) and machine-readable for tools harvesting licence metadata. While ccREL is not a general-purpose REL (it essentially covers the Creative Commons licence framework), its simplicity and widespread usage across the web make it a key component of open data environments. It has been fundamental in

enabling billions of digital works to carry machine-readable CC licence information, facilitating automated discovery of reuse permissions.

2.2.3. MPEG-21 Rights Expression Language (REL)

MPEG-21 [77] REL is an ISO/IEC standard [78] (part 5 of the MPEG-21 multimedia framework[13], developed for expressing rights in the context of digital media and DRM (Digital Rights Management). It provides an XML-based model for defining complex rights expressions and digital licences, including features such as temporal and conditional constraints, sublicensing, and monetary terms. MPEG-21 REL was based on the earlier XrML language and was designed to be unambiguously machine-actionable, integrating with hardware/ software enforcement mechanisms for content use[14]. However, its adoption has remained relatively limited when compared to ODRL. This is partly because of its complexity and the fact that it was primarily designed for commercial content distribution and digital rights management, rather than for open access and research environments. In the context of research data and Open Science, MPEG-21 REL's emphasis on detailed DRM controls has made it less suitable, with many opting instead for simpler tools to express licences and rights.

2.2.4. RightsStatements.org

RightsStatements.org provides a standardised and straightforward set of rights statements [18] tailored to the needs of cultural heritage organisations and research institutions. It offers both human- and machine-readable labels that help clarify how digital objects can be used, especially in cases where complete copyright details are missing or unclear.

RightsStatements.org is an initiative launched by a consortium including Europeana and the Digital Public Library of America (DPLA) to provide a set of standardised rights statements for cultural heritage institutions. Instead of a full ontology or rule language, it offers a controlled vocabulary of 12 predefined statements (of which Europeana uses 14 variants including Creative Commons tools) that describe the copyright and usage status of digital objects (e.g., "In Copyright," "No Copyright – Other Known Legal Restrictions")[15].

These statements are intentionally simple and high-level, prioritising clarity and consistency over expressiveness. RightsStatements.org statements typically indicate whether an item is in copyright, under copyright with specific conditions, or in the public domain, etc., without detailing fine-grained conditions. The simplicity of this approach lowers the implementation barrier for institutions with limited technical capacity and ensures that end-users get clear, multilingual guidance on reuse possibilities. While less granular than a full REL, the use of standard URIs for each statement makes them machine-readable and interoperable across platforms. This approach has been widely adopted in digital cultural heritage aggregators (e.g. Europeana) to improve legal clarity and multilingual consistency.

[13] https://www.iso.org/standard/44341.html

[14] https://www.loc.gov/standards/relreport.pdf#

[15] https://pro.europeana.eu/page/available-rights-statements#

2.2.5. DataCite Rights Metadata

DataCite, which provides a widely used metadata schema for research outputs (especially datasets), includes a dedicated Rights element in its schema [65] to capture usage rights. While not a standalone REL, the DataCite metadata schema allows depositors to specify a licence or rights statement for a dataset. Typically, this involves providing a human-readable rights description and, optionally, a URI pointing to a licence or terms of use. For example, one can supply the full title of a licence (e.g. *Creative Commons Attribution 4.0 International*) along with a URL to the licence text[16].

The DataCite schema supports multiple occurrences of the rights element, enabling description of complex rights if needed (such as multiple licences for different parts of a dataset). This integration of rights information into a general metadata schema forms a critical part of repository workflows: many research repositories (such as Zenodo, see below) rely on DataCite's rights field to store and propagate licence information. Although DataCite's approach is relatively simple (essentially referencing standard licences via text or URI), it is crucial for ensuring that basic machine-readable licence information travels with the dataset DOI metadata.

2.2.6. schema.org (CreativeWork.licence)

The schema.org vocabulary [82], broadly used in web metadata and Search Engine Optimisation (SEO), includes a licence property as part of the CreativeWork class. This property allows one to link a creative work (such as a dataset, publication, or media item) to a licence document, typically via a URL. In practice, embedding schema.org JSON-LD in a webpage or metadata record with a licence field can indicate, for example, that a dataset is available under a CC BY 4.0 licence by providing the licence's URL.

The schema.org approach is intentionally very lightweight – it does not provide an extensive rights model, only a hook to indicate the applicable licence. Nonetheless, this enhances discoverability on the web through semantic consistency [59] and enables basic rights expression in structured data formats like JSON-LD and microdata. Search engines and indexing services can parse this information to present licence info to users or filter results. The simplicity, however, means that schema.org alone cannot convey complex conditions or obligations; it usually complements more detailed RML frameworks by ensuring that at least a standard licence reference is present in web-exposed metadata.

2.2.7. Data Privacy Vocabulary and Privacy Statements

Data Privacy Vocabulary (DPV) [3] enables encoding of some policy provisions and privacy statements in machine-actionable formats. These statements overlap somewhat with the provisions that may be considered for a licence, and by and large the privacy statements published by repositories and infrastructures are not machine-readable [63]. Guidance on the measures required to ensure legal alignment is widely available [1], [43], [48].

[16] https://datacite-metadata-schema.readthedocs.io/en/4.5/properties/rights/#

2.3. Comparative Analysis of RMLs

A comparative analysis of existing Rights Markup Languages reveals significant variation in design philosophy, technical implementation, semantic expressiveness, domain specificity, and community adoption. Understanding these differences, which are outlined below, is essential for determining which model – or combination thereof – is most appropriate for the EOSC environment.

2.3.1. Syntax and Formalism

ODRL: ODRL uses a semantic web approach, with an RDF-based information model and support for JSON-LD and XML serialisations. This Linked Data compatibility allows ODRL policies to be integrated with ontologies and processed by inference engines. The formalism is highly expressive and backed by a well-defined vocabulary and ontology (W3C ODRL Vocabulary 2.2[17]) for permissions, prohibitions, and duties. ODRL's formal structure is flexible: it can be extended via profiles to meet community-specific needs without breaking overall interoperability.

MPEG-21 REL: In contrast, MPEG-21 REL is defined in XML with a strict schema (XSD). It was built around a rule-based approach to licence encoding, which makes it reliable but also somewhat inflexible. Its syntax is detailed and often difficult to work with, as it was originally intended to support a wide variety of multimedia licensing cases. Although it offers formal precision, the rigid XML schema can be a barrier when integrating with lightweight or modern web environments that tend to prefer formats like JSON or RDF.

ccREL: ccREL's syntax is minimalistic and web-native. It leverages RDFa, embedding RDF triples directly into HTML or XHTML. This choice prioritises ease of publishing on the Web – any HTML page can carry ccREL markup about its licence. The trade-off is that ccREL's formalism is tied to HTML contexts and is not a full-fledged standalone syntax; it piggybacks on host formats. ccREL does not aim for the same level of formal expressiveness as ODRL or MPEG-21; instead, it focuses on being easy to implement for Creative Commons licences.

In summary, ODRL offers a rich, linked-data-friendly formalism, MPEG-21 REL offers a heavy but strict XML schema, and ccREL offers a lightweight web-embedded approach. For EOSC's needs, the RDF/JSON-LD style of ODRL aligns well with modern interoperability requirements, whereas MPEG-21's XML might be too cumbersome, and ccREL, while simple, is limited to basic licence tagging.

2.3.2. Semantic Coverage

ODRL: ODRL provides the most comprehensive semantic model among the compared languages. It can specify detailed permissions and prohibitions and attach various constraints (temporal, spatial, purpose-of-use, etc.) and duties (requirements like attribution, share-alike conditions, etc.). ODRL policies are capable of expressing complex rules, including combined permissions with multiple conditions and even responses when a

[17] https://www.w3.org/TR/odrl-vocab/

prohibition is breached. One of its key strengths is the profile mechanism, which allows different communities to extend the core vocabulary with custom actions or constraints tailored to their needs. For instance, a genomics data profile might introduce a specific term such as "medical research use only." This kind of extensibility ensures that ODRL can adapt and expand its semantic coverage as new use cases emerge.

MPEG-21 REL: MPEG-21 REL also supports a rich set of terms, largely inherited from its basis in XrML. It can handle elaborate licence structures, including hierarchical or composite licences, and differentiate between various rights holders and conditions. It was designed to enable cases such as granting specific usage rights that might be transferred or super-distributed with content. However, the complexity of MPEG-21's semantics (and its verbose expression) makes it less human-readable and less adaptable without significant tooling. While powerful, its richness is difficult to leverage in practice outside of specialised DRM systems.

ccREL: ccREL's semantic scope is intentionally narrow. It essentially covers indicating which Creative Commons licence applies, who should be attributed, and where to find the licence text. It does not natively support additional conditions or any custom rights beyond what the CC licences already define. There's no facility for adding constraints such as time or location, or for expressing obligations beyond what, say, CC BY or CC BY-NC already mean in legal terms. Thus, ccREL is limited to the semantics of the Creative Commons licensing framework (which are themselves simpler than the full range of all possible rights expressions).

RightsStatements.org: This approach takes a very different path by offering *predefined* semantic statements rather than a freeform language. The semantic coverage of RightsStatements.org is limited to a small set of broad categories (e.g. "In Copyright," "Copyright Not Evaluated," "No Known Copyright", etc.). It does not support detailed conditions or exceptions, as each statement is predefined and carries a fixed, implicit meaning. The benefit of this approach is complete clarity, every statement has a documented and translated definition, which helps minimise ambiguity and ensures consistent interpretation across contexts. The disadvantage is that if a situation does not fit one of the standard statements, institutions have to choose the closest match (potentially losing nuance) or not use the framework. In other words, RightsStatements.org sacrifices granular expressiveness for consistency and simplicity.

In summary, ODRL has high semantic expressiveness (and extensibility), MPEG-21 REL also high but hindered by complexity, ccREL is low (focused on CC licences), and RightsStatements.org is very low (fixed statements only). The choice between them involves balancing the need for expressiveness with the need for simplicity and clarity.

2.3.3. Interoperability and Reusability

ODRL: As a W3C standard aligning with Linked Data principles, ODRL **scores strongly** on interoperability. ODRL policies are represented in RDF, which means they can be linked and merged with other data (e.g., connecting a data object's metadata to an ODRL policy via a URI). The vocabulary is well-documented and versioned by W3C, and there are open community discussions ensuring that updates remain backwards-compatible. ODRL has

been incorporated into various frameworks and projects—for example, it was used in EU-funded projects, such as MediaScape, to manage content usage policies. This shows that ODRL is being adopted more widely across different fields. This suggests that ODRL is gaining traction across various domains. However, its flexibility can also lead to interoperability problems. When different communities create their own profiles or interpret the model in slightly different ways, their policies may end up being incompatible. To address this, it's important to rely on well-established profiles and use validation tools that help keep implementations consistent. Establishment of a formal governance mechanism for EOSC profiles and vocabulary extensions will mediate this possible divergence.

MPEG-21 REL: Within its niche (digital media distribution), MPEG-21 REL offers interoperability in the sense that it's an ISO standard and can be implemented consistently given the strict schema. For example, different DRM systems that both implement MPEG-21 REL should, in theory, understand each other's licences. But in open science, few platforms natively support MPEG-21 REL, so cross-platform interoperability is minimal. The standard is not widely used in research data or open content repositories, which means an MPEG-21 licence expression might not be recognised or actionable if you move a dataset from one system to another. Thus, its interoperability is largely confined to certain industries (e.g., commercial multimedia frameworks) and is **weak** in the context of EOSC-like environments.

ccREL and schema.org: These are broadly supported in the web ecosystem at a basic level. Many websites and content management systems include CC licence metadata using ccREL or schema.org, and web indexers or aggregators (e.g. search engines, or services such as Google Dataset Search) can read these to some extent. This makes rights information *discoverable*. However, the depth of machine processing is limited – a system can know *which* CC licence via ccREL, but it can not for instance, automatically negotiate conditions beyond that. For EOSC, ccREL and schema.org provide a **moderate** level of interoperability: they ensure that at least licence links are uniformly understood across repositories and portals, but they do not support advanced policy reasoning. On the plus side, nearly every repository can handle a licence URL (schema.org style) or ccREL tag, so at least the licence info itself is portable and reusable. It provides a mechanism for linking to a richer expression of rights, for example by referencing an ODRL profile (effectively a context-specific licence).

In general, ODRL offers **strong** interoperability (especially if profiles are coordinated), ccREL/schema.org **moderate** (universal but limited), and MPEG-21 REL **weak** outside its domain. RightsStatements.org can be considered **moderate** in interoperability: it's used by many cultural heritage platforms in a consistent way, but outside that circle its statements might not be recognised.

2.3.4. Implementation Contexts

Different RMLs have seen adoption in different contexts, highlighting their respective strengths:

ODRL: Implementations of ODRL can be found in diverse settings. Some open data repositories and platforms (such as certain pilots by OpenAIRE, or the Zenodo-related projects) have explored using ODRL for more expressive licensing scenarios. ODRL is also used in media archives and publishing, e.g., the MediaScape project used ODRL to define

policies for multi-screen multimedia content. In software, ODRL has been mapped onto some licence frameworks to encode reuse policies. The adoption in core EOSC services is *emerging* rather than fully established, but interest is growing due to ODRL's alignment with data interoperability goals.

RightsStatements.org: This has become the de facto standard in many digital heritage institutions. Europeana and DPLA's leadership in this initiative means that countless museums, libraries, and archives use RightsStatements.org when sharing metadata through those aggregators. Its use is heavily concentrated in the cultural heritage sector (as opposed to scientific data). For EOSC, which deals more with scientific outputs, RightsStatements.org is less directly applicable except possibly for research collections that intersect with cultural heritage data, for example in the arts and humanities domains.

MPEG-21 REL: Implementations are mostly in DRM-heavy domains. For instance, certain video content distribution networks or e-book systems have used MPEG-21 REL to issue usage licences that devices enforce. However, in research data repositories, one would be hard-pressed to find an implementation of MPEG-21 REL. It remains niche and largely irrelevant for open science infrastructures, which lean towards open licensing rather than DRM.

DataCite schema and schema.org: These are *pervasively* implemented in general-purpose repositories. Zenodo (see below) uses DataCite's rights element to record licences. Many institutional repositories and data archives output schema.org JSON-LD (with a licence property) in their landing pages for Search Engine Optimisation (SEO) and indexing purposes. Because these are simpler mechanisms, they are present virtually everywhere: any repository that mints a DOI likely uses DataCite metadata, and any repository aiming to be indexed by Google uses schema.org. So, while one might not call them "RELs" in the strict sense, they are widely adopted in practice as a means of conveying rights information.

To summarise this comparative overview, Table 2.1 below highlights key characteristics of the major RML approaches.

This analysis highlights that no single standard satisfies all EOSC requirements out-of-the-box. However, ODRL emerges as the most promising base due to its flexibility, formal grounding, and compatibility with FAIR and Linked Data principles. At the same time, complementary use of lighter-weight solutions like schema.org (for broad discovery) and RightsStatements.org (for simpler cases or cultural heritage data) may offer practical solutions for different layers of the EOSC infrastructure. In effect, an optimal approach could involve a hybrid: ODRL for rich policy encoding where needed, and simpler metadata links or statements for routine or low-risk cases. The DALICC Project has also experimented with definition of a **composite vocabulary** based on ODRL, with custom, ccREL, Dublin Core, **schema.org**, **FOAF**, and other extensions as required.

Feature	ODRL	ccREL	MPEG-21 REL	RightsStatements.org	DataCite / schema.org
Syntax	RDF/JSON-LD	RDFa (HTML)	XML (XrML-based)	Controlled text (fixed)	Simple metadata field
Expressiveness	High (very granular policies)	Low (only CC licences)	High (detailed DRM policies)	Low (few broad categories)	Very Low (just link/licence)
Interoperability	Strong (Linked Data, W3C standard)	Moderate (web-wide support)	Weak (limited to DRM systems)	Moderate (within heritage networks)	Weak to Moderate (common but basic)
Customisability	High (via profiles)	None (fixed to CC)	Moderate (extensible within standard)	None (fixed statements)	None (limited to provided values)
Adoption in Open Science	Growing (pilot implementations)	High (for basic licences on the web)	Low (rarely used in OS)	High in cultural heritage (less in science)	Very High (almost all repositories use these)
Licensing Granularity	Full (per asset or fragment)	Licence-level only (whole item)	Full (detailed per item)	Statement-level (whole item)	Licence reference only (whole item)

Table 1: Comparison of Rights Markup Languages and related approaches

2.4. Implementation Examples

Several real-world implementations of rights metadata provide valuable insights into how RMLs (or simpler alternatives) are applied across different domains. These platforms, outlined below, demonstrate a range of approaches, from lightweight licence tagging to semantically rich models. Their practices inform the design of a fit-for-purpose RML for the EOSC by highlighting what has worked and where gaps remain.

2.4.1. Europeana

Europeana, the European digital platform for cultural heritage, was one of the first to adopt standardised rights statements to improve legal clarity and interoperability in its field. Working together with the Digital Public Library of America (DPLA), it helped develop the RightsStatements.org framework — a controlled set of rights statements specifically designed for cultural heritage materials. Every digital object aggregated by Europeana includes a machine-readable rights statement URI as part of its metadata. For example, an image in Europeana might be labeled with the statement "In Copyright – Educational Use Permitted," identified by a URI from the RightsStatements.org vocabulary. By enforcing that every item *must* have a **rights statement**, Europeana ensures there are no ambiguous cases – something critical for user trust and for automated reuse pipelines.

2.4.2. Zenodo

Zenodo, operated by CERN, is a general-purpose research repository that handles publications, datasets, software, and more. Zenodo's approach to rights is pragmatic: when depositing an item, users must choose a licence from a predefined list, and this licence is then stored in the item's metadata (following the DataCite schema). The list of licences covers all standard Creative Commons licences for publications/data and all OSI-approved open source licences for software, among others. The system default for datasets is CC BY 4.0 (for open data)[18], but users can select others or even add a custom licence if one of the recommended licences is not suitable[19].

The metadata stores the licence name and URL, which makes it machine-readable to an extent (e.g., harvesters know the licence URI). Zenodo does not implement a full REL like ODRL to capture complex conditions; however, it does allow some additional rights-related metadata: for instance, one can set an embargo date (after which the item becomes publicly available) or indicate access restrictions (open, embargoed, closed, restricted). Those are not expressed in a formal policy language but are handled by the repository's logic (e.g. an embargo date field). In summary, Zenodo represents a typical EOSC service case: it ensures that basic licence information is structured and exposed for every record, which is crucial for integration and discovery, but it currently stops short of encoding detailed machine-actionable policies.

[18] https://help.zenodo.org/docs/deposit/describe-records/licenses/#

[19] https://help.zenodo.org/docs/deposit/describe-records/licenses/#

2.4.3. OpenAIRE Graph

OpenAIRE Graph is not a single repository but an infrastructure that aggregates metadata from thousands of repositories across Europe and beyond. Through its guidelines, OpenAIRE strongly encourages (and in some cases mandates) inclusion of rights information in the metadata that repositories share. In practice, OpenAIRE harvests records via protocols such as **OAI-PMH** and looks at metadata fields for licence info (for example, the Dublin Core rights field, or the DataCite rights element if using DataCite XML). It accepts a variety of inputs: some repositories may provide a licence URI, others a text string like "CC BY 4.0", others might have an embargo status, etc. OpenAIRE has to normalise this information to make it searchable and comparable in its portal.

The OpenAIRE Guidelines for Literature Repositories and Data Archives [80] define recommended vocabularies for rights, essentially advising data providers to use standardised terms (such as Creative Commons URIs or **SPDX identifiers** for software licences). In practice, however, the aggregated metadata shows inconsistencies: not all repositories follow the guidelines strictly, and some provide only free-text descriptions of rights. This variability means that, for now, OpenAIRE's collected rights metadata is of uneven quality – some of it is machine-readable, some not. This illustrates a gap: a formal RML framework adopted across the board (with validation) would greatly enhance the interoperability of rights metadata that OpenAIRE and similar aggregators deal with.

2.4.4. DANS (Data Archiving and Networked Services)

DANS, the Netherlands' national repository and data services organisation, provides a model of rich rights metadata integration within research data repositories. Through platforms such as **DataverseNL**, the discipline-specific Data Stations and the former **EASY system**, DANS supports detailed licence selection and even user-defined access conditions. When a researcher deposits a dataset in a DANS repository, they can choose from a range of licence types (Creative Commons licences for data, open source licences for code, or even custom institutional licences if necessary). This choice is recorded with standard identifiers (e.g., a URL to the licence text, or an SPDX code).

Beyond simple licences, DANS also captures information such as whether personal data is present (triggering GDPR considerations), whether an embargo is in place, or if additional conditions apply (like a data usage agreement). If one or more files in a dataset require restricted access, these can be individually indicated and terms of access can be specified; the bespoke '**DANS licence**' is typically used for Restricted Access datasets and sets out the conditions for using datasets that are published under restrictions. This licence was developed by DANS and is not used by other data providers. The licence responds directly to several sensitive data-specific licence provisions, and work has been done to elaborate and possibly standardise these considerations [35] as 'Managed Access' licences.

DANS rights metadata is exposed via standard protocols (OAI-PMH, metadata APIs) so that aggregators and search portals can ingest it. DANS has been an active participant in European projects focusing on FAIR data (e.g., the FAIRsFAIR and FAIR-IMPACT projects) and EOSC pilot programs, and through these it contributes to developing best practices for rights

documentation. The experience of DANS demonstrates that it is feasible to integrate nuanced rights information (including ethical and legal constraints) into repository workflows in a way that is both researcher-friendly and machine-actionable. However, it also underscores the importance of wider standardisation — initiatives like that of DANS can only reach their full potential when other repositories adopt comparable practices, enabling better alignment across the EOSC.

2.5. Gaps and Challenges

While there has been clear progress in developing and applying Rights Markup Languages, important gaps remain that hinder their full potential to support scalable, interoperable, and ethically responsible research data infrastructures like the EOSC. A number of key challenges emerge, as discussed below.

2.5.1. Fragmentation of Standards and Practices

One major challenge is the **fragmentation** of existing standards. Multiple RELs and rights metadata frameworks have emerged, each with different levels of granularity, syntax, and domain specificity. ODRL, being very flexible and general, can be applied in many contexts – but this broad applicability can lead to inconsistent implementations unless community profiles and guidelines are strictly adhered to. Conversely, more prescriptive frameworks such as RightsStatements.org or ccREL provide clarity and consistency within their limited scope, but lack expressiveness needed for complex or cross-domain use cases.

The result is a lack of semantic alignment across the ecosystem: the same dataset could be described with an ODRL policy in one repository, a simple licence URL (schema.org/DataCite) in another, and a RightsStatements.org tag in yet another. These are not automatically equivalent, complicating interoperability. In EOSC, where data from many sources comes together, this fragmentation means additional work is required to reconcile and interpret rights information. Without convergence or at least mappings between these approaches, automated systems struggle to uniformly handle rights, and users face confusion.

2.5.2. Limited Machine-Actionability

Including a licence in metadata is now common, but true **machine-actionability** of rights information remains limited. Many repositories and data portals still store licence or rights information as unstructured text or as a hyperlink without further logic. For example, a repository might have a free-text field "Licence: Creative Commons Attribution" or even just embed the licence in a PDF document. Such approaches hinder automated processing. Even when a structured field is present (e.g. a licence URI), the systems consuming that metadata often do not automatically enforce or react to it. For instance, if a dataset is marked as "restricted access" in one repository's metadata, another system harvesting that metadata might not automatically prevent access because it does not "understand" that flag without human interpretation.

In an EOSC setting, machine-actionability is essential: when data moves between services, there should be automatic checks like "Is the user allowed to do this with the data under these licence terms?" or "This dataset can be integrated into that platform because its licence is compatible." Currently, those kinds of automated decision-making processes are underdeveloped because the rights metadata either is not expressive enough or the tooling to interpret it is missing.

2.5.3. Lack of Support for Ethical and Contextual Constraints

Existing RMLs tend to focus primarily on legal conditions derived from copyright or licensing frameworks. However, in Open Science—especially when dealing with sensitive data (e.g., personal data, health information, trade secrets, Indigenous data)—ethical and contextual obligations play a critical role. The FAIR Principles emphasise that '**FAIR is not equivalent to open**' and that there are many reasons why data may only be available under certain conditions, and/or to certain users.

Datasets containing personal or biomedical data, for instance, might be designated for "research use only" or "non-commercial use only" independent of copyright. Most current models and standards do not adequately support representation of these kinds of constraints.

The CARE Principles for Indigenous data go beyond the current open data /open science movement by calling for respecting collective benefit and authority of indigenous communities in data reuse. Most Indigenous data are held by non-Indigenous governments, institutions and agencies, and the rights are usually vested in the authors of the resulting research. The CARE principles state that Indigenous Peoples must have 'Authority to Control' to these data: firstly, they must have access to data that support Indigenous governance and self-determination. They must also have a role in using mechanisms to protect access to data as indicated by these communities, to minimise harms, and maximise benefit. The CARE principles complement and build upon the FAIR principles: the overarching goal is that Indigenous data can be both 'FAIR and CARE' [9].

There is no widely adopted standard mechanism to indicate, for instance, "community consent required before reuse" or "data use limited to specific geographic regions due to ethical agreements." While one could theoretically extend ODRL to cover some of these (e.g., add a custom constraint for "intendedPurpose: research"), such extensions are not yet standardised or common.

The lack of support for these ethical constraints leads to insufficient protection of sensitive data and undermines both the FAIR and CARE principles when data is shared widely. It also forces providers to handle these via out-of-band means (like contract agreements or manual vetting) rather than the metadata/RML layer.

2.5.4. Interoperability at Scale

Achieving true legal and policy interoperability at the scale of EOSC requires more than just vocabularies; it requires shared interpretation and infrastructure. Currently, infrastructures are not equipped with validation and reasoning **tools** for rights metadata. For example, if one repository encodes a complex ODRL policy, another service would need an ODRL engine to interpret it – which it likely does not have. Additionally, differences in national or jurisdictional rules (e.g., what "NonCommercial" means, or differences in data protection laws) are not captured by RMLs themselves and require context.

When metadata is harvested across platforms (as EOSC does via catalogs and aggregators), inconsistencies often arise because not everyone adheres to the same standard or level of detail. One repository might share very minimal rights info while another shares detailed info; when aggregated, the combined dataset metadata has uneven rights annotations, and some of that nuance can be lost. This creates a **loss of fidelity** and introduces risk in downstream reuse — for instance, a user might wrongly assume data is free to use because one portal did not display the restriction that was in another portal's metadata. To address interoperability at scale, efforts are needed in aligning standards (as noted in 2.5.1) but also in building crosswalks and translation layers (so a rights statement from System A can be mapped to an equivalent in System B), and in deploying services that can validate and enforce rights across systems.

2.5.5. Lack of Governance and Community Alignment

Finally, there is a general absence of coordinated governance around RML usage in the research community. While technical standards exist, their uptake is uneven and often ad-hoc. There is no central body or widely accepted coalition in EOSC that dictates: "This is how rights should be represented and these profiles must be used." Without community-endorsed profiles, clear implementation guidelines, and supported validation tools, different communities might diverge in how they use an RML, as discussed. The risk of divergence and incompatibility is high in the absence of governance. For example, two disciplines might both decide to use ODRL, but define their own profile with slightly different terms for similar concepts – later discovering their policies are not directly interoperable. Additionally, the lack of governance shows up in limited user education and tooling support.

Many researchers or repository managers are not experts in rights expression; they need easy-to-use tools (e.g. form-driven licence generators or policy templates) and guidance on how to apply them. Right now, such tools are scarce or rudimentary. Smaller institutions or individual researchers often just pick a licence from a list and do not engage with more detailed RML considerations, which is understandable given the complexity. This means that the potential of RMLs is underutilised. A governance framework under EOSC could help by providing training, maintaining canonical profiles (e.g., an "EOSC Data Policy Profile" of ODRL), and fostering an active community of practice.

The above analysis of existing Rights Markup Languages highlights both the technical maturity and the inherent limitations of current approaches in addressing the multifaceted needs of Open Science environments. Several standards provide valuable mechanisms for

expressing legal permissions and obligations, but gaps remain in achieving full interoperability, supporting ethical considerations, and enabling machine-actionable governance at scale. In the context of the European Open Science Cloud, these challenges must be addressed through a domain-specific rights framework that accounts for the legal, ethical, and practical realities of EOSC stakeholders.

Chapter 3 will expand on these findings by examining the rights and restrictions that are particularly relevant within the European Open Science Cloud (EOSC). It will provide a classification scheme for these rights, taking into account the diversity of research outputs as well as the expectations of data providers, users, and regulatory bodies. This classification will help inform the requirements for a Rights Markup Language tailored specifically for the EOSC, bridging the gap between existing standards and the operational needs of the EOSC.

3. Rights in EOSC: Scope and Classification

3.1. Types of Rights

In the context of the European Open Science Cloud (EOSC), the rights associated with research outputs are diverse and multidimensional. They encompass not only legal rights rooted in intellectual property law, but also ethical obligations, personal data protection constraints, and community-based governance principles. A robust Rights Markup Language (RML) must account for this diversity by enabling the clear, machine-readable representation of all these types of rights.

3.1.1. Intellectual Property Rights (IPRs)

Intellectual Property Rights are a well-known category in digital research outputs, governing ownership, access, and use. They include:

- **Copyright:** Protects original works of authorship (e.g., texts, images, software, datasets) where sufficient originality exists.
- **Database rights:** As established in EU Directive 96/9/EC, apply to structured data collections in which substantial investment has been made in obtaining, verifying, or presenting the content.
- **Software licences:** May be proprietary or open-source (e.g., MIT, GPL, Apache) and define how code can be reused, modified, or redistributed.

Within EOSC, the clear articulation of IPRs is essential to promote reuse while safeguarding creators' rights and ensuring legal certainty. Moreover, European Commission-funded research has to adhere to prescriptions in respect of intellectual property and open access to results [20].

3.1.2. Ethical Rights and Responsibilities

Beyond legal frameworks, many research outputs—especially in the humanities, social sciences, and health sciences—are governed by ethical considerations [87]. In addition, there exists a significant body of ethics guidance and codes of conduct in research [2], [9], [12], [13], [14], [15], [16], [22], [26] and these may require encoding in licences for transparency in respect of compliance or alignment. Domain-specific or international treaty-related expectations may also play a role [21]. Considerations include:

- **Informed consent:** Particularly for studies involving human participants, which limits how data can be reused and under what conditions.
- **Community rights:** Rights articulated by specific communities, such as those in the CARE Principles for Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, Ethics) (see Section 2.5.3 above)
- **Non-commercial use restrictions:** Often applied voluntarily to data from vulnerable communities, sensitive contexts, and/or creative practitioners (e.g. artists, writers, activists) limiting reuse for commercial purposes.

- **Research-purpose-only clauses:** Clauses that restrict reuse to scientific or academic purposes, excluding commercial exploitation.
- **Societal Benefit**: international agreements, of which the Nagoya protocol [21] serves as an example, predefine best practices in respect of ethical data sharing for some domains.
- **Right to Validate Reuse**: This is a right that overlaps with protection of sensitive data and indigenous community rights, often implemented by way of determining whether subject rights are impacted by prior to and/ or after creation of a derivative work.

Ethical rights may not always be codified in national law, but they are enforceable through institutional policies, funding agreements, and community norms. Representing these rights in metadata is critical for promoting responsible data stewardship. Encoding of ethical considerations may partly be supported by work in respect of ontologies for ethics considerations [7], [8].

3.1.3. Personal and Sensitive Data Protection Rights

Personal data is subject to strict regulation under the General Data Protection Regulation (GDPR) [1], [71], Digital Services Regulations [40] and other national or regional privacy laws. Subjects of study have further needs in respect of protection, for example in respect of disclosure that could harm communities, commercial interests, or the environment [35]. Data privacy considerations also partly overlap with ethical concerns [34]. In the EOSC ecosystem, relevant rights and constraints include:

- **Data subject rights:** Such as the right to be informed, the right of access, and the right to withdraw consent.
- **Data transfer limitations:** Restrictions on transferring data, especially across borders, when personal data are involved.
- **Data controller obligations:** Requirements on those handling data, including establishing a lawful basis for processing, data minimisation, and storage limitation. The Digital Services Regulations further require information to be available in respect of contact points in case of legal actions or orders.
- **Bibliographic Data**: Established practice regards bibliographic data, and by extension personal data references in citation metadata, as being in the public domain [29], [31]. Best practice requires this exemption to be included into privacy statements.

Although GDPR does not prescribe specific metadata vocabularies, any RML should be able to encode key indicators such as “contains personal data,” “can only be processed under controlled circumstances”, “data cannot be accessed, but derived results can”, “subject to data transfer restrictions,” or “requires explicit consent for reuse.”

3.1.4. Institutional, Producer, and Contributor Rights

Beyond individual and legal considerations, EOSC must also address rights asserted by institutions, data producers, and contributors. These include:

- **Embargo rights:** Conditions where data is withheld from open access for a specified period of time.
- **Attribution requirements:** Requirements that go beyond simple citation, mandating acknowledgment of certain funders, institutions, or contributors when the data is reused.
- **Usage monitoring and notification:** Conditions where repositories or data stewards request (or require) that any reuse of the resource be reported back to them[20].
- **Curator Rights**: Important to encode in a consistent and transparent manner: the right to improve metadata and data to align with community best practices, subject to proper version management, and the right to act on behalf of the depositor should the depositor be unable to attend to access requests (for example, due to illness, being deceased, or retired).

Such rights are often encoded in poorly structured repository policies or data management plans (DMPs). Recognising these in metadata is essential to capture the full spectrum of conditions attached to research outputs.

3.2. Examples of Rights in Open Data Contexts

Understanding how rights manifest in real-world open data contexts is essential for designing a fit-for-purpose RML. The examples in Table 2 illustrate the diverse ways in which legal, ethical, and personal rights are currently expressed (or sometimes omitted) across research infrastructures and repositories. They also underscore the challenges and inconsistencies that emerge without a harmonised rights vocabulary.

Aspect	Zenodo	EGA	DANS	Europeana
Description	General purpose repository operated by CERN and OpenAIRE	The European Genome-phenome Archive (EGA). Human genomics data that often include sensitive personal information	General purpose repository (Dataverse.nl), domain-specific repositories (Data Stations)	Europeana (the digital platform for European cultural heritage)
Access Control Mechanism(s)	Embargo, Restricted Access	Data access is tightly controlled through Data Access Committees	Embargo, Restricted Access	Rights Statements.org recommended
Guidance to Depositors	Yes	Yes	Yes	Yes
Licence Choices	Depositor selects from predefined list, or point to a		Depositor selects from a predefined list, which	"In Copyright" "No Copyright – Contractual

[20] For such feedback, one could consider machine-readable implementation based on the Data Usage Vocabulary [6].

	custom licence		includes the DANS Licence	Restrictions" "Public Domain" "Unknown Copyright Status"
Rights Metadata	Licence URI, no REL implementation	No REL implementation	Licence URI optionally supplemented by free text access conditions No REL	Rights Statements.ORG, implemented with stable URIs, recommends Creative Commons licences
Special Cases	None	Reuse for research purposes only. Prohibition of commercial use. Requirement to maintain anonymity and confidentiality. Prior approval from an ethics review board (e.g., IRB or REC) is required	'Managed Access' [35]: Identify end user for future communication, and/ or Verify derived dataset impact of subject rights prior to or after processing, and/ or limit use to controlled environments, and/ or limits to source data visibility	None
End User Aspects	End user policies and terms of use [89], [90]	Data Use Agreement [69], [70]	Policies and Terms of Use [62] Assent to Licence conditions recorded	Implementation Guidelines
Needs	Structured models to express obligations and constraints beyond basic licence types	Extending RML capabilities to capture ethical and context-specific rights	Extending RML to capture 'Managed Access' provisions, requires richer vocabularies to semantically express nuanced rights information	Limited granularity of these standard statements—designed for simplicity—may be insufficient for complex research outputs

Table 2: Implementation Diversity in Research Infrastructures and Repositories

3.3. Software Licensing Complexity

In many open science projects, research software is hosted on platforms such as GitHub. GitHub allows users to attach standard open-source licences (e.g., MIT, Apache, GPL) to their

repositories, but it does not enforce the inclusion of machine-readable metadata or ensure licensing consistency across different components of a project.

It is common for a single research project to include multiple components with different licences—for example, a core codebase under GPL that links to third-party libraries under MIT or even proprietary terms. This situation can lead to potential licence incompatibilities that are not evident from metadata alone. Resolving such conflicts often requires manual review, since these details are rarely captured in structured metadata.

This example demonstrates the growing need for RMLs that support modular, component-based rights modelling, especially in software-intensive research environments.

The DALICC project partly solves the licence compatibility problem by providing a service whereby licence incompatibilities can be identified. The DALICC **vocabulary of licence provisions** is fairly comprehensive, and includes many commonly used open licences in its scope, but it may not contain all of the ones relevant in the context of EOSC.

Overall, the case studies (Table 3.1) reveal a fragmented and inconsistent approach to rights expression across open science platforms. Some infrastructures (like Europeana and Zenodo) provide structured options for declaring usage rights, whereas others rely on informal mechanisms or manual documentation that hinder automation and scalability. Moreover, ethical considerations and data protection requirements—though critical—are often underrepresented in existing metadata models.

These observations point to a clear need for a harmonised, extensible, and machine-readable approach to rights documentation within EOSC. As the next section will explore, the capacity to systematically capture, interpret, and act upon rights information is essential to ensuring legal certainty, ethical integrity, and technical interoperability in open science.

3.4. Rights Metadata Needs in EOSC

The European Open Science Cloud aims to become a federated environment for the publication, access, and reuse of research outputs. To achieve this vision while maintaining legal certainty and ethical compliance, the integration of high-quality, structured rights metadata is not optional—it is essential. Rights metadata underpins the ability of EOSC services to inform users, enforce conditions, and promote the responsible reuse of digital assets.

3.4.1. Legal Certainty, Transparency, and Reuse Assurance

A key objective of EOSC is to foster legal clarity and trust across its ecosystem. Researchers, service providers, and third parties must be able to quickly determine several factors about any given resource, including:

- Whether the resource is legally reusable at all.

- Under what conditions, and by whom, it can be reused[21].
- What obligations apply to its reuse (e.g., requirements for attribution, share-alike, or non-commercial use).
- Whether additional permissions are required for reuse (for example, from data subjects, depositors, or institutions).
- Whether the end user inherits any obligations of the rights holder in respect of for example subject rights.

Without structured rights metadata, answering these questions is either a manual, error-prone process or simply impossible. This lack of clarity introduces friction and legal risk into data workflows, discouraging reuse or potentially leading to unintended violations.

3.4.2. Support for Machine Actionability

EOSC operates at a scale that demands automation. Machine-readable rights information allows software agents and services to:

- Harvest rights metadata across distributed platforms and registries.
- Interpret metadata to filter search results based on permitted uses.
- Integrate rights checks into access control mechanisms.
- Trigger automated workflows (e.g., releasing data when an embargo expires, or prompting users to agree to certain terms before download).

To enable this level of automation, rights metadata must be expressed in standard, machine-interpretable formats (such as RDF or JSON-LD) and align with recognised vocabularies/ontologies (e.g., ODRL, DataCite, or schema.org).

3.4.3. Representation of Diverse Rights Types

EOSC must accommodate more than just traditional copyright. A comprehensive rights metadata framework needs to cover a range of rights and constraints, including:

- **Intellectual property rights:** for publications, software, and datasets (e.g., copyright licences, database rights, patents).
- **Data protection constraints:** linked to personal or sensitive data (e.g., GDPR-imposed restrictions, data subject rights).
- **Ethical conditions:** such as usage limitations, consent requirements, or community-specific governance rights.
- **Institutional or contractual policies:** including embargo periods, “request access” models, or deposit agreements mandated by repositories or funders.

This diversity calls for a flexible and extensible metadata model that can capture various types of rights and their conditions without oversimplification.

[21] With the critically important requirement that the conditions are such that it can be determined independently whether access will be granted if the conditions are met.

3.4.4. Interoperability Across Domains and Services

EOSC brings together services and repositories from different scientific disciplines, legal jurisdictions, and technical environments. This diversity requires rights metadata that is interoperable across domains. In practice, rights metadata should:

- Be shareable via APIs and harvesting protocols.
- Be understood across different platforms, regardless of internal implementations.
- Support multilingual representation of rights information.
- Map to or align with existing community standards where possible.

Such interoperability ensures, for example, that a dataset from a health science repository in Finland can be legally and ethically reused in a climate research platform in Spain—without the need for manual legal intervention. Some cases, despite machine actionability of access conditions, will continue to require human evaluation (for example whether a derived dataset impacts the rights of subjects).

3.4.5. Governance, Provenance, and Validation

Effective rights metadata in EOSC must also support governance and quality control. In particular, it should enable:

- Provenance tracking (recording who assigned the rights, when it was done, and on what basis).
- Versioning and updates (capturing changes as licences are updated or embargoes expire).
- Validation rules (detecting incomplete, conflicting, or invalid rights declarations).
- Governance mechanisms (allowing service providers or communities to define profiles, policies, or constraints specific to their context).

These features are crucial for building a rights metadata infrastructure that is not only interoperable but also trustworthy and sustainable over time. While not an explicit requirement in respect of rights metadata, EOSC will also likely require the development of guidelines and best practice in respect of a record of access requests, outcomes of such requests, and any correspondence related to it.

3.5. Challenges in Classification and Expression

Classifying and formally expressing rights in digital environments presents a number of conceptual, technical, legal, and institutional challenges. In the EOSC context, where diverse stakeholders contribute heterogeneous research outputs, these challenges are amplified by requirements for scalability, automation, and cross-border interoperability.

3.5.1. Conceptual Ambiguity of Rights Categories

There is a lack of universally agreed definitions for many categories of rights. Terms such as “ethical rights,” “usage rights,” or “community rights” are often used inconsistently across disciplines, funders, and institutions. For example:

- The meaning of “non-commercial use” can vary by jurisdiction and context.
- “Research use only” might include or exclude industry-funded research, depending on the interpretation.
- “Sensitive data” might refer to personal data in one context, culturally restricted data in another, or confidential commercial data in a third, or to environmentally sensitive data.

Such ambiguities complicate efforts to create standardised taxonomies or controlled vocabularies. They also hamper the ability of machines to interpret rights metadata without human input.

3.5.2. Multiple and Overlapping Rights Layers

A single digital research output can carry multiple layers of rights and restrictions, including:

- Copyright governing the text of publications or the content of datasets.
- Software licences covering any accompanying code or scripts.
- Institutional policies imposing access restrictions or embargoes.
- Data protection obligations for datasets containing personal or sensitive information.
- Third-party content licences or restrictions for embedded materials.

These layers can interact in complex or even conflicting ways. For example, a dataset released under a permissive open licence may still contain personal data subject to GDPR—making unrestricted reuse legally problematic despite the open licence. Classifying and encoding such multi-layered rights in a coherent, machine-readable way remains a significant challenge.

3.5.3. Incomplete or Non-standard Metadata Practices

Even when repositories provide fields for rights metadata, these are often used inconsistently or incorrectly. Common issues include:

- Free-text entries for licence names (e.g., writing “CreativeCommons” instead of using a standard licence URI). SPDX [94], for example, provides a mechanism for registration of stable URIs for licences, but not all creators of licences make use of such a service.
- Ambiguous labels such as “Open Access” or “All rights reserved” without clear legal definition.
- Omission of key information like embargo end dates, reuse restrictions, or required ethical approvals.

Without standardised values and structured syntax, such metadata fields cannot support automated discovery, filtering, or enforcement of rights.

3.5.4. Jurisdictional and Legal Diversity

EOSC spans institutions across all EU member states and associated countries, each operating under its own legal frameworks and norms. For example:

- The scope of database rights and moral rights varies from one European country to another.
- National regulations define different thresholds[22] for anonymisation or for when explicit consent is required.
- The enforceability of open licences (such as Creative Commons licences) can differ by jurisdiction.

Designing a classification model that accounts for this legal diversity, while remaining harmonised and machine-processable, is a considerable undertaking.

3.5.5. Evolving Nature of Rights and Obligations

Rights and restrictions are not static – they can change over time as conditions evolve. Licences may be updated, embargoes may expire, or new ethical guidelines may emerge. A rights classification and metadata model must therefore support:

- Versioning and time-stamping of rights metadata (to record changes in rights or conditions over time).
- Dynamic expressions (e.g., metadata that indicates “embargoed until 1 January 2026,” and then updates when the date passes).
- Notification or audit mechanisms to reflect changes in status or policy (for example, alerting when a data use consent is withdrawn).

Most current metadata schemas lack support for such dynamic lifecycle management. This makes it difficult to ensure that rights information remains accurate and legally valid as circumstances change.

3.6. Proposed Classification Framework

In light of the above, a flexible yet standardised classification framework is proposed to accommodate the diversity of legal, ethical, and contextual obligations associated with research outputs. This framework is designed to guide the development of RML vocabularies and metadata schemas across EOSC services, supporting the structured representation, reasoning, and enforcement of rights.

[22] Which, in turn, may be at odds with domain best practices.

3.6.1. Classification Dimensions

The proposed model categorises rights along several orthogonal dimensions, as summarised in Table 3:

Dimension	Category	Source or Provision Examples
Nature of the RIght	Legal	Copyright, database rights, patents, …
	Ethical	CARE principles, informed consent, community governance…
	Data Protection	GDPR-related obligations, data subject rights (note: operates as a cross-cutting regulatory layer triggered by legal/ethical conditions).
	Institutional/Contractual	Embargoes, access policies, …
Entity the right applies to	Entire resource	Entire dataset or research output
	Component or fragment	An image within a dataset, or a specific section of a publication, …
	Associated metadata	Rights over the descriptive metadata or identifiers of the resource, …
Type of legal expression	Permission	E.g. reuse allowed under stated conditions
	Prohibition	E.g. commercial use forbidden
	Obligation	E.g. attribution required for reuse
	Constraint	E.g. access only after a certain date
Temporal Scope	Static	E.g., “always available under CC BY”
	Dynamic	E.g., “embargoed until January 2026”
	Conditional	E.g., “available upon approval by a Data Access Committee”
Jurisdictional Applicability	Global	E.g., a CC0 public domain dedication
	Regional	E.g., compliance with EU GDPR requirements
	National/local	E.g., subject to country-specific legal or ethical rules

Table 3: Proposed Dimensions of Rights Classification

This multi-dimensional approach allows the framework to be both granular and adaptable, enabling rights to be represented in a way that reflects real-world complexity without undue simplification.

3.6.2. Controlled Vocabulary and Encoding

The framework should be supported by a controlled vocabulary of rights types, reuse conditions, and ethical categories. Each term in this vocabulary should:

- Be associated with a persistent URI to enable unambiguous semantic linking.
- Be defined in multiple languages to support international adoption.
- Map to existing standards where possible (e.g. correspond to terms in ODRL, RightsStatements.org, ccREL, or DataCite's metadata schema for rights). It may be required to define additional vocabulary elements to account for provisions that are unique to the research landscape, for example in respect of trusted research environments or typical institutional policies.

In implementation, the framework should follow Linked Data principles, using formats such as RDF, JSON-LD, or XML. This will facilitate integration with EOSC services and registries and ensure that rights metadata is machine-interpretable across different platforms.

There may be utility in linking and reusing vocabularies and ontologies for which there is an overlap in concepts. These include the Data Use Vocabulary [5], [6] - dealing with feedback on and annotation of resources as a means of expressing intended and actual end use of a dataset, the Data Privacy Vocabulary [4], which may be useful in encoding of GDPR-related provisions, and emerging work in respect of encoding of ethics assessments [7], [8].

3.6.3. Practical Integration

For the classification framework to be adopted in practice, it must be:

- **Compatible** with existing EOSC metadata schemas (such as DataCite or schema.org).
- **Configurable** by institutions and service providers (allowing for local extensions or custom profiles where needed).
- **Validatable**, with mechanisms to detect incomplete, conflicting, or inconsistent rights declarations.
- **Maintainable**, incorporating versioning, governance processes, and community feedback to evolve the vocabulary over time.

One can imagine that provision of tools and API services that simplify the process of selection of licences and their provisions, and are able to detect conflicts in licence combinations will improve adoption.

This proposed classification framework is intended to serve as the foundation for the EOSC RML vocabulary and its supporting infrastructure. By adopting such a framework, EOSC can ensure that rights information is not only described, but also made actionable and enforceable—aligning with the values of open, responsible, and FAIR research.

4. Current Practices in EOSC-aligned Platforms

To provide a representative and practical overview of rights management practices within EOSC, this chapter focuses on a summary of four key platforms Zenodo, DANS / Dataverse NL, the European Genome Phenome Archive (EGA) and OpenAIRE Graph. These were selected based on their relevance, diversity of use cases, and the maturity of their metadata practices, ensuring a balanced cross-section of both general-purpose and domain-specific repositories.

The European Open Science Cloud encompasses a variety of repositories and services, each with its own approach to managing and expressing rights. This chapter provides an overview of how several prominent EOSC-aligned platforms currently handle digital rights and licensing metadata. The focus is on the above key examples to illustrate existing practices, strengths, and limitations in rights management. The detailed discussion and review of these case studies can be found in Appendix A.

Implementation Aspect	Zenodo	DANS	EGA	OpenAire Graph
Licensing - Data	Any predefined SPDX Licence URL to Custom Licence	Configurable List (SPDX subset) Includes 'DANS Licence'	Not in Use	Any predefined SPDX Licence URL to Custom Licence
Licensing - Metadata	CC0 by Default	CC0 by Inference	Not in Use	CC0 by Inference
Rights Metadata	Licence Element Access Level Free Text Conditions	Licence Element Access Level Free Text Conditions Sensitive Data Indicator	Standardised Access Conditions	Licence Element Access Level
Access Levels	Open, Embargoed, Restricted, Closed	Open, Embargoed, Restricted	Closed	Open, Embargoed, Restricted, Closed
File-Level Access or Licences	No	File-Level Access Control	No	No
Access Requests	Handled by Depositors	Formal request mechanism	Data Access Committees	Not Applicable
REL Implemented	No	No	No	No
Machine-Actionable Terms	No	No	Data Use Ontology (DUO) [5]	No
Specific Data Use Agreements	No	No	Yes, per request	No
Verification of Licence and Access Conditions	No	Licence acceptance recorded Human verification of access	Data Access Committee on a case-by-case basis	No
Limitations	Basic Implementation Free-Text Extensions No complex usage terms	Basic Implementation Free-Text Extensions No complex usage terms	Limited vocabulary Not interoperable	Basic Implementation High-Level Focus No complex usage terms
Planned Improvements	EOSC REL implementation	EOSC REL Pilot	Include DUO into EOSC RML	EOSC REL implementation

Table 4: Summary of Current Practices in Selected Repositories

5. Gaps and Limitations in Current RMLs and EOSC Practice

Significant gaps remain in the way rights are managed and expressed across the EOSC, despite the progress and practices observed in the repository cases we investigated. These gaps exist on multiple levels – legal, ethical, and technical – and they highlight limitations both in the available Rights Markup Languages (RMLs) themselves and in how repositories implement rights metadata. This chapter outlines key shortcomings that must be addressed to achieve a robust and comprehensive rights management framework in EOSC.

5.1. Legal Gaps

One set of gaps concerns the legal dimension: the discrepancies between what the law requires or permits and what current metadata and RMLs can express or enforce. A fundamental issue is licence compatibility and complexity. While many repositories record the licence of a digital object, there is often no mechanism to automatically reason about the compatibility of multiple licences when data are combined. For instance, a researcher may wish to merge two datasets with different licences, but determining whether the combination is legally permissible can be non-trivial. In practice, this is usually done manually. If done incorrectly, it can lead to a violation of terms or force the researcher to avoid using one of the datasets. A classic example is combining a dataset under a copyleft licence like CC BY-SA with another under a Non-Commercial licence (CC BY-NC); the share-alike terms conflict with the non-commercial restriction, preventing any unified redistribution. Current RMLs and repository workflows do not adequately alert or prevent such conflicts automatically.

Another legal gap is the inconsistent handling of jurisdiction-specific rights and restrictions. Intellectual property and data protection laws vary across jurisdictions, even within Europe. For example, Europe's *sui generis database right* applies in EU countries but not elsewhere; moral rights (like the right of attribution or the right to object to derogatory treatment of a work) have different scope in France than in the UK, etc. Moreover, privacy regulations like GDPR impose conditions (lawful basis for processing, cross-border transfer limitations, etc.) that may not be uniformly captured in metadata. Current repository metadata often simplifies these complexities – e.g., a dataset might just be labeled "contains personal data" with no further detail.

RMLs such as ODRL can, in theory, include constraints like geolocation or jurisdiction, but in practice these are seldom used. As a result, there is a gap in legal interoperability: a machine or user cannot easily tell if a dataset available in EOSC is cleared for use in another country or under what privacy constraints it operates. The lack of fine-grained metadata to denote legal nuances (like "this dataset cannot be used outside EU" or "only non-commercial research in Germany is allowed") is a limitation of current practice. Additionally, repositories have no common approach to expressing data protection conditions–for instance, whether

consent from data subjects is required for reuse, or whether data is fully anonymised (hence free to use) versus pseudonymised (thus regulated). These details are typically buried in documentation or not stated at all.

There are also gaps in coverage of certain legal rights. Traditional RMLs and licences focus heavily on copyright and usage permissions, but not on other legal constructs that matter in research data. *Moral rights* (the creator's right to be credited and to object to misuse) are generally not part of licence terms like CC-BY, yet are inalienable in many jurisdictions – and there is little support in metadata to flag works where, say, attribution is not just a licence obligation but a moral right. Another example is *patent or trademark considerations* in software or datasets–current metadata schemas do not include fields for "patent encumbrances" or similar, leaving a blind spot.

Lastly, current RML standards themselves have legal gaps in adoption. While ODRL is expressive enough for many scenarios, it has not been formally tested in courts or in widespread legal practice for research data. There may be uncertainty about how an ODRL policy would stand up as evidence of granting permission or imposing obligation as opposed to a conventional licence text. This could make institutions hesitant to rely solely on machine-readable policies without human-readable (and lawyer-readable) counterparts. In effect, we see a gap between legal text and code: the law is still largely conveyed in natural language contracts, and bridging that to a purely code-based expression is an ongoing challenge.

In summary, the legal gaps include a lack of automated licence compatibility checking, insufficient representation of jurisdictional differences and privacy requirements, incomplete coverage of all relevant legal rights in metadata, and an unease or unproven status of RMLs in legal enforcement. These gaps mean that legal certainty is not yet achieved in EOSC: users often must resort to manual legal analysis, and repository managers must issue human-readable policies to cover what metadata cannot. Improving RMLs and practices to fill these gaps is essential for EOSC's vision of seamless and trustworthy cross-border access (i.e. true legal interoperability).

5.2. Ethical, Soft-law, and Community Expectation Gaps

Beyond formal law, researchers and repositories must also contend with ethical norms, community guidelines, and "soft-law" principles (non-binding but influential frameworks). Current RML implementations and repository practices show clear gaps in how these aspects are handled.

One major gap is the expression of ethical usage conditions. Many datasets, especially in health, indigenous studies, or social sciences, come with ethical obligations: e.g., "for health research only," "not to be used to stigmatise or harm the community of origin," or "use requires approval by an ethics committee." These are often not encoded in any standard licence. For instance, consent forms for biomedical research might stipulate that data only

be used for a specific disease or only non-commercially, but a Creative Commons licence applied to the dataset cannot reflect those nuances. As discussed in Chapter 3, traditional licences fail to capture many such restrictions. The result is that repositories currently communicate ethical constraints through policy documents, user agreements, or simply not at all. EGA's use of Data Use Ontology (DUO) [5] is a rare example of standardising some ethical constraints (such as disease-specific use, or requiring Institutional Review Board (IRB)approval) in metadata. Generally, however, outside the genomics community, there is no widely adopted ontology for ethical conditions.

The CARE Principles for Indigenous Data Governance [9] have gained traction as a normative framework, but how would a repository indicate that a dataset involving indigenous communities requires adherence to CARE? At present, this would likely be handled through a custom disclaimer or an embargo until a proper agreement is in place–not through structured metadata or RML. Thus, the ethical gap is that machine-readable languages and schemas do not yet accommodate the subtleties of ethical consent, culturally sensitive access, or community-specific usage rights.

Related to this is the gap related to community expectations: guidelines and principles that are not encoded in law but are expected to be followed. In Open Science, examples of soft-law include the FAIR principles [10], TRUST principles [11], institutional codes of conduct for data sharing, or funding agency policies on open data. While these influence practice, they often do not translate into metadata. For instance, the FAIR principle that metadata should include clear usage licences is partially implemented (licence field present), but the principle that data should be "as open as possible, as closed as necessary" is an Open Science policy guideline that repositories attempt to follow without a direct way to represent the reasoning in metadata. We lack a vocabulary for statements like "this dataset is restricted because it contains locations of endangered species" or "access requires requester to be from academic institution as per funding mandate". Instead, repositories apply a blanket tag (e.g., restrictedAccess) and leave the justification hidden. This disconnect could lead to mistrust or misuse – consumers of data might assume "restrictedAccess" always means sensitive personal data, when in fact it could be for entirely different ethical reasons.

Another gap in current practice is the lack of representation of contributor or community rights. Researchers or communities who contribute data sometimes impose conditions (soft-law style) on its reuse. For example, a data collector might request to be informed of any secondary analysis (a courtesy, not a legal requirement), or a participant group might ask for a culturally appropriate consultation before reuse. In isolated cases, insistence on being listed as co-authors or contributors when a work is reused may be included in access conditions. These kinds of expectations are almost never captured in RML or metadata today. They appear, if at all, in textual agreements, unstructured terms of use, or memos.

There is a corpus of sensitive data that requires an agreed and consistent set of terms to describe elements of 'managed access' [35], expressed in all likelihood as a set of standardised prohibitions and obligations. The main focus of these terms is the protection of subject rights, and cover considerations such as recording the particulars of the end user,

verifying the research bona fides, validating the derived dataset before and/ or after creation of derived data to ensure subject rights are respected, and prohibition of use of the data outside controlled environments. In some cases, input data may not be seen by the end user, and only derived data is accessible. Aspects such as an accompanying anonymised or synthetic data set to enable research then becomes important [42], [44], and it will be useful to have a formal relationship defined between the closed dataset and its synthetic or anonymised versions. In effect, these managed access provisions obligate the end user to maintain the undertakings in respect of informed consent and/or ethics clearance related to the subjects of study.

Furthermore, ambiguities in terminology reflect an ethical gap as well as a legal one. Terms like “non-commercial” or “research use only” are ethically loaded (meant to enforce certain community norms) but poorly defined. Different communities interpret them in various ways – for example, does “non-commercial” permit use by a for-profit hospital for patient care research? Does “research use only” exclude research funded by industry? These ambiguities mean that even if such terms are used, a machine (or even another human) may misinterpret the allowed uses. Current RMLs do allow defining such terms more precisely via URIs, but the vocabulary must be agreed upon. Right now, *no consensus vocabulary exists for many of these soft-law constraints*. Projects and forums (such as the Research Data Alliance) have discussed standardised consent and use condition codes, but adoption is nascent.

In summary, the ethical and soft-law gaps in EOSC practice include: the inability to capture many non-legal but critical conditions in a structured way, the absence of widely adopted vocabularies for ethical constraints, and the risk of miscommunication due to vague terms. These gaps leave a range of important considerations “in the dark” for automated systems. As a result, compliance with ethical norms currently relies on human judgment and institutional oversight rather than being facilitated by metadata. To truly honor the principle of “responsible data sharing” in EOSC, future RML development must extend beyond legal permissions to encompass ethical permissions and prohibitions, bringing concepts from frameworks like CARE into the machine-readable realm.

5.3. Technical and Implementation Gaps

The third category of gaps lies in the technical realm–the limitations of current systems and standards in implementing a rich rights metadata framework. These gaps often explain why the legal and ethical gaps persist: if software platforms and standards do not support certain features, repositories can not easily adopt them. Key technical and implementation shortcomings include the following:

5.3.1. Inconsistent and Non-Standard Metadata Practices

As noted previously, repositories vary widely in how (and if) they record rights information. Some use Dublin Core’s simple dc.rights field with a free-text entry, others use DataCite’s rights property with a URI, and others have custom database fields. The result is that when aggregating metadata, one encounters everything from standard licence URLs to strings like

“CreativeCommons” or “All rights reserved”. Many records lack needed details such as the version of a licence or the expiration date of an embargo, simply because the local schema did not enforce those or the depositor omitted them.

This inconsistency is a major technical gap – even if we have RML standards available, they are not being uniformly applied. A related issue is lack of validation: repositories often do not validate the content of rights fields against a controlled list. A user might type “Creative Commons BY” or “CC-BY 4” and these would not automatically map to a proper identifier. Without validation and normalisation, the metadata cannot support automated reasoning. In essence, the current implementation gap is that rights-related metadata is frequently unstructured or non-interoperable by default, defeating the purpose of RML. As evidence, studies have found that a significant fraction of repository records have rights metadata that is either missing or not machine-understandable.

It is also not clear that given the same input considerations in respect of legal obligations, disclosure risk, ethical concerns, and policy-related drivers, that two repositories will select the same licence. Hence there is a need to map these considerations to the licences that will adequately support them, as proposed by DALICC and supported by published EU guidelines [37].

5.3.2. Limited Adoption of Existing RML Standards

While robust standards like ODRL exist, very few EOSC-linked repositories have implemented them in their workflows. Implementing an RML (like embedding ODRL policies in metadata records or allowing depositors to configure ODRL rules) requires significant development effort and user education. Currently, most repository software does not offer a user-friendly way to add complex rights rules. For example, no mainstream repository platform provides an interface for a depositor to say “allow reuse only in country X” or “obligation: cite project Y when using this data” in a structured way. These would have to be coded in an RML by an expert, which is not realistic for average researchers depositing data. Thus, a gap is the usability of RML tools: even if an advanced user knows ODRL, there is seldom a place to put ODRL XML/JSON in a repository submission form. Some initiatives (such as certain National Data Services) are toying with machine-actionable DMPs or using ODRL for data access policies, but these are pilots. The wider EOSC community has not yet integrated these into out-of-the-box repository solutions.

5.3.3. Lack of Integration with Infrastructure

Rights metadata (when it exists) is often not integrated into the active functions of infrastructure. For instance, ideally an access control system would read a machine-readable policy and automatically decide if a user can download a file. In practice, access control checks are usually binary (open or closed flag) and not tied to licence conditions. No mainstream repository will automatically, say, refuse a download because the user is commercial and the licence is non-commercial – because the system does not “know” the user’s status or reason about the licence. Similarly, discovery systems rarely use rights

information beyond filtering open vs closed. The more nuanced reasoning – such as "find datasets I can use commercially" or "alert me if I'm trying to combine incompatible data" – is not implemented. This is a technical gap where the **potential of RML is unrealised**: current EOSC practice does not yet incorporate rights reasoning engines or licence compatibility checkers into repository workflows.

5.3.4. Interoperability Limitations

Although OpenAIRE has standardised high-level categories for those repositories included in its graph resources, the example from the FAIR-IMPACT project shows that repositories still have bespoke notions of access levels [91]. A technical gap persists in mapping all these into a common framework beyond just open/closed. For truly granular interoperability, one repository's "safeguarded" or "controlled" access (terms used by, say, the UK Data Service) need to be translated to another's terms – and ultimately into a formal policy. At the moment, this mapping is shallow (just category labels). There is no global registry of standard licences and conditions that all EOSC services reference in a linked data manner. In other words, while efforts like SPDX (for software licences) or Creative Commons catalogs exist, EOSC lacks a **unified registry of rights terms** that covers the full spectrum (from well-known licences to local data use conditions). The DALICC project made some progress in this regard [93], but it is not a EOSC-endorsed, sustainable registry.

5.3.5. Dynamic and Life Cycle Management

An implementation challenge is presented by the dynamic nature of rights. Embargoes are lifted, consent can be withdrawn, policies can change. Current metadata is typically static; once a record is published, the licence field might not be routinely updated even if circumstances change. There is little technical infrastructure for versioning of rights metadata – for example, keeping track that "as of 2023 the data was restricted, as of 2025 it became open" in a machine-readable log. Guidelines in this respect exist [30], but are often not technically supported in all repositories, or not properly utilised if supported.

Without lifecycle management, automated systems might act on outdated information. Ensuring that updates propagate (e.g., if a depositor changes the licence of a dataset after initial publication) is another weak point. This raises governance questions (who has authority to change rights metadata after publication?) and technical ones (how to notify harvesters of changes). Most current repositories do allow edits to metadata post-publication, but there is inconsistency in whether such changes generate new identifiers or notifications. A comprehensive RML infrastructure would need to accommodate time-based and versioned policies, which is currently a gap.

There is also a real possibility that without periodic re-appraisal of rights and access provisions of repository holdings, changes in legislation, policies, or community expectations will not be applied. Trustworthy repositories [27], [28], for example as certified by CoreTrustSeal [23], actively curate and re-appraise legacy deposits, but they represent a small fraction of research data repositories. There is a high probability, for example, that

some legacy deposits have licences that do not support the legal obligations of GDPR properly.

In summary, the technical and implementation gaps highlight that while the *ideas* of machine-readable rights are known, the *execution* is lagging. We lack standardised, user-friendly implementations across repositories, integration of rights reasoning into systems, and consistent cross-platform metadata. Overcoming these gaps will require coordinated development – updating repository platforms to support richer rights metadata, establishing central vocabularies and registries, and building services that consume and act on this metadata. These challenges must be addressed for the vision of automated, EOSC-wide rights management to be realised.

5.3.6. Formalisation of Actors, Policies, and Links to Licences

National, network, funder, infrastructure, and domain or institutional policies typically impact on the rights and access afforded to end users of research outputs across several fronts, and the rights, duties and obligations of end users, curators, depositors, and subjects of study are more often than not scattered across multiple policies associated with a repository. It is a significant task to extract the provisions applicable to a specific actor or context from these documents, and it is further difficult to determine whether all relevant provisions have been captured adequately in a licence. ODRL [4] makes provision for formal linking of policy provisions and licence provisions, and both ODRL and schema.org [82] provides a partial formalisation of the actors (agents) involved in the rights and ethics landscape. Formalising and structuring policies [17], their links to licences, and explicitly defining and linking the actors involved is a possible and necessary improvement to established practice.

Group	Limitation or Gap	Description
Legal and Regulatory	Licence Compatibility	Inputs to a derived work may carry licences that are incompatible in combination
	Jurisdiction Specificity	Some provisions in law are not applicable globally, but regionally and in some cases, nationally
	No Vocabulary and/ or lack of granularity in expression	REL vocabulary originated from the perspective of copyright, and does not cover aspects of legal limitations, obligations, or restrictions
	Moral and Ethical Considerations	Moral and ethical obligations, such as proper attribution of input works, are not addressed adequately in some cases.
	Adoption gaps	Provisions such as ODRL are largely untested in many jurisdictions.
Ethical, Soft Law, and Community	Moral and Ethical Considerations	Moral and ethical obligations, such as proper attribution of input works, are not addressed adequately in some cases, and adequate vocabulary for its description is

Expectations Considerations		lacking
	Research Subject-related Considerations	Aspects such as alignment with CARE [9] principles cannot be adequately represented in terms of licence provisions
	TRUST and FAIR, Policies, Frameworks	Aspects such as alignment with FAIR [10] and TRUST [11] principles cannot be adequately represented in terms of licence provisions. The same applies to Open Science expectations or institutional or funder policies.
	"Managed Access"	Sensitive data requires new and standardised vocabulary to capture considerations related to managed access - recording identity of end users, verifying derived data sets for accidental disclosure, and processing datasets in controlled environments.
	Attribution and Acknowledgement	Research practice requires proper attribution, acknowledgement, and sometimes soft provisions include informing depositors about reuse, or, in isolated cases, an insistence on being listed as co-authors or contributors when a work is reused.
Technical and Implementation Considerations	Ambiguous terminology	Risk of miscommunication due to ambiguous terminology that is open to interpretation.
	Inconsistent metadata practices	A consistent term to identify rights and access information across metadata standards is lacking, and registry references (e.g. SPDX) to unambiguously identify licences are not commonly implemented.
	Limited adoption of REL	While robust standards such as ODRL exist, very few EOSC-linked repositories have implemented them in their workflows.
	Lack of Integration with Infrastructure	A technical gap where the potential of RML is unrealised: current EOSC practice does not yet incorporate rights reasoning engines or licence compatibility checkers into repository workflows.
	Interoperability Limitations	There is no global registry of standard licences and conditions that all EOSC services reference in a linked data manner, and EOSC lacks a unified registry of rights terms that covers the full spectrum (from well-known licences to local data use conditions).
	Change and Lifecycle Management	Most current repositories allow edits to metadata, including rights, post-publication, but there is inconsistency in whether such changes generate new identifiers or notifications, and provenance records are often not maintained. Re-appraisal to ensure that legacy deposits reflect current legal and other provisions are very rarely done.

Table 5: Summary of Gaps and Limitations

6. Synthesis: Findings and Recommendations

Bringing together the analysis from previous chapters, this chapter synthesises the key findings and outlines recommendations to align EOSC practices with an advanced rights management framework. It also discusses open issues and areas for future work. The overarching goal is to ensure that EOSC can support **machine-readable, interoperable, and comprehensive rights information** across its services, thereby enabling data to be shared and reused legally and ethically by design.

6.1. Key Findings

Several clear findings emerge from the state-of-the-art review:

- **Fragmentation of Practices:** Rights metadata practices in EOSC are currently fragmented and inconsistent. Some platforms (e.g., Zenodo, DataverseNL) enforce basic licence and access information, whereas others (e.g., many institutional repositories) may have minimal or ad hoc documentation of rights. Even within platforms that capture rights, the level of detail and structure varies. This fragmentation results in poor interoperability – a condition repeatedly highlighted in our analysis. A harmonised approach is needed to bridge these gaps.
- **Underutilisation of RML Standards:** Existing Rights Markup Languages (such as ODRL) and related standards (like Creative Commons REL or RightsStatements.org) are not yet deeply embedded in EOSC repository workflows. While the conceptual benefits of these standards (expressiveness, machine reasoning) are recognised, implementation is lagging. The study did not find any EOSC repository fully implementing ODRL policies in metadata at present – usage is limited to pilots or specific use cases. Instead, most rights information is captured in simpler metadata fields without the rich semantics an RML could provide.
- **Legal and Ethical Complexity vs. Simple Metadata:** EOSC content comes with a complex web of legal and ethical conditions, but current metadata captures only a small fraction of that complexity. Intellectual property rights (copyright, licences) are sometimes recorded (licence name/URL), but aspects such as data protection constraints, consent requirements, or ethical limitations are usually absent or hidden in documents. This leads to a disconnect: a dataset might appear “open” in a catalog because it has an open licence, yet in reality its use might be constrained by ethical norms not visible in the metadata (e.g., sensitive content requiring institutional approval or Indigenous data to which the CARE Principles are relevant). Without additional context, automated systems cannot distinguish such cases, posing a risk for compliance and lack of transparency for end users.
- **Importance of a Unified Approach:** A recurring theme is the need for a unified, machine-readable approach to rights in EOSC. The case studies and examples in Chapter 3 demonstrated how the lack of a common framework leads to friction and

manual effort. Key EOSC objectives such as facilitating cross-border data reuse and automating access depend on having a common language for rights. The findings affirm that developing an EOSC-specific Rights Markup Language (or profile of an existing one) is a logical step. This RML should be capable of expressing the variety of rights information identified: from standard licences to embargoes, from GDPR-related flags to community consent conditions.

- **Some Building Blocks in Place:** Despite gaps, certain building blocks can be leveraged. The widespread adoption of Creative Commons licences in open science is a strength – it provides a familiar baseline taxonomy of permissions/ obligations (BY, NC, ND, SA) that an RML can integrate (indeed, ODRL has vocabulary for such terms). OpenAIRE's metadata guidelines ensure that at least *access status* and *licence URI* are commonly shared metadata elements across many repositories. The DataCite schema (used by many EOSC repositories) has a slot for rights information that can carry URLs and text; this could convey richer RML statements if standardised. Furthermore, domain ontologies such as DUO (in EGA's case) show that community-specific rights vocabularies can be developed and used, pointing toward a federated but interoperable approach (different domains can have add-ons, but the core framework remains compatible).
- **Pilot Initiatives Indicate Momentum:** The fact that Zenodo and OpenAIRE are earmarked for pilot implementations of an EOSC RML is a key finding from the project's context. It demonstrates institutional buy-in and provides testbeds for implementation. Similarly, projects under the FAIR and EOSC umbrellas (e.g., FAIRsFAIR, FAIR-IMPACT) are actively exploring machine-actionable policy descriptions, indicating that the timing is ripe to push the envelope. The findings underscore that our work is aligned with a broader momentum in the community to solve these issues.

In summary, the state of the art reveals *both* a pressing need and a promising opportunity: EOSC needs a more sophisticated rights management approach to overcome current limitations, and many of the pieces required (awareness, standards, willingness, initial implementations) are already available or emerging.

6.2. Recommendations for EOSC Alignment

To address the identified gaps and build on the findings, we propose a set of recommendations aimed at aligning EOSC practices with a comprehensive rights management strategy. These recommendations span technical measures, policy-level actions, and metadata standardisation efforts, reflecting the multifaceted nature of the challenge.

1. **Develop and Adopt an EOSC-wide Rights Markup Language (RML) Vocabulary:** At the core, EOSC should finalise a dedicated RML vocabulary that extends or profiles existing standards (like ODRL) to meet EOSC-specific requirements. This RML should include terms for all key rights elements identified: standard licence permissions/ prohibitions, embargo periods, access conditions (open/ restricted/ closed),

obligations (attribution, citation of funding, notification of reuse), personal data indicators, ethical use constraints, and so on. Alignment with established ontologies is crucial – for example, reuse ODRL's classes and properties for permissions and duties, incorporate DataCite's terms for basic licence info, and include DUO or similar codes for consent requirements. The vocabulary should be published with persistent URIs and made available in machine-readable formats (RDF/OWL) to facilitate integration. Multilingual labels and definitions will help adoption across Europe. Technically, this vocabulary will serve as the *schema* for rights metadata in EOSC. We recommend a governance mechanism (perhaps under the EOSC Association or RDA) to maintain and evolve this vocabulary over time, as new needs arise.

2. **Agree on 'Managed Access' and New Vocabulary Terms Related to Sensitive Data**: While it will be possible for profiles and extensions of REL and RML to be implemented for specific cases, there is a corpus of sensitive data that requires an agreed and consistent set of terms to describe elements of 'managed access', expressed in all likelihood as prohibitions and obligations. The main focus of these terms are the protection of subject rights, and cover considerations such as recording the particulars of the end user, validating the derived dataset before and/or after creation to ensure subject rights are respected, and prohibition of use of the data outside controlled environments. In some cases, input data may not be seen by the end user, and only derived data is accessible.
3. **Integrate RML into Repository Platforms (Technical Integration):** Key EOSC services and repository platforms should be updated to support the new RML. This means modifying software like Invenio (Zenodo), Dataverse, DSpace, etc., to allow users to select and assign rights statements from the controlled vocabulary. For example, a depositor interface could present checkboxes or dropdowns not just for licence, but also for additional conditions (e.g., "Personal data present? Y/N", "Consent required for reuse? Y/N", "Non-commercial use only? Y/N"). The selections would then be encoded in the metadata as RML statements. We also recommend developing APIs and tools that can handle rights information: harvesting protocols (OAI-PMH, ResourceSync) should carry RML data, and packaging standards like RO-Crate should include a section for rights encoded in the RML. Ensuring that rights metadata is embedded at the technical level of EOSC – i.e., part of the fabric of data exchange – is critical. This could be facilitated by including RML support in the EOSC Core services; for instance, the EOSC Metadata Catalogue and PID services could have fields or linked data that reference the rights vocabulary.
4. **Mandate the Provision of Rights Metadata in EOSC Policy:** At the policy level, EOSC should mandate the provision of machine-readable rights metadata as a condition for repository onboarding or funding. Just as OpenAIRE's guidelines made *access rights* a required field, EOSC can raise the bar: any dataset, software, or resource in EOSC should include an RML-based rights statement. This does not force everything to be open; rather, it forces transparency. Even a restricted dataset should have an RML statement saying, for example, "restrictedAccess + reason (personal data)" instead of just being silently restricted. EOSC could incorporate this into its Rules of Participation and Data Quality guidelines. Additionally, funding agencies in Europe (through bodies like the EC) could require that projects use the EOSC RML for any

outputs they deposit. The benefit of a policy mandate is that it accelerates adoption and creates a network effect – if everyone is expected to do it, tools and support will quickly develop. Of course, a transition period and support (training, documentation) will be needed.

5. **Harmonise Metadata Schemas/Profiles to include new RML elements:** We recommend updating existing metadata schemas/profiles to include the new RML elements. For example, the next version of the OpenAIRE Guidelines should incorporate a section on how to embed the EOSC rights vocabulary (perhaps as an additional recommended field or as part of an extended rights element). DataCite could consider some of these elements in future schema versions or provide guidance on using relatedIdentifier or additional Metadata for rights. A crosswalk should be published showing how common repository metadata (Dublin Core, DataCite, schema.org) can convey the RML information, to ease implementation. In essence, the aim is that no matter what metadata standard a repository uses internally, there is a clear path to expose the standardised rights info to others. Part of this effort might include registering the EOSC RML in schema registries or adding it to the EOSC Interoperability Framework [24],[66],[73] as a recognised standard.
6. **Develop Technical Toolkit and Services:** EOSC should invest in developing a toolkit of services around the RML. One such service could be a licence compatibility checker: given two or more RML policies, it could algorithmically determine if they are compatible or highlight conflicts (e.g., two datasets being combined – the service could warn if one is CC BY-SA and the other is CC BY-NC, flagging a problem). Another service could be a policy reasoner integrated with discovery: users could query "show me datasets I can use for commercial purposes" and the system would filter by RML statements (selecting only those without a non-commercial restriction). Over time, more advanced uses are possible: automated enforcement where, for instance, a data access request system reads an RML and generates a tailored click-licence or access agreement for the user to agree to before download (covering any obligations like citation requirements or fees, if those were in the policy). These tools will make the RML "come alive" in daily use and demonstrate value to researchers and data providers.
7. **Establish Governance and Alignment with Legal Frameworks:** We also recommend establishing a governance body or working group to oversee the alignment of the RML with evolving legal and ethical standards. This group would liaise with legal experts (to ensure the RML remains legally sound), ethicists (to integrate new soft-law norms), and standards organisations like W3C or GA4GH for technical alignment. They would also handle versioning of the vocabulary. As laws change (e.g., say a new EU Data Act introduces data sharing obligations, or new copyright exceptions for text mining are enacted), the RML might need updates to include corresponding terms. Similarly, as the CARE principles get operationalised, the RML should include ways to mark data as subject to the CARE principles around access and restrictions. The governance group ensures EOSC's approach stays **up-to-date and authoritative**.
8. **Integrate into EOSC Interoperability Framework:** The EOSC Interoperability Framework (IF) [66] should include legal interoperability via RML as a core

component. This means when defining how EOSC resources interconnect, rights metadata is treated on par with technical metadata. The IF could provide guidelines on using the RML in all EOSC exchange layers. This integration at a high level would signal that rights metadata is not optional, but an infrastructure component of EOSC.

9. **Invest in Training and Community Engagement:** Finally, a recommendation on the human side: invest in training and community engagement regarding rights metadata. Researchers and repository managers need to understand *why* providing rich rights metadata is important and how to do it correctly. Data depositors need to be supported and trained on understanding and making access restrictions explicit and justified, and in determining how they will be applied in practice. The availability of data professionals, such as an institutional data steward, to provide this advice and support is important (see: https://doi.org/10.2218/ijdc.v19i1.1048). The data professionals, themselves, should be supported in their own knowledge of this evolving and complex area through 'train the trainer'-style engagement.

The introduction of the EOSC RML should be accompanied by clear documentation, webinars, possibly even integration into data steward curricula. Community feedback should be sought to refine the vocabulary – ensuring it covers real needs and is not overly burdensome. Early adopters (like Zenodo, DANS, etc.) should share their experiences in implementing the RML to guide others.

Collectively, these recommendations aim to create a virtuous cycle: clear standards and requirements (so everyone knows what to do), embedded in technology (so it's easy to do), enforced by policy (so it gets done), and supported by services (so the benefits are tangible). If implemented, EOSC would move towards a state where every digital object comes with a transparent, machine-readable "rights profile" that travels with the object and can be understood and honored by any service in the ecosystem. This will greatly enhance legal and ethical compliance in automated workflows and reduce friction for users trying to reuse content across the cloud.

6.3. Open Issues and Future Work

While the recommendations above provide a roadmap, there remain open issues and areas requiring further investigation as EOSC moves forward with enhancing rights management:

- **Handling of Complex and Evolving Policies:** One open challenge is how to manage very complex policies that might be conditional or context-dependent. For example, consider a dataset that is open for research uses for two years, after which it becomes completely open for any use (a time-based condition), or a software tool that is free for academic use but requires a fee for commercial use (a conditional monetary obligation). Encoding such multi-layered logic in RML is possible but can become complicated. Ensuring that the RML and associated tools can handle conditional, temporal, and composite policies gracefully is future work. This ties into the need for policy versioning and evolution over time. Mechanisms for updating rights metadata and propagating those changes (with proper version stamps and

perhaps notifications to users) need to be designed and implemented. Approaches like attaching validity periods to RML statements or linking to a canonical policy object that can be updated are potential solutions to explore.

- **User Identity and Credential Integration:** A machine-readable policy often needs to be evaluated against who the user is (for instance, "allowed for non-profit institutions" or "only accessible to EU citizens' data"). The concept of user attributes and credentials (like the GA4GH Passport and Visas approach for data access permissions) may need to interface with RML[23]. In other words, enforcement of rights might require the system to know something about the requestor. EOSC's AAI (Authentication and Authorisation Infrastructure) could be extended to carry attributes (like affiliation type, or researcher status) that can be matched with RML constraints (like "nonCommercialUseOnly" could be interpreted by checking if the user's affiliation is academic or if they agreed to a non-commercial pledge). Designing this integration is an open area – it likely will involve collaboration with identity providers and perhaps the use of attribute-based access control in combination with RML policies.
- **Scalability and Performance:** Introducing policy checks and reasoning into data discovery and access workflows raises questions of scalability. A future where every data request triggers a licence compatibility check or a dynamic consent evaluation means extra computation. We need to ensure that any RML reasoning engine is efficient, or that policies are structured to be easily checkable. Caching decisions, pre-computing compatibility matrices between common licences, and simplifying policy logic for machine consumption are techniques to consider. Benchmarking different approaches (ODRL in JSON-LD vs a rule-based engine vs hardcoded licence logic) could be a future task to ensure EOSC can operate at scale with these features enabled.
- **Community Acceptance and Behavioral Factors:** Even with technical solutions in place, an open issue is how the research community will respond. Some researchers might worry that too much emphasis on machine-readable rights could introduce additional bureaucracy or restrictions ("Will I have to fill out a complex form for every dataset deposit now?"). It's crucial to monitor the community's reception and iteratively refine the approach to maximise buy-in. Part of future work should involve pilot studies and user surveys to gauge understanding and comfort with the new system. It may turn out that certain terms are confusing or that users commonly misapply a particular restriction. This feedback loop will be necessary to adjust the vocabulary and guidance.
- **Legal Validity and Policy Coordination:** There is an open policy question about the ***legal status of machine-readable agreements***. For example, if a dataset is accompanied by an ODRL policy that says "no commercial use", is that enforceable in court by itself, or does it need a human-readable licence text to back it up? While projects like Creative Commons have tackled this (their licences have a legal text and a machine-readable tag, with the text being authoritative), our EOSC RML may include novel combinations that have not been tested. It might be worthwhile for EOSC to

[23] https://academic.oup.com/nar/article/50/D1/D980/6430505#

obtain legal opinions on certain representative policies or even to advocate for legal recognition of standardised machine-readable licences. Engaging with initiatives such as the Open Knowledge Foundation or W3C's legal group could help solidify the legal robustness of the RML approach. In parallel, coordination with global efforts (e.g., the CODATA-RDA legal interoperability guidelines, GA4GH consent frameworks, etc.) will ensure EOSC is not an island but part of a wider movement towards machine-actionable rights.

- **Addressing New Types of Digital Objects:** Future work should also consider that EOSC is not static; new types of digital objects and data use cases will emerge. For instance, how do we manage rights for AI models trained on EOSC datasets? If a model is essentially a derivative work of many data sources, can we propagate rights metadata through it? Or consider real-time data streams and IoT data in EOSC – licensing those for reuse is an evolving frontier. The RML and policies might need extension to cover such scenarios (perhaps more emphasis on attribution stacking for aggregated data, or licences for AI outputs). Keeping an eye on emerging technologies and their intersection with rights will be an ongoing task.
- **International Alignment:** EOSC, while European, will interact with non-EU repositories and initiatives. Ensuring the RML aligns or is at least interoperable with international standards is important for global research collaboration. This could be future work in standardisation bodies. For example, working with W3C to formally standardise the EOSC RML profile of ODRL, or with ISO if a need for an ISO standard arises (similar to MPEG-21 REL in concept, but updated for open science). The goal should be that EOSC's work can influence and integrate with global frameworks, not remain a custom solution only understandable within Europe.
- **Maintenance of Consent and Ethical Metadata:** A specific open question is how to keep ethical metadata current. Ethical approvals or community consents can be withdrawn. Should EOSC have a mechanism for a community to flag, for instance, that a dataset should be taken down or its usage narrowed because new concerns arose? This is partly a governance issue. Future work might involve creating channels for ethical oversight – e.g., a mechanism where Indigenous communities or data subject groups can provide feedback on the use of their data via EOSC, and that feedback could update the RML conditions (perhaps adding a new restriction). While this goes somewhat beyond pure technology, it underscores that rights management in EOSC will always have a human and societal element that needs integrating with the technical system.

In conclusion, the journey to a fully realised EOSC Rights Markup Language and aligned practice is just beginning. This report provided the state-of-the-art foundation and a blueprint moving forward (Deliverable D13.3), but much implementation and community work remains en route to the final EOSC digital rights ecosystem (to be delivered in D14.3). The key open issues identified – from technical scalability to legal validity – should be addressed in the next phase of the project. By tackling these challenges with the same interdisciplinary approach (legal, technical, ethical) taken in this research, EOSC can deliver a pioneering solution for digital rights management that not only serves Europe's researchers, but also sets an example for open science infrastructures worldwide.

References

[1] GDPR.EU (2025), Writing a GDPR-compliant privacy notice, https://gdpr.eu/privacy-notice/

[2] European Commission, 2018. Ethics and Data Protection, https://commission.europa.eu/system/files/2020-06/5._h2020_ethics_and_data_protection_0.pdf

[3] W3C (2025), Community Group Report - Data Privacy Vocabulary (DPV), https://w3c.github.io/dpv/primer/

[4] W3C. (2018). Open Digital Rights Language (ODRL) Version 2.2 Ontology. World Wide Web Consortium (W3C). https://www.w3.org/ns/odrl/2/

[5] Various (2025), Data Use Ontology, https://github.com/EBISPOT/DUO

[6] W3C (2016), Data on the Web Best Practices: Dataset Usage Vocabulary, https://www.w3.org/TR/vocab-duv/

[7] Vázquez-Flores, K. L., Montiel-Ponsoda, E., & Poveda-Villalón, M. (2022). EA-Ontology: Ethical Assessment Ontology. EKAW (Companion), https://ceur-ws.org/Vol-3256/paper7.pdf

[8] Vázquez-Flores, K. L. (2025), EAOntology, https://krnlet.github.io/

[9] Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020) 'The CARE Principles for Indigenous Data Governance', Data Science Journal, 19(1), p. 43. Available at: https://doi.org/10.5334/dsj-2020-043

[10] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[11] Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. Scientific Data, 7, 144. https://doi.org/10.1038/s41597-020-0486-7

[12] CoDATA (2025), Policy Brief: Data ethics and structural inequities in science, https://drive.google.com/drive/folders/13kbKucOcIbflTMzygaAZc-l0XtkqblY6

[13] CoDATA (2025), POLICY BRIEF:Ethics and Indigenous Data Governance, https://drive.google.com/drive/folders/13kbKucOcIbflTMzygaAZc-l0XtkqblY6

[14] CoDATA (2025), POLICY BRIEF: Privacy, https://drive.google.com/drive/folders/13kbKucOcIbflTMzygaAZc-l0XtkqblY6

[15] CoDATA (2025), POLICY BRIEF: Data Ethics and Research integrity, https://drive.google.com/drive/folders/13kbKucOcIbflTMzygaAZc-l0XtkqblY6

[16] ALA (2022), Cataloguing Code of Ethics, https://alair.ala.org/server/api/core/bitstreams/9923a196-d345-4244-a07c-19450965f167/content

[17] Hugo, W and Beck, J. Deliverable number: 5.2 - Guidance on Data Policy and Supporting Open Licences - Data Policy Considerations, Project: 730995 - Supporting EU-African Cooperation on Research Infrastructures for Food Security and Greenhouse Gas Observations (SEACRIFOG), 2019, https://dx.doi.org/10.13140/RG.2.2.11940.04485

[18] RightsStatements.org. (2019). Standardised International Rights Statements.

https://rightsstatements.org/en/

[19] European Union. (2016). 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). Official Journal of the European Union, L 119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[20] European Union. (2021). Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013. Official Journal of the European Union, L 170, 1–68. https://eur-lex.europa.eu/eli/reg/2021/695/oj

[21] Secretariat of the Convention on Biological Diversity. (2011). Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilisation to the Convention on Biological Diversity: Text and annex. United Nations Treaty Series, 3109, 3. https://www.cbd.int/abs/

[22] General Services Administration. (2020). Federal Data Strategy – Data Ethics Framework. U.S. Federal Government. Retrieved from https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf

[23] CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance (V01.00). Zenodo. https://doi.org/10.5281/zenodo.7051096

[24] European Commission, Directorate-General for Research and Innovation; EOSC Executive Board, FAIR & Architecture Working Group. (2021). EOSC Interoperability Framework (Report from the EOSC Executive Board Working Groups FAIR and Architecture). Publications Office of the European Union.

[25] European Commission, Directorate-General for Research and Innovation, & EOSC Executive Board. (2021). EOSC Rules of Participation. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1

[26] ALLEA – All European Academies. (2023). The European Code of Conduct for Research Integrity (Revised ed. 2023). Berlin: ALLEA. https://allea.org/code-of-conduct/

[27] Consultative Committee for Space Data Systems (CCSDS). (2024). Reference Model for an Open Archival Information System (OAIS) (Magenta Book, CCSDS 650.0-M-3, Issue 3, December 2024). Washington, D.C.: CCSDS Secretariat.

[28] Consultative Committee for Space Data Systems (CCSDS). (2011). Audit and Certification of Trustworthy Digital Repositories – Recommended Practice (Magenta Book, CCSDS 652.0-M-2, September 2011). Washington, D.C.: CCSDS Secretariat. https://public.ccsds.org/Pubs/652x0m2.pdf

[29] Cox, Krista L., "Metadata and Copyright: Should Institutions License Their Data about Scholarship?" (2017). Copyright, Fair Use, Scholarly Communication, etc.. 59. https://digitalcommons.unl.edu/scholcom/59

[30] van Horik, R., & Hugo, W. (2024). D3.3 - Guidelines for creating a user tailored EOSC Compliant PID Policy (V2.0 - DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION). Zenodo. https://doi.org/10.5281/zenodo.14092489

[31] Thorogood, A. (2022). GDPR brief: Data protection implications of publishing metadata to enable discovery. Global Alliance for Genomics and Health. https://www.ga4gh.org/news_item/ga4gh-gdpr-brief-data-protection-implications-of-publishing

-metadata-to-enable-discovery/

[32] RDA-CODATA Legal Interoperability Interest Group. (2016). Legal Interoperability of Research Data: Principles and Implementation Guidelines. Zenodo. https://doi.org/10.5281/zenodo.162241

[33] EOSC Association AISBL. (2025). EOSC Federation Handbook (First Edition, March 27, 2025). Zenodo. https://doi.org/10.5281/zenodo.14999577

[34] European Commission, Directorate-General for Research and Innovation. (2018). Ethics and data protection (14 November 2018). Luxembourg: Publications Office of the European Union. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/ethics-guide-data-protection_en.pdf

[35] Hugo, W., & van Kemenade, J. (2023, September 21). Harmonising access procedures for sensitive data [Conference presentation]. EOSC Conference. Retrieved from https://symposium23.eoscfuture.eu/wp-content/uploads/2023/09/Harmonising-Access-Procedures-for-Sensitive-Data.pdf

[36] PREMIS Editorial Committee, Rights Working Group. (2024). The 2022 revision of PREMIS rights white paper. Library of Congress. https://www.loc.gov/standards/premis/premis-rights-whitepaper-20240408.pdf

[37] Publications Office of the European Union. (n.d.). EU Data Licensing Assistant. Publications Office of the European Union. Retrieved August 8, 2025, from https://data.europa.eu/licensing-assistant/licences?locale=en

[38] CPP (to be provided by WP1)

[39] Center for Research Libraries, & Online Computer Library Center. (2025). Trustworthy repositories audit & certification: Criteria and checklist (TRAC) (Version 1.1, January 2025). Center for Research Libraries. https://www.crl.edu/sites/default/files/2025-01/trac.pdf

[40] European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act). Official Journal of the European Union, L 277, 1–102. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065

[41] Google Cloud. (n.d.). Data Loss Prevention. https://cloud.google.com/dlp

[42] Information Commissioner's Office (ICO). (2012). Anonymisation: Managing Data Protection Risk Code of Practice. ICO / UKAN. https://ico.org.uk/media/1061/anonymisation-code.pdf

[43] OECD. (2016). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. OECD Publishing. https://doi.org/10.1787/9789264255692-en

[44] UNECE. (2007). Managing Statistical Confidentiality & Microdata Access: Principles and Guidelines of Good Practice. United Nations. https://unece.org/statistics/publications/managing-statistical-confidentiality-microdata-access

[45] European Medicines Agency (EMA). (2018). External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. EMA/90915/2016. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en.pdf

[46] Statistics Canada. (2019). Statistics Canada's Modernization of Disclosure Control: Current Practices and Research Directions. Statistics Canada. https://www150.statcan.gc.ca/n1/pub/12-539-x/12-539-x2019001-eng.htm

[47] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., & de Wolf, P.-P. (2010). Handbook on Statistical Disclosure Control. ESSNet SDC / Eurostat.

https://ec.europa.eu/eurostat/cros/content/handbook-statistical-disclosure-control_en

[48] U.S. Department of Health & Human Services (HHS). (2012). Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. HHS. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

[49] Williamson, E., Walker, A. J., & Goldacre, B. (2020). OpenSAFELY: Transparent and Reproducible Analyses of Electronic Health Records. BMJ, 369, m1848. https://doi.org/10.1136/bmj.m1848

[50] Templ, M., Kowarik, A., & Meindl, B. (2015). Statistical Disclosure Control for Microdata: Methods and Applications in R. Springer. https://doi.org/10.1007/978-3-319-15072-9

[51] Hundepool, A., van de Wetering, A., Ramaswamy, R., Franconi, L., Polettini, S., Capobianchi, A., ... & Giessing, S. (2012). μ-ARGUS User Manual. Statistics Netherlands. https://research.cbs.nl/casc/documentation.html

[52] ICPSR. (2022). ICPSR Deposit Agreement. Inter-university Consortium for Political and Social Research. https://www.icpsr.umich.edu/web/pages/deposit/agreement.html

[53] CESSDA ERIC. (2017). CESSDA Deposit Guidelines. https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide

[54] UK Data Service. (2023). Depositing Data: Guidance for Depositors. https://ukdataservice.ac.uk/manage-data/deposit-data/

[55] OECD. (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD Publishing. https://doi.org/10.1787/9789264034020-en

[56] Hugo, W. (2024). Case Study Analysis: FAIR-AWARE (0.4 Draft). Zenodo. https://doi.org/10.5281/zenodo.13486326

[57] R.A.K. Richardson, S.S. Hong, J.A. Byrne, T. Stoeger, & L.A.N. Amaral (2025), The entities enabling scientific fraud at scale are large, resilient, and growing rapidly, Proc. Natl. Acad. Sci. U.S.A. 122 (32) e2420092122, https://doi.org/10.1073/pnas.2420092122.

[58] Abelson, H., Doctorow, C., & Lessig, L. (2008). ccREL: The Creative Commons Rights Expression Language. Creative Commons. https://opensource.creativecommons.org/ccrel/

[59] Berners-Lee, T., & Fischetti, M. (1999). Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor. Harper San Francisco. https://dl.acm.org/doi/10.5555/554813

[60] Creative Commons. (n.d.). About CC licences. https://creativecommons.org/share-your-work/cclicences/

[61] Creative Commons. (n.d.). ccREL: The Creative Commons Rights Expression Language. Retrieved from https://wiki.creativecommons.org/wiki/CcREL

[62] DANS-KNAW. (2019). DataverseNL general terms of use. https://dans.knaw.nl/wp-content/uploads/2023/03/DANS-DataverseNL-ToU.pdf

[63] DANS-KNAW. (n.d.). Privacy declaration. https://dans.knaw.nl/en/privacy-declaration/

[64] DataCite Metadata Working Group. (2024). DataCite Metadata Schema 4.6. DataCite. https://schema.datacite.org/meta/kernel-4.6/

[65] DataCite. (2021). Metadata Schema for the Publication and Citation of Research Data, Version 4.4. https://schema.datacite.org/meta/kernel-4.4/

[66] EOSC Executive Board. (2021). EOSC Interoperability Framework v3.0. https://eosc.eu

[67] European Commission. (2019). European Open Science Cloud (EOSC) Strategic Implementation Plan. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/2c1e5ff9-9b69-11e9-9d01-01aa75ed71a1

[68] European Commission. (2021). Turning FAIR into Reality. European Union Publications Office. https://op.europa.eu/s/ovXx

[69] European Genome-phenome Archive (EGA). (n.d.). How to request data. https://ega-archive.org/access/request-data/how-to-request-data/

[70] European Genome-phenome Archive (EGA). (n.d.). Policy documentation. https://ega-archive.org/access/data-access-committee/policy-documentation/

[71] European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). Official Journal of the European Union. https://gdpr-info.eu/

[72] Europeana. (n.d.). Usage guidelines for metadata. Europeana. https://www.europeana.eu/en/rights/usage-guidelines-for-metadata

[73] FAIRCORE4EOSC Consortium. (2023). Deliverable D2.3: EOSC Interoperability Specification Alignment. https://faircore4eosc.eu/

[74] Pellegrini, T., Höhne, J., de Boer, V., In der Maur, M., & Ingensand, S. (2019). Genealogy and Classification of Rights Expression Languages: Preliminary Results. (Preprint), https://penni.wu.ac.at/papers/IRIS%202018%20Genealogy%20and%20Classification%20of%20Rights%20Expression%20Languages%20Preliminary%20Results.pdf?utm_source=chatgpt.com

[75] Global Indigenous Data Alliance. (2019). CARE Principles for Indigenous Data Governance. https://www.gida-global.org/care

[76] Iannella, R., & Villata, S. (2018). ODRL 2.2 Ontology. W3C Recommendation. Retrieved from https://www.w3.org/TR/odrl-vocab/

[77] International Organization for Standardization (ISO). (2004). ISO/IEC 21000-5:2004: Information technology – Multimedia framework (MPEG-21) – Part 5: Rights Expression Language. https://www.iso.org/standard/36095.html

[78] MPEG Group. (2006). MPEG-21 Part 5: Rights Expression Language. ISO/IEC 21000-5:2004.

[79] OpenAIRE. (2023). Zenodo Metadata Guidelines v4.5. https://zenodo.org/help/metadata

[80] OpenAIRE. (n.d.). OpenAIRE guidelines for literature repositories, data archives, and CRIS managers. https://guidelines.openaire.eu/

[81] Pielke, R. A. (2007). The Honest Broker: Making Sense of Science in Policy and Politics. Cambridge University Press.

[82] Schema.org. (n.d.). Schemas for structured data on the internet. https://schema.org/

[83] Schultes, E. A., & Wittenburg, P. (2019). FAIR Data and Semantic Interoperability. Data Intelligence, 1(1), 40–47.

[84] Szabo, N. (1997). Formalizing and Securing Relationships on Public Networks. First Monday, 2(9). https://doi.org/10.5210/fm.v2i9.548

[85] Thorpe, D. ,Braukmann, R., and Maineri, A. (2025) . "Improving the Transparency of Data Access Conditions in the SSH Domain: Recommendations based on a small-scale analysis of the conditions applied to restricted access datasets." International Journal of Digital Curation 19.1 (2025). https://doi.org/10.2218/ijdc.v19i1.1048

[86] TRIPLE Project. (2022). D7.4: Policy Recommendations for SSH Open Science.

https://project.gotriple.eu

[87] Van Reisen, M., et al. (2022). Ethical Considerations in Digital Rights Management in EOSC. Journal of Open Science Policy, 3(1), 45–61.

[88] Zenodo. (n.d.). About records. https://help.zenodo.org/docs/deposit/about-records/

[89] Zenodo. (n.d.). Terms of use. https://about.zenodo.org/terms/

[90] Zenodo. (n.d.). Zenodo policies. https://about.zenodo.org/policies/

[91] Bell, D. (2024), Advancing access interoperability with ODRL, https://fair-impact.eu/use-cases/advancing-access-interoperability-odrl

[92] Rouchon, O., Kraaikamp, E., Gonzalez, E., Fink Kjeldgaard, A. S., Pedersen Tenderup, N., Davidson, J., Hodson, S., Rettberg, N., & Scharnhorst, A. (2024). D6.2 - Core metadata schema for legal interoperability (Version v1). Zenodo. https://doi.org/10.5281/zenodo.11104269

[93] DALICC (2025), Vocabulary Documentation, https://dalicc.github.io/#dalicc-vocabulary-documentation

[94] SPDX (2025), The System Package Data Exchange, https://spdx.dev/

Appendix A

A.1 Zenodo

Zenodo is a general-purpose open science repository operated by CERN and supported by OpenAIRE, widely used for sharing datasets, publications, and software.

Licensing Metadata: Zenodo requires depositors to specify a licence for every publicly accessible[24]. By default, the recommended licence is Creative Commons Attribution 4.0 International (CC BY 4.0), though users can choose from a broad list of standard licences (sourced from the SPDX list) or even enter a custom licence if needed[25]. This licence information is stored in the metadata and is exposed via Zenodo's APIs and OAI-PMH feeds, typically as part of the DataCite metadata schema[26]. All metadata in Zenodo is made openly available under CC0, ensuring it can be harvested and reused without restriction[50].

Access Levels and Restrictions: Zenodo supports multiple access levels for uploaded files: open access, embargoed access, restricted access, and closed access[50]. Open access files are immediately downloadable by anyone. Embargoed files remain temporarily closed and are automatically released on the specified embargo end date, a feature that ensures time-limited restrictions are honored in a consistent way. For restricted files, Zenodo allows depositors to share content on a case-by-case basis: a requester must be approved by the depositor to gain access. Completely closed files (intended for archival or non-distribution) can be deposited as well, although in such cases Zenodo permits the licence information to be provided in the free-text description field since the files themselves are not publicly accessible. In general, however, use and reuse of any Zenodo-hosted content is governed by the licence chosen by the depositor, and users are expected to comply with those terms.

Current Limitations: While Zenodo's enforcement of a licence field improves consistency and ensures every public record has an associated reuse permission, the platform's rights metadata is relatively basic. It records the licence (typically via a standard URI or identifier) and an access category (open/restricted/etc.), but does not natively capture more granular conditions or obligations. Complex usage terms—such as an obligation to notify the creator upon reuse, limitations to research-only purposes, or jurisdiction-specific conditions—are not expressible in Zenodo's metadata schema beyond what the licence itself covers. For example, Zenodo does not implement a full-fledged Rights Expression Language; any additional restrictions beyond standard licence terms (e.g. "available upon request to certain users") must be handled manually or noted in prose. This means that Zenodo currently lacks machine-actionable encoding of detailed policies such as ethical restrictions or specific Data Use Agreements. As noted in the previous chapter, Zenodo's approach represents a "partial" implementation of rights metadata: it ensures a basic level of information (licence and access status) is always present, but richer semantics (e.g. obligations or context-specific restrictions) are not yet supported in a structured way.

---

[24] uploadhelp.zenodo.org

[25] help.zenodo.org

[26] about.zenodo.org

Nevertheless, Zenodo provides a strong foundation for rights metadata within EOSC. It aligns with OpenAIRE guidelines (exposing standardised fields for access rights and licence[27]) and is actively exploring enhancements. Notably, Zenodo is one of the platforms identified for pilot implementation of an EOSC-specific Rights Markup Language in future work. This indicates a recognition that while current practices cover the basics, there is room to integrate more expressive, machine-readable rights descriptions into Zenodo's infrastructure.

A.2 DANS / DataverseNL/ Data Stations

DANS (Data Archiving and Networked Services) is a leading research data repository manager in the Netherlands and operates **DataverseNL** (a generalist repository) and 4 domain-specific repositories ('**DANS Data Stations**). These platforms are based on Harvard's Dataverse software. As such, its practices shed light on how an institutional data archive approaches rights management within EOSC.

Structured Metadata for Rights: The Dataverse platform used by DANS provides dedicated metadata fields for licensing and access conditions at both dataset and file level. Depositors are required to choose a licence for their dataset, much like Zenodo. The interface allows selection of standard licences (e.g. Creative Commons licences) and also supports the addition of custom licence text if needed. In DANS's workflow, once a depositor fills in the basic metadata and saves a draft, they can specify the usage terms under a "Terms" or "Terms of Use" section. This includes selecting or editing the licence and setting any access restrictions. The selection of a licence (or waiver) is mandatory for open datasets, aligning with the principle that reuse conditions must be transparent.

Access Restrictions and Embargoes: DANS supports both dataset-level and file-level access controls. Depositors may mark content as open access, embargoed, or restricted. A unique aspect of DataverseNL is the option to enable "access requests" for restricted files. By default, if a file is restricted, the system allows potential users to send a request to the repository (and thus to the data owner or curator) to gain access. DANS actively promotes the use of the request mechanism as the standard approach for restricted data, rather than outright denial of access. If a depositor chooses to disable the "enable access request" option (opting not to allow even requests), then they are required to supply a custom Terms of Access statement explaining the conditions under which the data might be available. This information is stored as a controlled text field – it ensures that some documentation of restrictions is present, but it is not a formally standardised vocabulary.

On download, irrespective of access limitations or not, the end user is requested to confirm that they understand the licence conditions and will abide by it. This assent is logged by the system.

DANS also permits embargoes to be set on files, meaning data will be automatically released after a specified date. Embargo periods can be defined (the repository recommends a maximum of two years) and, while the system does not strictly enforce a hard limit on the length, curatorial practice at DANS encourages reasonable embargo

[27] about.zenodo.org

durations. During the embargo, files remain inaccessible to the public, and their metadata will typically indicate the embargo end date. Once that date passes, the files become open access without manual intervention.

Personal Data and Ethical Information: Reflecting a concern for GDPR and ethics, DataverseNL's deposit interface asks the depositor to declare whether the dataset contains personal data (Yes/No/Unknown). This flag, while not a licence or rights per se, is an important piece of rights-relevant metadata, as it signals potential privacy constraints. A curator will verify this declaration, and it informs the handling of the dataset (for instance, a dataset flagged as containing personal data may undergo extra scrutiny or require appropriate access controls). Aside from this, any additional ethical or legal conditions (such as informed consent requirements, or allowed usage contexts) typically must be written in the free-text "Terms of Access" or in accompanying documentation, since the standard metadata schema has limited structured fields for such nuances.

Current Limitations: DANS's approach illustrates an advanced implementation of rights metadata within a repository platform: it goes beyond a single licence field and incorporates embargo dates, access request workflows, and even flags for personal data. The use of controlled fields (like checkboxes or dropdowns for access status and licence) improves interoperability – for example, the metadata can be harvested in a consistent way (DANS exposes its records via OAI-PMH in compliance with OpenAIRE guidelines, including fields for access rights and licence). However, there are still limitations. Custom access conditions such as "available upon request" or "restricted to registered users" are essentially encoded as text phrases rather than as part of a formal ontology. This means external systems or automated agents cannot easily interpret those conditions – they would not know, without human intervention, what "available upon request" entails or how to act on it. The lack of a machine-readable policy expression (e.g. something analogous to ODRL) for these conditions is a gap. Moreover, like many repositories, DANS relies on depositors (and curators) to correctly apply the appropriate restrictions and licences. As a recent small-scale study led by DANS indicates, clear procedures and decision-making guidelines on the application of restrictions are in fact often not applied [85]. Misclassification or omissions (such as neglecting to set a restriction on sensitive data, or using an ambiguous licence label) can occur, and there is no automated validation against an external rights vocabulary to catch such issues.

In summary, DANS demonstrates many best practices (mandatory licence, embargo and restriction options, personal data flagging) which align well with FAIR principles. It highlights that EOSC repositories are actively trying to balance openness with necessary restrictions. Yet it also underscores the need for a richer rights markup: one that could standardise terms like "restricted access – available on request" in a way that machines can uniformly interpret, and that could convey complex conditions (e.g. "for non-commercial research only") beyond what a simple licence name can achieve. These insights from DANS will feed into the requirements for an EOSC-wide Rights Markup Language.

A.3 European Genome-Phenome Archive (EGA)

The European Genome-Phenome Archive (EGA) represents a very different use-case within the EOSC ecosystem: it is a federated repository for human biomedical data, much of which is sensitive personal data (genomic sequences linked to phenotypic information). EGA's content is typically not openly accessible; instead, it operates under a controlled access model to protect privacy and comply with ethical and legal requirements. Examining EGA's practices provides insight into how highly regulated data is managed with respect to rights and conditions.

Controlled Access via Data Access Committees: Unlike Zenodo or DANS, where depositors set licences and basic access conditions, EGA outsources the access decision to Data Access Committees (DACs). Each dataset (or study) in EGA is associated with a DAC – usually a body of experts or stewards from the data-providing institution–who evaluate requests from researchers wishing to use the data. All data in EGA that contains personal information is encrypted and cannot be downloaded unless the user has been granted explicit permission by the DAC. The process is as follows: a prospective user finds a dataset of interest on the EGA catalog, then applies to the relevant DAC (contact info is provided on the dataset's page) with a usage proposal or agreement[28]. The DAC reviews the application (often ensuring the requestor's research intent is in line with the consent under which the data was collected) and if approved, EGA will allow that user to access the files via a secure account loginebi.ac.uk. Along with approval, the requestor typically must sign a Data Access Agreement – a legally binding document that stipulates how the data can be used, stored, and shared. This agreement might include clauses such as: the data can only be used for the approved research purpose, no attempt will be made to re-identify individuals, no commercial use is permitted without further consent, results must not infringe privacy, etc. Importantly, these conditions are decided on a case-by-case basis by the DAC and are not enforced by technical means in the EGA platform; rather, they rely on the legal agreement and trust in the researcher's compliance.

Data Use Ontology (DUO) Annotations: Recognising the need to communicate usage restrictions more systematically, EGA has adopted the Data Use Ontology (DUO) [5] for annotating **datasets**. DUO is a standardised vocabulary developed by the Global Alliance for Genomics and Health (**GA4GH**) to codify common data use restrictions in a machine-readable way (consent codes and requirements)ega-archive.org. For example, DUO includes terms like "General Research Use" (meaning the data can be used for any research purpose), "Health/Medical/Biomedical research only", "Non-Commercial Use Only", "Ethics Committee approval required", "Time limit on use", etc., each with a unique identifier. EGA curators apply relevant DUO codes to datasets based on the consent forms and access policies in place. These codes are then displayed on the dataset's landing page to inform potential requestors about core restrictions before they apply. For instance, a dataset might be tagged with DUO:0000042 (General Research Use permitted) and DUO:0000019 (Publication Required, meaning researchers must publish results) and DUO:0000021 (IRB approval required). A search interface can leverage these codes to allow users to find datasets they are eligible to use (e.g. searching for datasets open to general research versus

[28] ebi.ac.uk

those limited to disease-specific research). The DUO implementation is a significant step toward machine-readable rights metadata in the EGA context, as it standardises the way conditions are described.

Current Limitations: Despite DUO, the actual enforcement of restrictions in EGA remains procedural and human-mediated. The repository itself does not automatically prevent a researcher from doing disallowed actions with the data (for example, nothing technically stops someone from attempting a use outside the agreed purpose once they have the data). Compliance is ensured through the combination of legal agreements and the oversight of DACs. From a metadata perspective, EGA's use of DUO is exemplary for discoverability and clarity, but DUO covers only a set of predefined consent scenarios. If a dataset has highly specific conditions (e.g., "must collaborate with the data provider" or "destroy data after 3 months"), these might not be fully captured by current DUO terms and may instead be written in free text or only in the Data Access Agreement. Moreover, EGA's model is inherently closed: it does not integrate with general-purpose metadata aggregators for rights information because each dataset is essentially "closedAccess" with qualifications. In OpenAIRE's terms, most EGA datasets would be classified as restrictedAccess or closedAccess, with no licence URL, since no blanket licence applies (the 'licence' is individually negotiated per user).

That said, EGA's experience underscores several important points for EOSC rights management. First, certain data (e.g., human genomic data) will likely always require bespoke access control and human judgment in the loop due to ethical and legal complexities. Machine-readable policy languages will need to interoperate with these human processes (for example, a future EOSC RML could incorporate DUO codes as part of its vocabulary, to at least label sensitive datasets with standardised restrictions). Second, EGA demonstrates a best practice in using ontologies to bridge legal text and metadata: the DUO codes effectively translate consent form permissions into tags that machines can understand. This approach could be broadened in EOSC to cover other domains (e.g., an ontology for indigenous data sovereignty or for dataset-specific embargo policies). In summary, EGA's current practice is a hybrid of formal ontology use and traditional agreements, reflecting the state of the art for sensitive data management in EOSC.

A.4 OpenAIRE Graph

OpenAIRE Graph is not a repository for data or publications per se, but rather a federated infrastructure that aggregates metadata from hundreds of repositories and publishes guidelines to ensure interoperability. It plays a key role in EOSC as a metadata broker and as a setter of standards for repository managers. Therefore, OpenAIRE's practices regarding rights information are crucial for aligning EOSC-wide approaches.

Metadata Guidelines for Rights: OpenAIRE's guidelines mandate that every metadata record exposed to the OpenAIRE Graph harvester include a standardised indication of Access Rights. In practice, this means repository managers must provide a metadata field specifying whether a resource is open access, embargoed, restricted, or closed[29]. OpenAIRE relies on a

[29] https://guidelines.openaire.eu/en/latest/

controlled vocabulary for this field, specifically the info:eu-repo/semantics/[30] terms: openAccess, embargoedAccess, restrictedAccess, and closedAccess[31]. This standardisation is incredibly important for discovery: it allows the OpenAIRE Graph portal (and by extension EOSC search tools) to filter results by availability (e.g., showing only open materials vs. those requiring login or permission). Unlike some local repository categorisations which might use custom labels, the OpenAIRE vocabulary ensures a common understanding across all participants in the networkfair-impact.eu. In fact, including an access rights tag is mandatory in OpenAIRE's Application Profile (even though it might be an optional field in the native schema like Dublin Core or DataCite) – this requirement has driven many repositories to add or map that information.

In addition to the access category, OpenAIRE strongly encourages exposing the licence of the resource in a machine-readable way. The guidelines suggest using the rights element (or a specific sub-element like licence in DataCite) to include a licence URI (for example, a Creative Commons URL) and a human-readable name. An example in the guidelines shows a record having two rights entries: one for the access rights and one for the licence (**Table 4**). This dual approach distinguishes between the availability of the object (open vs closed) and the permissions to reuse it (licence terms). In OpenAIRE's metadata aggregation, the licence is not strictly required, but including it greatly enhances the reusability and clarity of the metadata.

Rights	<rights rightsURI="info:eu-repo/semantics/openAccess"/>)
Licence	<rights rightsURI="https://creativecommons.org/licences/by/4.0/"> Creative Commons Attribution 4.0 International </rights>

Table 6: Examples of OpenAIRE Rights and Licence References

Policy and Harmonisation Role: As a central EOSC service, OpenAIRE not only aggregates metadata but also disseminates best practices. Its guidelines have effectively harmonised how rights are labeled across European repositories. For instance, repository software platforms such as DSpace, EPrints, and Invenio (which Zenodo uses) have plugins or configurations to comply with OpenAIRE's rules, meaning they map local fields (like DSpace's dc.rights or Invenio's access flags) to the openAccess/embargoedAccess vocabulary. This is a form of soft standardisation: without enforcing a particular licence or policy, OpenAIRE ensures the description of the rights status is uniform. It also means that if one repository uses a term like "available upon request" and another uses "restricted access", both would be normalised under restrictedAccess in OpenAIRE's index, avoiding confusion.

Current Limitations: OpenAIRE's focus is on high-level access status and basic licence info, which leaves out a lot of nuance. The controlled vocabulary does not express why something is restricted or what conditions apply. For example, two datasets might both be

[30] https://guidelines.openaire.eu/en/latest/literature/field_accesslevel.html

[31] guidelines.openaire.eu

labeled “restrictedAccess”; one might be because it contains personal data (hence only accessible via a Data Access Committee), another might be because of a commercial embargo by a publisher, and a third simply because the depositor chose to limit access. These differences are not captured by the simple tag. Similarly, the licence field typically only captures standard licences; it does not cover bespoke conditions or the presence of any obligations beyond the licence text. OpenAIRE is aware of these gaps – its guidelines point out that repositories can add additional rights statements (using Dublin Core or DataCite fields) for more information[32], but this is not yet standardised across repositories.

From the EOSC perspective, OpenAIRE serves as an important bridge: it can propagate any new standards for rights metadata across a wide network. Indeed, OpenAIRE is involved in EOSC projects looking at machine-readable rights. Zenodo and OpenAIRE have been identified as pilot platforms to apply the proposed EOSC Rights Markup Language. One could envision, for example, an extension to the OpenAIRE guidelines in the future that would allow or require an RML policy expressed in a format like ODRL, attached to each record. For now, however, the state of practice is that OpenAIRE ensures minimum compliance (everything tagged with access rights, licence where possible) and interoperability (everyone uses the same tags), but not full expressiveness.

In summary, OpenAIRE’s current practices enforce a baseline of consistency in rights information across EOSC repositories. This baseline is a necessary foundation: it makes clear which resources are open vs. closed and under what licence they are released, facilitating basic legal interoperability (e.g., allowing portal users to filter search results by “open access” and to see the licence before downloading). The challenge that remains is moving beyond this baseline to incorporate richer rights metadata. The groundwork laid by OpenAIRE will be invaluable for the adoption of an EOSC-wide Rights Markup Language, as any new metadata will likely be integrated into OpenAIRE’s harvesting and guidelines to achieve widespread uptake.

[32] guidelines.openaire.eu