

Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	<a href="https://fair-impact.eu/">https://fair-impact.eu/</a>

## D6.2 – Core metadata schema for legal interoperability

Work Package	<b>WP 6, Interoperability</b>
Lead Author (Org)	<b>Olivier Rouchon (CNRS)</b>
Contributing Author(s) (Org)	<b>Esteban Gonzalez Guardia (UPM), Nicolaj Pedersen Tanderup (DeiC), Emilie Kraaikamp (DANS-KNAW), Joy Davidson (UEDIN, DCC), Simon Hodson (CODATA), Najla Rettberg (RDA), Anne Sofie Fink (DeiC), Andrea Scharnhorst (DANS-KNAW)</b>
Due Date	<b>2024-05-31</b>
Date	<b>2024-05-02</b>
Version	<b>V1.0</b>

### Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

## Versioning and contribution history

Version	Date	Author	Notes
0.1	2023.05.12	Olivier Rouchon (CNRS)	TOC and V0.1
0.2	2023.10.30	Olivier Rouchon (CNRS), Nicolaj Pedersen Tanderup (DeiC), Emilie Kraaikamp (DANS-KNAW), Joy Davidson (UEDIN, DCC), Simon Hodson (CODATA), Najla Rettberg (RDA), Andrea Scharnhorst (DANS-KNAW), Esteban Gonzalez Guardia (UPM)	Section #3 and #4 added, ready for internal milestone M6.4
0.3	2024.02.28	Olivier Rouchon (CNRS), Anne Sofie Fink (DeiC), Esteban Gonzalez Guardia (UPM)	Section #5 added
0.4	2024.03.22	Olivier Rouchon (CNRS), Anne Sofie Fink (DeiC), Esteban Gonzalez Guardia (UPM)	Section #6 and conclusion, version ready for internal review
0.5	2024.04.12	Olivier Rouchon (CNRS), Anne Sofie Fink (DeiC), Esteban Gonzalez Guardia (UPM), Liisa Marjamaa-Mankinen (CSC), Agnes Jasinska (DCC), Giacomo Cannizzaro (SURF)	Feedback from T6.2 internal review integrated, ready for final review
1.0	2024.05.02	Olivier Rouchon (CNRS), Anne Sofie Fink (DeiC), Esteban Gonzalez Guardia (UPM), Emilie Kraaikamp (DANS-KNAW), Marlon Domingus (EUR), Javier de la Cueva (Instituto de Empresa University)	Feedback by invitation of FAIR-IMPACT PCO

### Disclaimer

FAIR-IMPACT has received funding from the European Commission's Horizon Europe funding programme for research and innovation programme under the Grant Agreement no. 101057344. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.



## Table of content

---

1. Executive Summary	6
2. Introduction	7
3. Illustration of the complexity of the legal interoperability landscape	9
4. Schemas and vocabularies under consideration	12
4.1 Evaluation criteria	12
4.2 Dublin Core Metadata Element Set & Terms	13
4.3 DataCite Metadata Schema	14
4.4 Research Outputs Metadata Schema	16
4.5 Asset Description Metadata Schema	17
4.6 Open Digital Rights Language	18
4.7 Data Privacy Vocabulary	20
4.8 FAIRSharing Policy Schema	21
4.9 Data Catalog	23
4.10 OpenAIRE Guidelines	24
4.11 Data Documentation Initiative	26
4.12 RO-Crate	28
5. Use cases	30
5.1 EMBL - BY-COVID use case	30
5.2 SIKT - ESS use case	31
5.3 UKDS - CESSDA use case	34
5.4 UPM - RELIANCE use case	37
6. Core metadata schema for legal interoperability	39
7. Conclusions and next steps	41

## Terminology

Terminology/Acronym	Description
ADMS	Asset Description Metadata Schema
AP	Application Profile
AT	Authority Table
CC	Creative Commons
CCE	Common Conditions of use Elements
CDI	Cross-Domain Integration
CESSDA	Consortium of European Social Science Data Archives
CRIS	Current Research Information System
DCAT	Dataset Catalogue
DDI	Data Documentation Initiative
DCMI	Dublin Core Metadata Initiative
DCMES	Dublin Core Metadata Element Set
DMP	Data Management Plan
DPV	Data Protection Vocabulary
DoA	Description of Action
DUC	Digital Use Conditions
DUO	Data Use Ontology
EC	European Commission
EEA	European Environment Agency
EGA	European Genome Archive
EIF	European Interoperability Framework
EMBL	European Molecular Biology Laboratory
ENVRI	Environmental Research Infrastructures
EOSC	European Open Science Cloud
ESS	European Social Survey
FAIR	Findable Accessible Interoperable Reusable
GA	Grant Agreement to the project
GDPR	General Data Protection Regulation
ICO	Informed Consent Ontology
ISNI	International Standard Name Identifier
JSON	JavaScript Object Notation
MSCR	Metadata Schema and Crosswalk Registry
ODRL	Open Digital Rights Language
ORCID	Open Researcher and Contributor ID
RDF	Resource Description Framework
RIOXX	Research Output Metadata Schema
RO-Crate	Research Object Crate



Terminology/Acronym	Description
RPO	Research Performing Organization
SEMIC	Semantic Interoperability Community Europe
SSHOC	Social Sciences & Humanities Open Cloud
UKDS	UK Data Service
URI	Uniform Resource Identifier
VIAFID	Virtual International Authority Identifier
W3C	World Wide Web Consortium

## 1. Executive Summary

---

FAIR-IMPACT aims to support the implementation phase of the European Open Science Cloud. To this end, FAIR-IMPACT has a focus on the EOSC Interoperability Framework. The perspective for this deliverable is on legal interoperability for core metadata schemes for research datasets which refers to the ability to share, access, and use research data across different legal jurisdictions and institutions while complying with relevant laws, regulations, and policies.

At first, the deliverable illustrates the impact of the GDPR on legal interoperability in practice. The deliverable then presents as a landscape analysis the foundation for a legal interoperability framework, by evaluating the relevance of existing widely adopted metadata schemas and controlled vocabularies used in data repositories for the description of legal constraints. Their use and implementation are illustrated with specific use cases, which also validate the findings by highlighting the challenges faced and addressed by different communities.

As a core metadata schema to foster legal interoperability between research datasets the DCAT-standard stands out as the best-suited tool to describe datasets, facilitates interoperability between data catalogues, portals, and repositories, and allows for the enrichment of metadata in the analysis. Additional controlled vocabularies support aspects of legal interoperability: international licences (AT-licence), access control policies (ODRL), copyright and intellectual property (DCMI, PROV-O), data protection and privacy (DPV).

Few mappings between metadata schemas are available, they could be improved with new ones in addition to the ones considered here. The FAIRCore4EOSC has deployed the MSCRregistry, which, thanks to its ease of access, could enhance the uptake of such mappings effectively.

The recommendation of the DCAT standard as building block for legal interoperability supported by additional controlled vocabularies could also be explored by the new, to-come EOSC Interoperability Task Force, and support programs could be put together to stimulate the adoption of good practices in the domain.



## 2. Introduction

---

FAIR-IMPACT has a focus on the EOSC Interoperability Framework with the four layers - technical, semantic, organisational, and legal interoperability. The perspective for this deliverable is on legal interoperability for core metadata schemes for research datasets.

In the perspective of the FAIR Principles the legal interoperability of research datasets refers to the ability to access research data across different legal jurisdictions and institutions while complying with relevant laws, regulations, and policies. It implies ensuring that legal barriers and restrictions do not impede the collaboration and exchange of research data between different entities, such as researchers, institutions, and research performing organisations.

This poses several key challenges, including the proper management of data protection and privacy regulations, intellectual property rights, contractual and collaboration agreements, and ethical considerations and consent.

The ambition of this task in FAIR-IMPACT is high and formulated as “From the data layer to the organisational layer, this task will provide a complete legal framework facilitating technical interoperability.”<sup>1</sup>

The challenges are manifold: legal regulations are dynamic - they develop over time. They are implemented hierarchically (usually from higher governance bodies to concrete organisations on the ground). Different countries and regions have varying data governance and/or legal frameworks governing research data. Moreover, their ‘timeline’ of implementation is usually not synchronised in time. The legal frameworks are inherently complex. They have different stakeholders. They usually negotiate different interests, and in consequence, they may include specific requirements, restrictions, or exceptions that hinder seamless data sharing and interoperability. Identifying and comparing these frameworks to establish compatible agreements and to identify those elements, which are core for legal interoperability is crucial but challenging.

The "legal frameworks" themselves do not change very much over time. In case law clarifications are given with regards to e.g. the legal scope of GDPR terminology. The real issue is the translation of the GDPR principles and 'open norms' in sector specific guidance and, for instance, a common view on risk identification, risk assessment and risk mitigation.

This deliverable aims to create the foundation for a legal interoperability framework by providing an evaluation of the relevance of existing widely adopted metadata schemas and controlled vocabularies used in data repositories for the description of legal constraints. In particular, we analysed the schemas and vocabularies listed in [Section 4](#).

Specific use cases were then chosen to illustrate their use and implementation and validate the findings of the previous phase. The core of this deliverable is a proposal for a core set of metadata or elements for a general legal interoperability framework.

---

<sup>1</sup> FAIR-IMPACT Description of Action (DOA), internal document

In lieu of a motivation, this report starts in the next chapter with an illustration of the impact of the GDPR<sup>2</sup> on repositories and the reuse of data looking at data services provided by DANS-KNAW.

---

<sup>2</sup> <https://gdpr-info.eu/>



### 3. Illustration of the complexity of the legal interoperability landscape

---

There is an obvious impact of the GDPR on research data infrastructures (including repositories) across Europe as mirrored by recommendations for ‘GDPR-compliance for personal data’ and related to the recommendation of alignment between Member States national legislations and EOSC.

In a recent dissertation, Alisson Tyler (Tyler 2022) analysed this impact, executing four case studies of data repositories in the social sciences. Those are the Czech Social Science Data Archive<sup>3</sup>, the Finnish Social Science Data Archive<sup>4</sup>, data services of the Data Archiving and Networked Services<sup>5</sup>, an institute of the Royal Netherlands Academy of Arts and Sciences and the Dutch Research Organisation, and GESIS<sup>6</sup>, a German Leibniz Institute for the Social Sciences. Tyler started her research from the observation that while the GDPR and related national legislation aimed for a better protection of personal data of individuals, the implication of such regulations for the science system, and especially for those research communities, which work with personal data, was less thought through. In particular, she investigated *“how the changes required to comply with the GDPR have affected the data archives’ abilities to operate”*<sup>7</sup> Tyler’s description highlights the complexity of the legal frameworks under which data repositories work. Her study also reveals that legal frameworks are not static: the frameworks are not static in the sense that they themselves change. Their implementation is often taking place as a slow(er) process and pace across time and various institutional layers, too. Moreover, those institutional layers follow their own dynamics, and one can observe both an independence among and an interdependence between layers.

In the case of EASY<sup>8</sup>, a former data service of DANS-KNAW the process of the GDPR implementation lasted several years. Legally, DANS is an institute of the Royal Netherlands Academy for Arts and Sciences (KNAW). Thus, the way the KNAW absorbed the GDPR after its acceptance by the Dutch national legislation, established the boundary conditions for DANS to adopt its own policies. The implementation process should not be imagined linear and straightforward: At each stage, translations (in the sense of adopting certain rules into new kinds of systems) took place and were accompanied by negotiations.

Based on such concrete investigations, we attempt to summarise the various aspects of such a process. In the figure below we show a schematic overview of the impact of the GDPR on repositories, their depositors and end users, which shows how this EU legislation leads to different national practices and does not directly have legal interoperability as effect.

---

<sup>3</sup> <https://www.soc.cas.cz/en/departement/czech-social-science-data-archive>

<sup>4</sup> <https://www.fsd.tuni.fi/en/>

<sup>5</sup> <http://www.dans.knaw.nl/>

<sup>6</sup> <https://www.gesis.org/en/home>

<sup>7</sup> Quotation from the PhD thesis abstract see here: <https://deepblue.lib.umich.edu/handle/2027.42/175592>

<sup>8</sup> <https://dans.knaw.nl/en/data-services/easy/> - a research data repository for self-archiving of datasets

## Legal interoperability for data repositories?

### The case of the GDPR

<b>EU and national legal arrangements</b>  In 2018 the General Data Protection Regulation (GDPR) was implemented in the EU. This included EU wide legislation and national implementation laws.	<b>Phase 1</b>
<b>Phase 2</b>	<b>Depositors have to act in line with the GDPR</b>  As a result, depositors have to align their data processing with the GDPR and decide on policy matters in relation to research. <ul style="list-style-type: none"> <li>• GDPR compliance</li> <li>• Setting up policies</li> <li>• Contractual arrangements with repositories</li> </ul>
	<b>Repositories have to act in line with the GDPR</b>  Repositories have to make sure they are GDPR compliant, working on the technical set up as well as the organizational aspects. <ul style="list-style-type: none"> <li>• GDPR compliance</li> <li>• Setting up policies</li> <li>• Contractual arrangements with depositors</li> <li>• Explanation on contractual arrangements and use of the service</li> </ul>
<b>End users have to act in line with the GDPR</b>  Reuse of data requires careful assessment and further processing under GDPR terms. <ul style="list-style-type: none"> <li>• Contractual arrangements with depositor or repository (whoever is controller)</li> <li>• GDPR compliance</li> </ul>	<b>Phase 3</b>

### The GDPR Impact - challenges for interoperability

Time and resources:

Member states are not all at the same pace with implementation. Depending on the type of repository it may require significant effort to achieve GDPR compliance.

Differences in member state implementation law:	What is allowed in one member state may not be allowed in another.
Differences in policies nationally and internationally:	The specification on how data can be processed and shared is not everywhere the same and may even differ nationally between research domains.
Serving an international public/end users:	Due to the variety of where data originates, and under what terms reuse is possible, repositories need an effective way of communicating the reuse possibilities to end users.

Whenever it comes to the discussions about further steps towards legal interoperability (see for instance “Which legal aspects should be considered when creating, providing or using novel data-driven solutions in data spaces?”<sup>9</sup>) existing experiences in actual data archiving practices should be taken into account when considering how the recommendations for legal interoperability can be implemented.

<sup>9</sup> Farrell, E., Minghini, M., Kotsev, A., Soler Garrido, J., Tapsall, B., Micheli, M., Posada Sanchez, M., Signorelli, S., Tartaro, A., Bernal Cereceda, J., Vespe, M., Di Leo, M., Carballa Smichowski, B., Smith, R., Schade, S., Pogorzelska, K., Gabrielli, L. and De Marchi, D., European Data Spaces - Scientific Insights into Data Sharing and Utilisation at Scale, EUR 31499 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-76-53522-5, <https://data.europa.eu/doi/10.2760/400188>, JRC129900.

## 4. Schemas and vocabularies under consideration

This initial phase toward the recommendation of a framework for legal interoperability is the evaluation of relevant metadata schemas or controlled vocabularies for this purpose with the objective to identify the most appropriate ones. A list of widely-used schemas and vocabularies has been identified, and a set of criteria has been put together to evaluate them.

**Table 1 – List of schemas and vocabularies evaluated**

Name	Type
Dublin Core Metadata Element Set/Terms	Schema
Datacite Metadata Schema	Schema
Rioxx	Schema
Asset Description Metadata Schema	Schema
Open Digital Rights Language	Schema + vocabulary
Data Privacy Vocabulary	Vocabulary
FAIRSharing Policy Schema	Schema
Data Catalog	Schema
OpenAIRE Guidelines	Schema
Data Documentation Initiative	Schema
RO-Crate	Container

### 4.1 Evaluation criteria

The EOSC Interoperability Framework<sup>10</sup> recommends that a minimum metadata model should be proposed to ease discovery over existing federated research data and metadata to facilitate findability and to support interoperability not only within domain-specific services or repositories, but also across domains and communities.

EOSC IF recommends a minimum metadata set to describe metadata records for FAIR digital objects where it is possible to use various existing metadata schemes and aggregators' guidelines. This applies for any rights and terms of access information for this digital object. Typically, rights information includes a statement about various property rights associated with the digital object, including intellectual property rights. The EOSC IF recommends to refer to a rights statement with a URI, or a literal value (name, label, or short text) if the former is not possible or feasible.

These metadata elements are essential for the reusability of the digital object and consist of the:

- Standardised version of the licence name, URI of the licence;
- Copyrights holder (use authority control databases for persons or institutions if available and persistent identifiers - e.g., ORCID, ISNI, VIAFID - or national authority database of personal and corporate names);
- Access rights;

<sup>10</sup>

<https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1/language-en>

- Confidentiality declaration;
- Special permissions, restrictions, conditions, disclaimers;
- Citation requirements;
- Level of access, access type, authentication, authorisation, access method, and granularity.

The schemas and vocabularies in the next sections will be evaluated against these recommendations, i.e. their ability to convey this information for scientific datasets.

In order to ease the process, these elements have been gathered in three categories:

- Access policies (access rights, level of access, access type, access method, etc.);
- Copyright and intellectual property (copyrights holder, licence, citation requirements, etc.);
- Data protection and privacy (confidentiality declaration, special permissions, etc.).

In addition, other non-technical criteria will be considered during the evaluation, such as:

- The level of adoption and the existence of a community of practice,
- The provision of support,
- The availability of tools.

## 4.2 Dublin Core Metadata Element Set & Terms

The DCMES (table 2) is a simple and widely used framework which originated in 1995 as an effort to create a simple and standardised way of describing digital resources to facilitate resource discovery and metadata sharing. The initial set of 13 metadata elements was quickly expanded to 15 as the DCMI was officially formed to continue the development of the Dublin Core. These elements are meant to provide a basic level of description for various types of digital resources, making them more discoverable and understandable. Additionally, qualifiers, refinements and terms can be used to provide more context and specificity when necessary.

**Table 2 – DC Metadata Element Set & Terms fact sheet**

Name	Dublin Core Metadata Element Set & Terms
Author	DCMI
Purpose	Representation of usage conditions of digital assets
Type	Schema
Current version	1.1 (Schema) 14 (Terms)
Issue date	20/08/2020
References	
Formalism	XML, RDF
URL	<a href="https://www.dublincore.org/specifications/dublin-core/">https://www.dublincore.org/specifications/dublin-core/</a>

Over 25 years, the DCMES schema has evolved to accommodate changing needs, emerging technologies, and the increasing complexity of digital resources. It has become a foundational framework for metadata description and resource discovery across various domains, and its ongoing development reflects the commitment to ensuring its relevance in an ever-evolving information landscape.

## ***Access Policies***

The DCMES schema does not include a specific element designed explicitly for describing access policies. However, existing qualifiers (e.g. the “accessRights” sub-property of the “Rights” element) can be used to convey this information indirectly.

## ***Copyright and Intellectual Property***

The DCMI terms include the “Rights” element, which allows to convey information about copyright and rights-related information as well as the legal and intellectual property rights associated with the resource. It can be used to provide additional information such as the type of licence, usage permissions, and any relevant terms and conditions. If more details are required, qualifiers, controlled vocabularies, or an extension of the DCMI Terms vocabulary (e.g., with the Open Data Rights Language, which defines properties for specifying, among others, copyright statements and copyright notices – see below) can be used to accommodate specific needs, the latter being a more sophisticated approach.

## ***Data Protection and Privacy***

The DCMI terms do not include a specific element designed explicitly for describing data protection and privacy information. Similarly to the access policies, existing elements and qualifiers such as the “Rights” element can be used to convey information related to data protection and privacy indirectly, even if it is more broadly used to describe legal and intellectual property rights. To convey more detailed data protection and privacy information, extended vocabularies, controlled vocabularies, or annotations may be relevant.

## ***Conclusion***

The DCMES strengths lie in its simplicity, broad applicability, making it suitable for basic descriptive needs and facilitating collaboration. However, the small set of elements implies lack of richness, limited specificity, and ambiguity due to semantic imprecision: as an example, only one element and a property may be relevant to describe information from the three types above. This can make it less suitable for describing complex or specialised resources such as datasets.

Other schemas (e.g. Datacite, Rioxx, DCAT - see below) makes extensive use of DCMI terms to which they sometimes associate controlled vocabularies (e.g., ADMS - see below) to address this ambiguity problem and enhance interoperability.

### ***4.3 DataCite Metadata Schema***

The DataCite Metadata Schema (table 3) is designed to provide a standardised way to describe research data, making it easier to discover and cite datasets. Since its founding in 2009, DataCite has grown and now spans the globe from Europe and North America to Asia and Australia. The aim of DataCite is to provide domain-agnostic services to benefit scholars in a wide range of disciplines.

Key to the DataCite service is the concept of a long-term or persistent identifier, i.e., an association between a character string and a resource. Resources can be files, parts of files, persons, organisations, abstractions, etc. DataCite uses Digital Object Identifiers (DOIs).

The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. The resource that is being identified can be of any kind, but it is typically a dataset.

**Table 3 – Datacite Metadata Schema fact sheet**

Name	Datacite Metadata Schema
Author	Datacite
Purpose	Documentation for publication and citation of research data and other research outputs
Type	Schema
Current version	4.4
Issue date	30/03/2021
References	Dublin Core
Formalism	XML
URL	<a href="https://schema.datacite.org/meta/kernel-4.4/">https://schema.datacite.org/meta/kernel-4.4/</a>

Some correspondences exist with the Dublin Core Metadata. Some DataCite definitions have borrowed liberally from the DCMI ones - see below.

### ***Access Policies***

Like DCMI, the Datacite schema does not include a specific element designed explicitly for describing access policies. However, the existing “Rights” element can be used to convey this information indirectly, e.g., including a statement for access rights for the resource or referencing a service providing such information. This can include embargo information if applicable, but the schema does not impose specific vocabularies to describe them.

### ***Copyright and Intellectual Property***

The Datacite Metadata Terms include the “Rights” element, which allows to convey information about copyright and rights-related information as well as the legal and intellectual property rights associated with the resource. It can be used to provide additional information such as the type of licence, usage permissions, and any relevant terms and conditions. The element is complemented by sub-properties such as “rightsURI”, “rightsIdentifier”, or controlled values such as “RightsHolder” which allow to include standardised values for licence types and names, or property rights.

### ***Data Protection and Privacy***

There is somehow a tacit idea in Datacite Metadata Terms that IPR are the only rights that may exist in a dataset, so they do not include a specific element designed explicitly for describing data protection and privacy information either. Similarly to the access policies, existing elements and qualifiers such as the “Rights” element can be used to convey information related to data protection and privacy indirectly, even if it is more broadly used to describe legal and intellectual property rights.

## Conclusion

While there is some overlap in metadata elements between DataCite and Dublin Core, DataCite is specifically tailored for research data and datasets, while Dublin Core is a more general-purpose metadata standard that can be applied to various types of resources, including text-based documents, images, and multimedia. As a result, DataCite includes additional elements and extensions that are relevant to research data management.

### 4.4 Research Outputs Metadata Schema

The RIOXX (table 4) was developed for institutional repositories to share metadata about the scholarly resources they contain. Originally designed to meet the reporting requirements of the now defunct Research Councils UK<sup>11</sup>, Rioxx has also proven to be generally useful as a standard for sharing metadata between repositories and network services such as large-scale metadata aggregators (e.g. Core<sup>12</sup>).

The Rioxx metadata application profile focuses on applying consistency to the metadata fields used for identifiers of scholarly outputs, people and organisations, research funders and project/grants, and is designed to support the consistent tracking of open-access research publications across scholarly systems. It was developed for repositories to share metadata about the scholarly resources they contain. It has been deployed as a metadata application profile in approximately 70 institutional repositories in the UK, and has software implementations in DSpace, ePrints and Pure.

In 2019 the United Kingdom Council of Research Repositories<sup>13</sup> formed a Governance Group to oversee revisions and development.

**Table 4 – RIOXX fact sheet**

Name	Research Outputs Metadata Schema
Author	RCUK / UKCORR
Purpose	Share metadata between repositories and network services such as large-scale metadata aggregators
Type	Schema
Current version	3.0
Issue date	23/06/2022
References	Dublin Core
Formalism	XML
URL	<a href="https://www.riox.net/profiles/v3-0-rc-1/">https://www.riox.net/profiles/v3-0-rc-1/</a>

The Rioxx metadata schema is primarily designed for the purpose of interoperability in the context of academic research outputs, specifically for repositories and their content, but may not have a direct application to legal constraints.

## Access Policies

<sup>11</sup> <http://www.rcuk.ac.uk>

<sup>12</sup> <http://www.core.ac.uk>

<sup>13</sup> <https://www.ukcorr.org/>



The Rioxx schema no longer includes elements about the type of access (e.g., open access, restricted access) or associated embargo periods, as the “Rights” element has been removed in its latest version.

### ***Copyright and Intellectual Property***

The Rioxx schema includes fields (e.g., “license\_ref”) to capture information about copyright holders, licences, and embargo periods associated with research outputs.

### ***Data Protection and Privacy***

Likewise access policies, the Rioxx schema does not include fields to record information about data protection any longer.

### ***Conclusion***

While the Rioxx metadata schema itself may not directly address legal aspects, it provides a lightweight structured framework for capturing and managing information about research outputs.

Initially developed in a UK-specific context, Rioxx is evolving with a global perspective in its version 3. The changes have mostly been introduced to support some general use-cases, such as facilitating automated discovery of resources described by the metadata records (through explicit linking and semantic typing of resources using schema.org). This is essential for making repositories more machine-friendly and thus supporting basic use cases of harvesters, aggregators and other down-stream services, and reporting on events and characteristics relating to the open-access status of resources, primarily of interest to funders. This may include inspection of associations between resources and their related or underlying research entities, such as grants and/or projects, both of which are semantically differentiated within Rioxx v3.

The Rioxx 2.0 has been mapped to OpenAIRE Guidelines<sup>14</sup> for Literature Repository Managers 3.0 - see below. Rioxx is "lighter" than OpenAIRE, presenting a flatter expression of repository resources.

## ***4.5 Asset Description Metadata Schema***

The ADMS (table 5) is a vocabulary to describe so-called “semantic assets”, that is, highly reusable metadata and reference data that are used for eGovernment system development. It enables ICT developers to explore and search for interoperability assets such as controlled vocabularies (e.g., code lists, taxonomies, dictionaries, vocabularies) and metadata schemas (e.g., XML schemata, generic data models).

**Table 5 – ADMS fact sheet**

Name	Asset Description Metadata Schema
Author	W3C
Purpose	Profile for describing semantic assets

<sup>14</sup> [https://www.riox.net/mappings/crosswalk\\_riox\\_2\\_0\\_openaire\\_3\\_0/](https://www.riox.net/mappings/crosswalk_riox_2_0_openaire_3_0/)

Type	Vocabulary
Current version	2.0
Issue date	2023
References	DCAT, Dublin Core, SKOS, FOAF
Formalism	XML, RDF
URL	<a href="https://semiceu.github.io/ADMS/releases/2.00/">https://semiceu.github.io/ADMS/releases/2.00/</a>

ADMS includes solutions covering the political, legal, organisational and technical interoperability layers as defined by the EIF<sup>15</sup>. It originated from the ISA2 programme<sup>16</sup> supported by the EC, and the first draft versions were released in 2011. In 2013, an ADMS working group was created within the W3C and a note published. At that time, it was tightly coupled to DCAT (see below) for which it was developed as an AP with slightly different objectives: DCAT is designed to facilitate interoperability between data catalogues, and the catalogue itself is at the heart of the vocabulary. ADMS is rather focused on the assets within a catalogue.

### Access Policies

With the version published in 2024, a significant number of secondary concepts have been removed (e.g. licence) of ADMS. However Dublin Core or DCAT properties can be integrated to reflect the licence type.

### Copyright and Intellectual Property

There is no explicit mention of this in ADMS, but again Dublin Core or DCAT properties can be used to describe the rights associated with a dataset.

### Data Protection and Privacy

This is not covered by the vocabulary.

### Conclusion

ADMS is endorsed by the EC and its key implementations are Interoperable Europe<sup>17</sup>. However, it is specifically focused on the description of “interoperability assets” such as metadata schemas or vocabularies - a subset of what is envisioned with datasets. There has been a break in the support of ADMS between 2016 and 2023, but it is now supported by SEMIC<sup>18</sup>, an eGovernment service initiated by the EC, to foster community uptake. However, it is too specific to the description of semantic assets and relies on other schemas that will be more adapted for datasets management.

## 4.6 Open Digital Rights Language

ODRL (table 6) is a language promoted by the ODRL Community Group in order to model policies for digital content and media. To do so, ODRL offers a Core Vocabulary to specify the

<sup>15</sup> [https://ec.europa.eu/isa2/sites/default/files/eif\\_brochure\\_final.pdf](https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf)

<sup>16</sup> [https://ec.europa.eu/isa2/home\\_en/](https://ec.europa.eu/isa2/home_en/)

<sup>17</sup> [https://ec.europa.eu/isa2/news/new-level-cooperation-isa-building-interoperable-europe\\_en/](https://ec.europa.eu/isa2/news/new-level-cooperation-isa-building-interoperable-europe_en/)

<sup>18</sup> <https://joinup.ec.europa.eu/collection/semic-support-centre>

minimum set of terms suitable to model the policies and a Common Vocabulary of general terms to model, for example, actions regulated by the obligations, permission and prohibitions expressed in the policies.

**Table 6 – ODRL fact sheet**

Name	Open Digital Rights Language
Author	W3C
Purpose	Representation of usage conditions of digital assets
Type	Schema, vocabulary
Current version	2.2
Issue date	15/02/2018
References	Dublin Core, SKOS, FOAF
Formalism	RDF
URL	<a href="https://www.w3.org/TR/odrl-model/">https://www.w3.org/TR/odrl-model/</a>

It models different types of policies, making a distinction between

- a policy which is an agreement between an assignor and an assignee,
- a policy which is an offer from an assigner to an undefined wide audience and
- a policy which is a generic set of rules with no specified assignor and assignee.

Concerning the deontic logic, ODRL allows the expression of the effects associated with the non-compliance of an obligation, the effects of the noncompliance of some preliminary duties to obtain a permission and the duties to be accomplished for remedying a violated prohibition. Finally, it is possible to associate a policy with some meta-information concerning, for example, its creator, its coverage (i.e., the jurisdiction applied upon the policy) and the reference to older versions of the policy.

### ***Access Policies***

The ODRL model has a specific class to model policies. Users can establish a set of rules based on Permissions, Duties and Prohibitions. These concepts are related to the concept of Action. For example, with Permissions, you can execute an Action over an Asset, while Prohibitions can't execute an Action over an Asset. Examples of actions can be use or transfer.

### ***Copyright and Intellectual Property***

There is no specific class to implement copyright and properties. Nevertheless, these concepts can be controlled by combining the Party and Action concepts. A party is an entity (or entities) such as a person, a collection of people and organisms. In this case, you can control the access although you can not define who is the owner of the property. In the case of the concept Action, indicates an operation on an Asset.

Copyright exploitation rights respond to the right holder's legal capacity to define which activities (actions) may be exercised over a work. These actions are to copy, to make derivative copies (transformation), to distribute (tangible objects) and to communicate to the public (intangible objects).

Therefore, combination of Party and Action, Permission, Duty and Prohibition classes would provide the possibility to define copyright right holders and uses.

The benefits of utilising an ODRL representation include the facilitation of summarization, where key actions permitted or prohibited are succinctly captured within a policy, promoting efficiency in policy management. Moreover, ODRL enables streamlined compliance checking, allowing businesses to assess the compliance status of their processes promptly, enhancing regulatory adherence and operational effectiveness.

## Data Protection and Privacy

Through the Rule concept, we can define the entity that can access the resource. It is an indirect way to express the privacy of the property. But you can not define the licence of the property, only who can access or the actions you can realise based on your profile.

## Conclusion

This language offers a powerful mechanism to control the access to the resources, as well to express the actions that can be performed. Also, it is very flexible, adapting to different scenarios.

## 4.7 Data Privacy Vocabulary

The DPV (table 7) is a vocabulary and an ontology serialised using semantic-web standards to represent concepts associated with privacy and data protection, primarily derived from GDPR. DPV enables expressing machine-readable metadata about the use and processing of personal data based on legislative requirements.

**Table 7 – DPV fact sheet**

Name	Data Privacy Vocabulary
Author	W3C Data Privacy Vocabularies and Controls CG (DPVCG)
Purpose	To assist in the representation of information concerning privacy in the context of personal data processing by acting as a core framework of ‘common concepts’ that can be extended to represent specific laws, domains, or applications.
Type	Vocabulary, ontology
Current version	1.0
Issue date	05/12/2022
References	OWL, SKOS
Formalism	RDF
URL	<a href="https://w3id.org/dpv">https://w3id.org/dpv</a>

## Access Policies

As the Data Privacy Vocabulary is a representation of concepts associated with privacy and data protection it does not include any concepts or relations pertaining to access policies.

## Copyright and Intellectual Property

As the Data Privacy Vocabulary is a representation of concepts associated with privacy and data protection, concepts and/or relations exclusively pertaining to copyright or intellectual property are not covered.

## ***Data Protection and Privacy***

The DPV is explicitly designed to represent concepts associated with privacy and data protection, primarily the GDPR. As such, it enables representation of which personal data categories are undergoing what kind of processing by a specific data controller and/or transferred to some recipient for a particular purpose, based on a specific legal basis with specified technical and organisational measures and restrictions in place. The scope of the DPV is limited to personal data and other types of data cannot be represented using DPV as of the current version. DPV is, however, designed to describe policy classes at a top-level and the vocabulary is intended to be extendable. As such, future extensions of the vocabulary may allow descriptions of classes and concepts belonging to other types of data.

## ***Conclusion***

Due to its highly specialised nature, DPV has a limited, but nevertheless significant, potential for supporting legal interoperability, specifically in relation to privacy and personal data protection. As such, the relevance of DPV will be limited to the context of use cases, research domains and datasets processing personal data. Additionally, being a vocabulary, implementation of DPV will also be dependent on a metadata schema able to sufficiently encompass and describe the concepts contained in the DPV.

## ***4.8 FAIRSharing Policy Schema***

FAIRsharing<sup>19</sup> (table 8) is a community-driven resource with users and collaborators across all disciplines. FAIRsharing works with a range of stakeholders to enable the FAIR Principles by promoting the value and the use of standards, databases and policies. FAIRsharing provides a data policy registry that stores descriptions of data policies from a number of different sources including journal publishers, institutions and funders. In 2022, FAIRsharing extended their updated FAIRsharing data model to reflect the recommended fields suggested by FAIR-enabling Data Policy Checklist<sup>20</sup> and RDA's Journal Policy Features<sup>21</sup> to make more detailed and structured policy descriptions accessible to both humans and machines. FAIRsharing also enables relationships to be made between policy records and relevant standards, repositories and databases that are also described within FAIRsharing. Registered data policies can also be nested, allowing one data policy to extend another data policy(ies).

**Table 8 – FAIRSharing Policy Schema fact sheet**

Name	FAIRsharing Policy Schema
------	---------------------------

<sup>19</sup> <https://fairsharing.org/>

<sup>20</sup> Davidson, J., Grootveld, M., Verburg, M., van Horik, R., O'Connor, R., Engelhardt, C., Garbuglia, F., Vieira, A., Newbold, E., Proudman, V., & Horton, L. (2022). FAIR-enabling Data Policy Checklist (1.0). <https://zenodo.org/doi/10.5281/zenodo.6225774>

<sup>21</sup> Hrynaskiewicz, I, et al. 2020. Developing a Research Data Policy Framework for All Journals and Publishers. Data Science Journal, 19: 5, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2020-005>

Author	FAIRsharing and Digital Curation Centre
Purpose	Make detailed and structured research data policy descriptions accessible to both humans and machines
Type	Schema
Current version	
Issue date	2022
References	FAIRsharing released a blog post <sup>22</sup> to describe the updated policy metadata schema
Formalism	
URL	<a href="https://api.fairsharing.org/model/policy_schema.json">https://api.fairsharing.org/model/policy_schema.json</a>

## ***Access Policies***

The FAIRsharing schema has three properties relating to accessibility of research outputs. These focus on the sharing of research data (sharing\_data), metadata (sharing\_metadata), and research software (sharing\_research\_software). For each, there are predefined options to support machine readability and to make clear whether the property is required, suggested, not covered by the policy. 'Required' is used when the policy clearly states that sharing of the output (data, metadata or research software) is required and provides clarity on legitimate exceptions. 'Suggested' is used when sharing research software is only encouraged. 'Not Covered' is selected when the policy addresses this topic poorly or not at all. Otherwise, 'Other' is used. There is an opportunity to provide a free text note for each of these properties to support human readability.

There is also a property to describe whether the policy requires the provision of a data availability statement. 'Yes' is applicable when the policy makes clear that a Data Availability Statement is required. When the policy only encourages such a statement, or does not address this topic, 'no' is used.

## ***Copyright and Intellectual Property***

The schema includes a field (licenses\_for\_outputs) to make clear whether the policy has preferred licence types for research outputs covered by the policy. There is also a property to describe data citation (data\_citation) to make clear whether the policy has expectations around the citation of a broad range of research outputs and actors including data, software, researchers and roles.

## ***Data Protection and Privacy***

The schema includes a field relating to whether the policy has a requirement relating to data protection (data\_protection). This is a binary option (Yes/No) to support machine readability and there is an opportunity to provide a free text note. There is also a property to describe whether exceptions to data sharing (exceptions\_to\_data\_sharing) are permitted within the context of the policy. 'Yes' is applicable when the policy allows exceptions to the stated data sharing methods, otherwise 'no' is used.

## ***Conclusion***

---

<sup>22</sup> FAIRsharing and DCC collaborate to align policy metadata <https://blog.fairsharing.org/?p=451> (11 November 2022)

The FAIRsharing policy metadata has potential to support interoperability of organisational data policies specifically in relation to research data. FAIRsharing enables the creation of collections meaning that, in theory, national level collections could be created to include the policies of RPOs, funders, data service providers and publishers. Such an approach could support a more automated means of monitoring of the policy landscape across EOSC Association members in comparison with the annual survey<sup>23</sup> that is currently issued. FAIRsharing policy metadata has potential to support machine actionable DMPs by making the policy content of different actors machine readable and comparable. While there is good potential, there remains work to be done to increase the number of RPO and funders' policies registered within FAIRsharing which is currently more populated by publishers' policies. To this end, FAIR-IMPACT WP2 is currently considering a support offer to run a number of policy sprints with FAIRsharing to increase the number of policies in the registry.

## 4.9 Data Catalog

The DCAT (table 9) specification, or Data Catalog vocabulary was initially developed by the W3C as a way to describe datasets in data catalogues. It enables a publisher to describe datasets and data services in a catalogue using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogues. This can increase the discoverability of datasets and data services.

**Table 9 – DCAT fact sheet**

Name	Data Catalog
Author	W3C
Purpose	Facilitate interoperability between data catalogues published on the Web
Type	Schema
Current version	3.0
Issue date	14/01/2024
References	Dublin Core, FOAF, ADMS
Formalism	RDF
URL	<a href="https://www.w3.org/TR/vocab-dcat-3/">https://www.w3.org/TR/vocab-dcat-3/</a>

DCAT makes a distinction between a dataset, which represents a collection of data resources associated with a specific project, topic, or theme, and data Distribution, which represents a specific downloadable or accessible version of a dataset. A dataset can have multiple data distributions, each representing a different format, version, licence or medium for accessing the data.

## Access Policies

The “accessRights” field - which is inherited from the Dublin Core Terms - indicates the level of access allowed for the dataset. It can specify whether the dataset is public, restricted, or other access levels. Values must be picked in the EU Access Rights Named Authority List<sup>24</sup>.

<sup>23</sup> <https://eosc.eu/wp-content/uploads/2023/08/Monitoring-Framework.pdf>

<sup>24</sup> <https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/access-right>



## ***Copyright and Intellectual Property***

The “licenseDocument” field - another input from the Dublin Core Terms - allows for linking to a document that contains the full text of the licence associated with the dataset. It provides users with direct access to the legal terms and conditions. It includes the “type” property that is used to provide information about the licence associated with the dataset. ADMS licence type vocabulary<sup>25</sup> is mandatory to describe a reference to the licence type in a controlled way. More terms inherited from Dublin Core, such as “creator” (IP gives original rights to the creator of a work, thus it is the first rightholder), “publisher” (publishers exercise public communication rights of the resource) can also be relevant.

## ***Data Protection and Privacy***

The “provenance” field provides information about the origin or source of the dataset. This can include legal information related to the data's origin, which can be important for establishing data ownership and compliance. Some implementations of DCAT might use the “legalBasis” field to provide information about the legal basis or authority under which the dataset is published. This can help users understand the legal context in which the dataset is made available. These two fields are Dublin Core Terms, but DCAT doesn’t require the use of particular controlled vocabularies to populate them.

## ***Conclusion***

DCAT gained traction not only in Europe but also globally as many countries and organisations derived the vocabulary in AP specifications to adapt it to suit their local contexts<sup>26</sup>. Its focus on semantic web technologies and structured data made it suitable for promoting data interoperability and sharing across diverse ecosystems.

DCAT and the associated DCAT-APs continue to evolve, with newer versions addressing changes in data practices, technologies, and user needs. The development process often involved collaboration among governments, organisations, and experts to ensure that the profile remained relevant and effective.

### ***4.10 OpenAIRE Guidelines***

The OpenAIRE Guidelines (table 10) for Data Archives have been initiated in 2012, in order to provide recommendations and an application profile for data archive and repository managers to ensure that research data are properly managed and described, and include metadata compatible with the OpenAIRE infrastructure. They are published along with other guidelines: literature, institutional, and thematic repositories, software repositories, CRIS, and other research products.

OpenAIRE has adopted the DataCite Metadata Schema as the basis for harvesting and importing metadata about datasets from data archives which will be exposed when related

<sup>25</sup> <http://purl.org/adms/licencetype/1.0>

<sup>26</sup>

<https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe/about>



to an open access publication e.g. a dataset cited by an article. It builds on its version 3.1 by making some of the otherwise optional DataCite properties mandatory, as well as enforcing specific encoding schemes on the values of some DataCite properties.

**Table 10 – OpenAIRE guidelines fact sheet**

Name	OpenAIRE guidelines
Author	OpenAIRE
Purpose	Expose archives metadata in a way that is compatible with the OpenAIRE infrastructure
Type	Schema
Current version	2.0
Issue date	01/04/2014
References	DataCite
Formalism	XML
URL	<a href="https://guidelines.openaire.eu/en/latest/data/application_profile.html">https://guidelines.openaire.eu/en/latest/data/application_profile.html</a>

## Access Policies

OpenAIRE uses the access rights to enable a better user experience by declaring the access rights clear and explicit in the portal. Access rights are specified using the rightsURI attribute of the Rights property, and referring to the info:eu-repo-Access-Terms vocabulary<sup>27</sup>.

## Copyright and Intellectual Property

Licences can be mentioned in the Rights property as well. This is a free-text field, but the rightsURI attributes can further indicate this information using standard values which can be machine-actionable. Creator, Publisher and Contributor items may also be useful for legal interoperability when it comes to Copyright holder metadata.

## Data Protection and Privacy

There are no specific recommendations, but again the Rights element could be used here.

## Conclusion

The last update of OpenAIRE Guidelines for Data Archives dates back to 2014. The Guidelines for Literature Repositories - which have been updated more frequently - include more properties (e.g., dateEmbargo, accessrights, licenceCondition) associated with controlled vocabularies (e.g. COAR access right vocabulary<sup>28</sup>) which would be relevant as well for datasets, but are essentially intended for open access and non-open access publications. The latter constitute the entry or gateway to EOSC as repositories which are compatible with the versions 3.0 or 4.0 of the OpenAIRE Guidelines will be onboarding the EOSC Portal Catalogue and Marketplace integrated platform. In 2022, OpenAIRE put together a working group to ensure a community-driven global governance for the guidelines, which announced<sup>29</sup> a year later the simplification and merging of all the

<sup>27</sup> <http://purl.org/eu-repo/semantics/#info-eu-repo-AccessRights>

<sup>28</sup> [https://vocabularies.coar-repositories.org/documentation/access\\_rights/](https://vocabularies.coar-repositories.org/documentation/access_rights/)

<sup>29</sup> <https://zenodo.org/records/8094906>

guidelines except the CRIS one, to form the OpenAIRE Interoperability Guidelines for Repositories.

#### 4.11 Data Documentation Initiative

The DDI (table 11) is an International standard for describing data from the social, behavioural, and economic sciences. Two versions of the standard are currently maintained in parallel:

- DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving.
- DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible.

**Table 11 – DDI Codebook fact sheet**

Name	<b>DDI Codebook</b>
Author	DDI Alliance
Purpose	Light-weight version of the DDI standard, intended primarily to document simple survey data
Type	Schema
Current version	v2.5
Issue date	15/1/2014
References	Dublin Core
Formalism	XML
URL	<a href="https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html">https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html</a>

DDI is a very detailed standard developed in line with the needs of data archives and producers in the social, behavioural, and economics sciences, as well as official statistics. It is especially useful for big survey-style datasets interlinked in various ways. The ERIC CESSDA members are committed to applying either of the two main DDI standards (Codebook and Lifecycle). Additional adopters of the standard is the Dataverse project<sup>30</sup> and ICPSR<sup>31</sup> as well as many national statistical offices. The DDI is in line with the FAIR Principles including support for interoperability.

The new DDI-CDI standard - DDI CDI: Cross-Domain Integration | Data Documentation Initiative<sup>32</sup> - is a new specification from the DDI Alliance which has been developed to support cross domain research, and is intended to support interdisciplinary research in large-scale infrastructure ecosystems such as EOSC<sup>33</sup>.

DDI-CDI is presented as a specification which fills in important gaps in the areas covered by other standards, driven by the needs of implementers. Often, structural metadata differs across domains, and acts as a way of incorporating domain semantics with the data. DDI-CDI provides a model for integrating the disparate data structures and semantics, to facilitate

<sup>30</sup> <https://dataverse.org/>

<sup>31</sup> <https://www.icpsr.umich.edu/web/pages/index.html>

<sup>32</sup> <https://ddialliance.org/Specification/DDI-CDI/>

<sup>33</sup> <https://zenodo.org/record/4707263>

analysis of integrated inputs. It was built around a set of external projects being conducted by Alliance members and participants, which could test and feedback input into the model as it evolved.

Requirements for the DDI-CDI standard to integrate sensor data, event data, and key-value “big” data were prominent as was the need to provide a more complete description of data provenance. As such, it is a domain agnostic specification and can interact with a range of metadata standards. The EOSC Future “Climate Neutral and Smart Cities” project illustrates how DDI-CDI can be used for the description of provenance and processing in combination with domain standards<sup>34</sup>.

### **Access Policies**

A set of controlled vocabularies is developed by the DDI Alliance to be used to describe specific aspects of a dataset in DDI specifications and for other use. The controlled vocabularies cover specific domains and cross-domain subjects. They support interoperability and machine-actionability. Most of these controlled vocabularies are integrated in the CESSDA Vocabulary Service<sup>35</sup>. Unfortunately, none of the DDI vocabularies support legal interoperability.

### **Copyright and Intellectual Property**

There is no mention of this in the schema.

### **Data Protection and Privacy**

There is no mention of this in the schema.

### **Conclusion**

Neither the DDI Standards nor the DDI-CDI standard provides metadata on legal aspects. However, especially with reference to the DDI-CDI standard there are obvious possibilities for integration to other standards/metadata schemas better suited for addressing legal aspects.

*“DDI-CDI offers a new type of specification which could help to realise the capture, interchange, and use of metadata throughout the EOSC data-sharing infrastructure, and could do so in ways which are scalable and machine-actionable. It operates at the needed level of granularity and would work to heighten the utility of semantic mapping and approaches to the full utilisation of data.”<sup>36</sup>*

<sup>34</sup> <https://www.europeansocialsurvey.org/esslabs>

<sup>35</sup> <https://vocabularies.CESSDA.eu>

<sup>36</sup> The Role of DDI-CDI in EOSC: Possible Uses and Applications <https://zenodo.org/record/4707263>

## 4.12 RO-Crate

RO-Crate<sup>37</sup> (table 12) is a community created to define a lightweight specification for packaging research objects. This specification is based on schema.org<sup>38</sup> annotations in JSON-LD. Schema.org provides a vocabulary to describe structure data on the Internet. Concepts as Organizations, Persons, Digital Objects and Actions can be modelled in a standard vocabulary, as well as their associated metadata. RELIANCE members are part of the ro-crate initiative<sup>39</sup> and have contributed to its specification process.

The use of this specification reports some advantages:

- All the research objects are homogeneous.
- The research objects generated can be used by a wider community
- The research objects can be manipulated for other tools generated by the community. Some examples can be found online<sup>40</sup>.

The research objects used JSON-LD as format files. JSON-LD is designed to represent linked data on the Internet allowing in our case to express the relations between the different resources of a research object. Also, the use of this format facilitates the consumption of research objects for other services, it means, gives support to a M2M (Machine to Machine) model.

According to RO-Crate, a research object has the following structure:

- RO-Crate Metadata file. This file must be written with the format JSON-LD 1.0 and has to be named like ro-crate-metadata.json.
- RO-Crate Website. In this case, a human readable version of the research object can be attached. In our case, this version is supported by the platform ROHub, where our ROs are shared.
- Payload files and directories.

**Table 12 – RO-Crate fact sheet**

Name	<b>RO-CRATE Metadata Specification</b>
Author	researchobject.org community
Purpose	Model the Research Objects
Type	Recommendation
Current version	1.1
Issue date	2023-04-26
References	Schema.org, Dublin Core
Formalism	JSON-LD
URL	<a href="https://www.researchobject.org/ro-crate/1.1/">https://www.researchobject.org/ro-crate/1.1/</a>

## Access Policies

The RO-Crate specification does not include a specific element designed explicitly for describing access policies.

<sup>37</sup> <https://www.researchobject.org/ro-crate/>

<sup>38</sup> <https://schema.org/>

<sup>39</sup> <https://www.researchobject.org/ro-crate/community.html>

<sup>40</sup> <https://www.researchobject.org/ro-crate/tools/>

## ***Copyright and Intellectual Property***

The RO-Crate specification includes a special property to describe the licence of the work, following the concepts defined by schema.org in the concept CreativeWork<sup>41</sup>. You can add the specific url with the definition of the licence using the property @id. This attribute should always be a url. Also, there is another attribute to refer to the entity of the property rights, copyrightHolder. RO-Crate can differentiate the author of a CreativeWork from the owner of the rights.

## ***Data Protection and Privacy***

The RO-Crate specification does not include a specific element designed explicitly for describing data protection and privacy. The element 'Contextual Entity' could be of (limited) use for legal data protection purposes.

## ***Conclusion***

This specification is specially designed to model research objects and it is mostly based in schema.org. For this reason, the presence of metadata related with legal interoperability is limited. However, based on the nature of a research object, it can be used to define rights in each component of a research object, avoiding the limitation imposed by the nature of a digital object. We must highlight that this specification has attributes used to define the provenance of a digital object, which potentially can be used to establish mechanisms to allow the definition of access policies or privacy rules.

---

<sup>41</sup> <https://schema.org/CreativeWork>

## 5. Use cases

---

This phase will comprise the description of how project partners have implemented the schemas and controlled vocabularies previously considered to reflect legal or organisational constraints associated with their datasets. These use cases will provide feedback to be used as input for the final recommendations for a core metadata framework for legal interoperability.

### 5.1 EMBL - BY-COVID use case

#### Overview

BY-COVID with WP3 is focussed on services for the discovery and integration of COVID-19 data by delivering a flexible, tiered metadata discovery system across different domains, metadata standards, and maturity/robustness levels of data sources. This enables the linking of FAIR data and metadata on SARS-CoV-2 and COVID-19, on other related viruses and diseases, and on socio-economic consequences, across research fields, from omics, clinical, and epidemiological research, to social sciences and humanities.

#### Introduction

The key aim of this use case is to establish the basis for the further development of the COVID-19 Portal metadata discovery and identify a workable level of granularity that can be made visible through the COVID data portal to describe data access attributes.

#### Challenges to be addressed

Researchers searching for data to reuse through the COVID data portal can find it difficult to know which data are openly accessible and which may only be accessible following the submission of a formal data access request. By improving the description of the data access conditions for data sets made visible through the COVID data portal, researchers are better able to make decisions as to which data they will use. A range of approaches are currently used by the contributing data repositories. For instance, EGA makes use of DUO ontology<sup>42</sup> while the Dutch metadata catalogue employs DUC<sup>43</sup> which includes the DUO ontology along with the CCE<sup>44</sup> terms and the ICO<sup>45</sup> ontology. Previous attempts to harmonise data access descriptions across the CESSDA<sup>46</sup> archives have shown that there are a plethora of local approaches in use. For this use case to be viable, the COVID data portal must ensure that their selected approach is flexible enough to ensure that the different data access schema can be accommodated.

The BY-COVID team started looking at how DCAT and ODRL could be used to provide sufficient information about data access to researchers using the COVID data portal while keeping the burden on the contributing data repositories to a minimum. Given the

---

<sup>42</sup> <https://ega-archive.org/data-use-conditions>

<sup>43</sup> <https://duc.le.ac.uk/>

<sup>44</sup> <https://zenodo.org/records/8200079>

<sup>45</sup> <https://github.com/ICO-ontology/ICO>

<sup>46</sup> <https://www.cessda.eu/>

heterogeneity of the sources and the time constraints, it was decided to stick to the Bioschemas DataCatalog<sup>47</sup> profile already in use on the COVID data portal and map them (DCAT, DDI).

### ***Expected impact***

The anticipated outcome is that searching for relevant data to reuse is made much easier for researchers while respecting the variety of metadata approaches currently being used by the contributing repositories from the Life Sciences (EGA, Dutch Metadata catalogue) and the Social Sciences (CESSDA).

### ***Expected outputs***

These metadata have been integrated from their original representation in resource-specific formats (e.g., FAIRsharing, DCAT, etc.) demonstrating the flexibility of the underlying indexing platform. However, legal constraints associated with the resources remain described superficially, as aggregating heterogeneous sources implies a common set of attributes which lowers the level of details possible. Thus, policies such as access or reuse rules remain described at a high level, generally in free-text fields which prevents the machine-actionability of the information. The lack of uptake and expertise of specific controlled vocabularies or ontologies (ODRL, DUO, etc.) in the different communities involved also proved to set back the legal and organisational interoperability.

## **5.2 SIKT - ESS use case**

### ***Overview***

The use case provides an example of establishing an interdisciplinary data resource, combining social, climate, and environment data for use by researchers, primarily within the social science domain. Data is made accessible via the ESS LABS European Social Survey<sup>48</sup>.

The key aim here is to identify how aspects of legal and organisational interoperability have been approached in the use case. Some of the issues which could be addressed more effectively in future have been highlighted, along with possible solutions.

The anticipated outcome is that searching for and using relevant environmental and climate data together with social science survey data is made much easier for researchers, students and others while still providing credit to and building on the reputation of the contributing partners from the SSHOC and ENVRI clusters for the data in which they specialise. Ultimately, this leads to greater trust in the data on the part of researchers.

### ***Introduction***

<sup>47</sup> [https://bioschemas.org/profiles/DataCatalog/0.3-RELEASE-2019\\_07\\_01](https://bioschemas.org/profiles/DataCatalog/0.3-RELEASE-2019_07_01)

<sup>48</sup> <https://www.europeansocialsurvey.org/esslabs>

The Science Project 9 (SP9) *Climate Neutral and Smart Cities* is a Horizon 2020 funded Science Project<sup>49</sup> under EOSC Future Task 6.3: EOSC Future Climate-Neutral and Smart Cities<sup>50</sup>.

This project aims to contribute to this Horizon science mission by combining resources from two Science Clusters<sup>51</sup>. As part of the SSHOC consortium<sup>52</sup>, ESS<sup>53</sup> collects data related to political and social trust, health and health inequality, attitudes towards climate change and energy, understandings and evaluations of democracy and digital communication at work and with family, amongst many other topics related to the smart agenda. There is environmental data describing climate and air quality conditions related human attitudes and conditions as part of the ENVRI Cluster<sup>54</sup>.

Scientists from the SSHOC and ENVRI clusters have been collaborating to produce new outputs, including indicators for environmental indices and computational methods. European Social Survey data are being integrated with environmental indicator variables created for the project in a selection of major European city regions.

Access to data and indicators for areas new in the context of the ESS, as well as to a prototype tool for describing the provenance of the data (data harvesting, indicator production, data integration etc.) is provided by the project. Here the focus is on legal and organisational aspects for (meta)data enhancing/hindering interoperability for the use case.

From the task deliverable it is said that: “Results show the needed form of interoperability at several levels: organisational, scientific, semantic, and technical.” For this use case the need for legal and organisational interoperability will be addressed here.

### ***Challenges to be addressed***

The use case will give insights to the handling of cross-domain and cross-disciplinary data and of various provenance of relevance to organisational aspects by the EOSC Future Science Project Climate Neutral and Smart Cities<sup>55</sup>.

The project has chosen the DDI-CDI metadata standard for documentation of data provenance. This covers a certain aspect of organisational interoperability, as far as all business steps in the data integration process is described in detail, using the provenance application prototype developed through the project. This enables detailed exploration of procedures used by researchers, as well as for institutions performing similar tasks.

However, the standard does not support legal interoperability in ‘ready-to-use’ ways.

<sup>49</sup> The Climate Neutral and Smart Cities Science Project of EOSC Future Task 6.3 has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101017536

<sup>50</sup> <https://eoscfuture.eu/data/climate-neutral-and-smart-cities/>

<sup>51</sup> <https://www.eosc-hub.eu/eosc-hub-and-esfri-cluster-projects>

<sup>52</sup> <https://sshopencloud.eu/>

<sup>53</sup> <https://www.europeansocialsurvey.org/>

<sup>54</sup> <https://envri.eu/envri-fair-a-new-cluster-project-of-environmental-ris-has-been-launched/>

<sup>55</sup> <https://ess.sikt.no/en/study/71586b4f-ef66-4b90-aed7-e7e7ad7406ce>



Copyright/ownership, licensing and Intellectual Property rights is dealt with manually in the EOSC Future project. Data Protection and Privacy issues are not dealt with in the 'Climate Neutral and Smart Cities' Science Project, because data from the ESS are considered as anonymous on the level of geographic granularity that the data are made available on.

Copyright/ownership, licensing and Intellectual property rights is dealt with by Task 1.4 in EOSC Future, and a related matrix is to be filled in by all Science Projects (SPs) with help from Task 1.4 and an expert on EU laws. This includes information about the core deliverables, foreground, background, third-party and side ground IP, copyright/ownership, and licences, all expressed in free text.

While EOSC infrastructure components developed under EOSC Future are regarded as EOSC Future exploitable IP results, and need a contract in the form of a single ownership declaration or joint ownership contract for e.g. code, the same does not apply to most results from the science projects. The results of the SPs are very valuable for dissemination purposes for EOSC Future however, meaning the dissemination artefacts do form part of the EOSC Future IP results.

All textual products e.g. web sites and reports are attributed to the authors. The expert on EU laws supported a similar procedure for authors of reports. Related research papers are owned by authors, who have the intellectual properties.

The project has set up a Git repository for technical outputs such as code, description etc. including information about owner and licence.

For the provenance application prototype a single ownership declaration has additionally been used.

In addition software has been granted an open licence in accordance with the EC contractual documents.

For the integrated data being produced by the project there is no question of IP rights as they are based on data from open sources see e.g. Copernicus - Access to data<sup>56</sup> and EEA Data Policy<sup>57</sup> and ESS. Further - and more generally - data are not copyrightable per se, being considered "facts" from some legal perspectives. There is a lack of clarity regarding IP issues around data, which happily does not cause problems in this use case given that all the source data is open.

However, licensing and attribution of data is still needed.

The cross-domain data is given a licence as the lowest common denominator of the CC licence<sup>58</sup>. [CC BY NC for ESS-data] + [CC BY for data on air pollution and air quality] => CC BY NC for the integrated cross-domain data. For intermediate data sets that include only environmental data, before integration with the ESS data, a CC BY licence is used.

<sup>56</sup> <https://www.copernicus.eu/en/access-data>

<sup>57</sup> <https://www.eea.europa.eu/en/datahub/eea-data-policy>

<sup>58</sup> <https://creativecommons.org/share-your-work/cclicenses/>

From the perspective of FAIR, where the ability of machines to better support manual efforts regarding data access is concerned, there are still some issues. In this use case, there is no automation involved. The licence is given manually (see licensing for data<sup>59</sup>), but could potentially be provided in a machine-actionable fashion. This would require a well defined vocabulary for licensing to be defined, which does not exist today. Even the simplest functions, such as finding the text of a licence programmatically (and determining that such text is available, even) requires agreement between the parties providing and accessing the data on the mechanism for doing so.

Some work in the area of data privacy has been started in the DPV group at W3C, but there have been some difficulties in describing GDPR. Other related W3C frameworks like ODRL also exist, but depend on agreements about licensing conditions which do not yet exist.

There is thus no standard way to find, navigate, or describe licences and related conditions of use. Even the ability to minimally support these functions is lacking, requiring manual effort.

### ***Expected Impact***

Even if there has been a strong focus on legal interoperability in particular in EOSC Future, the use case shows the lack of potential for machine-readability/actionability metadata for supporting these aspects of data reuse.

In practice, practical and ‘low-tech’ solutions have been found e.g. description in matrixes and free as the ‘modus operandi’ for the time being.

This shows the need to focus further on finding good semantic and technical solutions related to legal interoperability that could help to fill this gap.

This issue is starting to be addressed in some other EU projects. Notably, WorldFAIR is producing the “Cross-Domain Interoperability Framework” which will recommend a best practice regarding data access. Among the recommendations is the use of some of the basic ODRL fields regarding the publication of licence text in a standard fashion for static data sets and services, at the level of a DCAT vocabulary “Distribution” (DCAT being widely adopted in Europe). This would be one approach to supporting better machine-readability. Further recommendations in that work have identified the same need for agreed semantics around conditions of use and access, and that topic has allegedly spurred some work in European academic circles, notably at Twente in the Netherlands where there is a strong focus on ontology. (As of this writing, there is no citable output, however.)

## **5.3 UKDS - CESSDA use case**

### ***Overview***

The UKDS is a partnership between the Universities of Essex, Manchester, UCL, Edinburgh and Jisc and the UK service provider to CESSDA.

<sup>59</sup> <https://ess.sikt.no/en/study/71586b4f-ef66-4b90-aed7-e7e7ad7406ce/54>

The focus of this use case is on enabling machine actors to better interpret the currently ambiguous semantics of digital objects' access and usage conditions and secondly, to provide more specific guidance on how to encapsulate the definition and execution of access and usage conditions in FAIR signposting practices<sup>60</sup>. The latter reference "licence" as a link type but apart from referencing natural language licence statements, this mechanism currently provides little scope for subsequent machine-actionable negotiation and execution of access/usage conditions for a digital object.

Access and usage conditions are typically specified, asserted and managed locally. Beyond classifying these with some shared, loosely understood categories such as "Open" and "Closed", such categories are largely bespoke to a particular repository. For example, UKDS has three top level categories: Open, Safeguarded and Controlled and also supports embargoes. By way of contrast, OpenAire has openAccess, restrictedAccess, embargoedAccess and closedAccess<sup>61</sup>. Such access categories are pivotal for FAIR however they encapsulate and signify a set of complex attributes, constraints and workflows but in a currently non-normative way. For the purposes of long-term global interoperability, such locally-defined high-level access categories are currently of little practical use beyond simple discovery, as they only signify precise meaning locally to the repository that assigns them.

An access category is normally assigned as the end result of a (typically human) assessment of the intersection of (a) the regulatory/legal context, (b) rights and usage prescriptions of the data owner, and (c) the disclosure risk of the data (itself a function of inherent properties of the data in isolation as well as emergent properties when the data is combined with other data). In most cases, these assessments are often non-deterministic.

W3C standards such as ODRL<sup>62</sup> have emerged, which allows natural language rights statements to be formally represented as structured RDF data. This use case will create the first comprehensive coverage of ODRL statements for a national collection in the landing pages of UKDS "studies"<sup>63</sup> (the primary object that acts as a container for datasets and documentation). This is a foundational first step to providing a machine-actionable corollary to hitherto natural language based artefacts, such as licences and data sharing agreements.

## Introduction

For researchers, access to data, particularly sensitive data, is too complex and takes much more time than it should. Much effort has been devoted to machine-actionable implementations of the FAIR principles but in the access arena, less progress has been made. Access and usage conditions are derived from the intersection of a number of factors: legal & statutory obligations, rights management assertions, external prescriptions from data owners, and intrinsic properties of the digital object e.g. more disclosive data will inevitably require more stringent access protocols. With the global recognition that interoperability will lead to better global services for researchers, access is no longer a second-order problem. Mediating researcher access to data has become a topic we can no longer leave

<sup>60</sup> <https://signposting.org/FAIR/#reltypes>

<sup>61</sup> <https://wiki.surfnet.nl/display/standards/info-eu-repo#infoeu-repo-AccessRights>

<sup>62</sup> <https://www.w3.org/TR/odrl-model/>

<sup>63</sup> <https://beta.ukdataservice.ac.uk/datacatalogue/studies>

primarily to humans' best administrative efforts, still largely informed by natural language licence artefacts. Rights statements, legal obligations and access workflows need to be systematically modelled and implemented in metadata and code, in order to be executed at scale by machines. ODRL, while not complete in its coverage of all aspects of rights and access management, is currently the most practicable way forward to deliver better access interoperability.

### ***Challenges to be addressed***

Attempting to harmonise top-level access categories across domains and repositories is unlikely to be a fruitful course of action: consider that CESSDA data access policy<sup>64</sup> took several years to reach agreement on the most coarse-grained access categories. We will pursue a more granular, bottom-up approach that establishes core vocabularies for the key ODRL classes i.e. Parties, Permissions, Obligations, Prohibitions and Actions and best practice for representing this in ODRL policies. In practice, there are a finite number of items in these core vocabularies for the majority of access-related repository activities. Once they are available to deploy in ODRL policies, this will be a significant step forward in effectively modelling the definition of traditionally prose-based access/usage conditions statement. It is a precursor to a future goal (not in the scope of this use case) of connecting ODRL Actions to machine-actionable workflow definitions modelled in, for example, Common Workflow Language, leading to full end-to-end machine-actionable messaging and process choreography between repositories.

The machine-actionable access arena is relatively immature – compared to practices around discovery, for example. Simply communicating why this is important is not a trivial task to communities administratively and culturally accustomed to dealing with researchers' access to data as a largely human-mediated activity.

As well as providing a real-world production implementation of ODRL, we will provide guidance, both technical and more governance-related: the terminological overlaps and relationship between licensing, rights management, access and usage (among others) remain a barrier to precise articulation of problem statements and the design of solutions in response.

### ***Expected impact***

Working with our partners in CESSDA in the newly created Sensitive Data Working Group, UKDS will endeavour to be the exemplar for an initial real-world implementation of ODRL and will encourage and advocate for the uptake of similar practice by other Service Providers in CESSDA. The expected benefits for data consumers include:

- Medium term:
  - transparency and efficiency in requesting data
  - consistency of access experience across different service providers
- Longer term:
  - automated processes and services across service providers
  - foundational infrastructure for future B2B federation of access workflows

<sup>64</sup> <https://doi.org/10.5281/zenodo.6722000>

For service providers, the standardised and structured approach to access through ODRL and associated controlled vocabularies will provide:

- Medium term
  - Guidance on minimal best practice and design patterns for new systems development
  - Equity, and transparency in processing access requests
- Longer term
  - The ability to track and evaluate access requests more systematically, helping provide more robust evidence to inform improvements to access management practices.
  - Opportunities for service providers to participate in multi-organisational and cross-domain collaborations

### ***Expected outputs***

The following tangible outcomes or solutions will be sought:

- Best practice documentation for embedding machine-actionable ODRL statements in resource landing pages and how this interacts with current FAIR signposting practice.
- Production implementation in UKDS catalogue<sup>65</sup>.

## **5.4 UPM - RELIANCE use case**

### ***Overview***

RELIANCE is an european project to extend the EOSC service offering with a set of industry-strong, innovative, interconnected services for the open, efficient, and cross-disciplinary management of the research lifecycle. Users will be able to pack all your research products in a research object, following the ro-crate specification.

The focus of this use case is to explore the use of privacy rules in a research object. In this case, we face some challenges related to the nature of the Research Object. How can the licences of each resource of the RO be combined/integrated? How can the specification used by the RO be extended?

As a result, a guideline has been produced with a collection of recommendations to extend the legal framework of a RO based on the analysis of the licences of the ROs present in the platform ROhub.

### ***Introduction***

A Research Object is a composition of research artefacts (digital objects used on a research), defining the relations among them. From the legal point of view, this sets out the problem about how we can define the property of each research artefact, and how we can establish the property rights of the Research Object. Below are some examples:

- Changes in air and water quality during the Covid-19 Lockdown in the Venice Lagoon<sup>66</sup> (rohub.org)

<sup>65</sup> <https://beta.ukdataservice.ac.uk/datacatalogue>

<sup>66</sup> <https://www.rohub.org/eec6faaa-e133-47d4-b377-44f7d06a9654?activetab=overview>

- Impact of the Covid-19 Lockdown on Air quality over Europe<sup>67</sup> (rohub.org)

### ***Challenges to be addressed***

Two main challenges in this use case have been identified

- The harmonisation of the licences and rights of each artefact with the licence and rights of the research object. This involves the accessibility of the resources by researchers and other artefacts. For it, we will analyse the licences used in the RO-hub platform.
- As seen in [section 4](#), RO-Crate has some limitations to express legal metadata. Based on the analysis done in this document, we plan to extend the legal framework of RO-Crate with new metadata to allow us to express the legal complexity of a Research Object.

### ***Expected outputs***

A small report will be produced analysing the licences used in the platform, classifying them based on the domain and the nature of the resource (publication, data, software, etc.). This information will be very valuable to the platform Rohub in order to improve the quality of their Research Objects. Also, a list of recommendations to extend the RO-Crate specification will be included in the report in order to improve the legal interoperability among the Research Objects.

---

<sup>67</sup> <https://www.rohub.org/53aa90bf-c593-4e6d-923f-d4711ac4b0e1?activetab=overview>

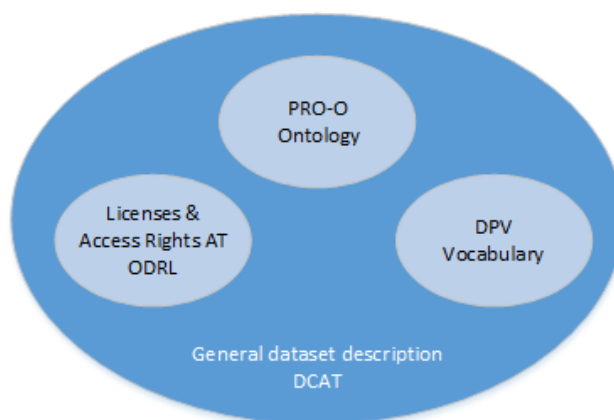
## 6. Core metadata schema for legal interoperability

The landscape analysis summarised in [Section 4](#) was an opportunity to identify and evaluate the different tools relevant to overcome the challenges. The use cases in [Section 5](#) highlighted the issues faced by different communities and their attempts to address them.

From the metadata standards designed specifically for describing datasets in a machine-readable format which have been evaluated, DCAT seems to be the tool of choice as it provides a standardised way to describe datasets, facilitates interoperability between data catalogues, portals, and repositories, and allows for the enrichment of metadata by providing a rich, expandable set of properties to describe various aspects of datasets. Its uptake by different communities is large, and a number of application profiles exist to reflect specific usage and implementation. It integrates widely-used other generic standards (DublinCore) and has already been mapped to others (e.g., Datacite<sup>68</sup>, ADMS, schema.org<sup>69</sup>, etc.). DCAT editors<sup>70</sup> and validators are available online, which makes it handier to integrate.

In order to enable the comparison and machine-actionability of the legal constraints bound to datasets, controlled vocabularies are the most effective and flexible tool at our disposal and they should complement the DCAT metadata schema efficiently. Different sub-topics have been identified, for which solutions exist, namely: licences, access policies, intellectual property rights, data privacy.

Figure 1 – Core metadata schema for legal interoperability



### Access Policies

The Access Rights<sup>71</sup> AT is a controlled vocabulary listing the access rights or restrictions to resources. It is designed for but not limited to DCAT simple descriptions of access rules for datasets (e.g., public, sensitive, confidential). These two authority tables are maintained by the Publications Office of the European Union on the EU Vocabularies website.

<sup>68</sup> <https://ec-jrc.github.io/datacite-to-dcat-ap/>

<sup>69</sup> <https://ec-jrc.github.io/dcat-ap-to-schema-org/>

<sup>70</sup> <https://rdforms.com/editors/dcat/>

<sup>71</sup>

<https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/access-right>

For more elaborated constraints, ODRL enables the definition of access control policies that express permissions and/or prohibitions associated with data. Some recommendations and examples about the integration of ODRL in DCAT are available<sup>72</sup>.

### ***Copyright and Intellectual Property***

The Licence<sup>73</sup> AT is a controlled vocabulary which exhaustively lists the different licences available internationally which can be tagged on publications, datasets, databases or software. DCMI terms include a RightsHolder property that can suffice for simple cases, but PROV-O<sup>74</sup> allows more complex descriptions, e.g. the organisation to which the rights holder belongs, or the context of the generation of the dataset. It has been mapped<sup>75</sup> to DCMI, and can be integrated<sup>76</sup> to DCAT by creating a DCAT catalogue as a subclass of a PROV-O entity.

### ***Data Protection and Privacy***

DPV can be utilised as a controlled vocabulary for invoking privacy and data protection-specific terms which can be used in ODRL. A specific AP exist for that purpose<sup>77</sup>.

---

<sup>72</sup> <https://w3c.github.io/dxwg/dcat/#license-rights>

<sup>73</sup> <https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/licence>

<sup>74</sup> <https://www.w3.org/TR/prov-o/>

<sup>75</sup> <https://www.w3.org/TR/2013/NOTE-prov-dc-20130430/#dct-rightsHolder>

<sup>76</sup> <https://github.com/w3c/dxwg/blob/gh-pages/dcat/rdf/dcat-prov.ttl>

<sup>77</sup> <https://besteves4.github.io/odrl-access-control-profile/oac.html>



## 7. Conclusions and next steps

---

The main takeaways from the use case descriptions above are the concrete existence of legal constraints in the research communities, and the lack of technical means or expertise to address them.

This proposal for a core metadata schema is an option to foster legal interoperability between research datasets, but others may exist. Few mappings between metadata schemas are already available, they could be improved with new ones which have been considered in this study (e.g., OpenAIRE guidelines) or others. The FAIRCore4EOSC has deployed the MSCR<sup>78</sup> registry which could ease the access and thus enhance the uptake of such mappings effectively.

This recommendation could also be used as input for the new, to-come EOSC Interoperability TF (current Semantic Interoperability, Technical Interoperability and AAI task forces will likely be merged in a larger TF with subgroups), and support programs could be put together to stimulate the adoption of good practices in the domain.

---

<sup>78</sup> <https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr>