

## D3.4: Open Science Policy Registry

|           |  |
|-----------|--|
| Author(s) | Leonard Mack (Jisc), Elli Papadopoulou (ARC) |
| Status    | Final version                                |
| Version   | v1.0   |
| Date      | 24/10/2018                                   |

### 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)  |

### Abstract:

This report introduces the framework for the EOSCpilot Open Science Policy Registry, a (semi-) automated system to register and submit, assess, and validate policies of EOSC service providers and stakeholders. Together with the Open Science Monitor and Open Science Policy Toolkit, it complements the policy-supporting services of EOSCpilot WP3 – Policy.

The Policy Registry framework is designed to serve four overarching functions, based on a modular architecture. These functions include the registration and submission of policy metadata; the assessment and validation of policy compliance; the storage of submitted policy (meta-)data in the registry's database; and the provision of data for the Open Science Monitor and other secondary data users (e.g. EOSC governance bodies).

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

| Document identifier: EOSCpilot -WP3-D3.5 |                                 |
|--|---------------------------------|
| Deliverable lead                         | Jisc                            |
| Related work package                     | WP3 - Policy                    |
| Author(s)                                | Leonard Mack, Elli Papadopoulou |
| Contributor(s)                           | George Papastefanatos           |
| Due date                                 | 30/08/2018                      |
| Actual submission date                   | 24/10/2018                      |
| Reviewed by                              | Donatella Castelli, Ilaria Fava |
| Approved by                              | Mark Thorley (UKRI)             |
| Start date of Project                    | 01/01/2017                      |
| Duration                                 | 24 months                       |

## Versioning and contribution history

| Version | Date       | Authors  | Notes  |
|---------|------------|--|--|
| 0.1     | 2018-05-10 | Leonard Mack   | Initial draft structure set up                         |
| 0.2     | 2018-06-25 | Leonard Mack   | Data model revised                                     |
| 0.3     | 2018-08-07 | Leonard Mack   | Further revisions of data model and use case workflows |
| 0.4     | 2018-09-09 | Leonard Mack   | Pre-review draft completed                             |
| 0.5     | 2018-09-24 | Leonard Mack, George Papastefanatos, Elli Papadopoulou | Revised data model and conceptual model                |
| 0.6     | 2018-10-15 | Leonard Mack   | Reviewer comments addressed and revisions applied      |
| 0.7     | 2018-10-17 | Leonard Mack   | Added Appendix A                                       |
| 0.8     | 2018-10-17 | Leonard Mack   | Submission-ready version                               |
| 1.0     | 2018-10-24 | Mark Thorley   | Minor typographic edits before submission              |

**Copyright notice:** This work is licensed under the Creative Commons CC-BY 4.0 license. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0>.

**Disclaimer:** The content of the document herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the document is believed to be accurate, the author(s) or any other participant in the EOSCpilot Consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the EOSCpilot Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the EOSCpilot Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

## TABLE OF CONTENTS

|   |           |
|---|-----------|
| <b>EXECUTIVE SUMMARY .....</b>                              | <b>5</b>  |
| <b>1. INTRODUCTION .....</b>                                | <b>6</b>  |
| 1.1. Overview of the Open Science Policy Registry .....     | 6         |
| 1.2. The Policy Registry in context.....                    | 7         |
| 1.3. Structure and purpose of this document .....           | 9         |
| <b>2. FOUNDATIONS OF THE POLICY REGISTRY .....</b>          | <b>10</b> |
| 2.1. Policy types.....                                      | 10        |
| 2.2. User groups and use cases .....                        | 13        |
| 2.3. Requirements of the Policy Registry .....              | 14        |
| 2.3.1. Requirements for the Policy Registry model.....      | 15        |
| 2.3.2. Workflow and processes of the Policy Registry.....   | 16        |
| 2.3.3. Policy Registry Services.....                        | 16        |
| <b>3. POLICY REGISTRY MODEL AND ARCHITECTURE .....</b>      | <b>19</b> |
| 3.1. Conceptual model.....                                  | 19        |
| 3.2. Architecture.....                                      | 20        |
| 3.3. Data Model .....                                       | 23        |
| <b>4. POLICY REGISTRY WORKFLOWS AND DECISION FLOWS.....</b> | <b>27</b> |
| 4.1. General workflow .....                                 | 27        |
| 4.2. Use case 1 action and decision flow.....               | 28        |
| 4.3. Use case 2 action and decision flow.....               | 31        |
| <b>5. DISCUSSION AND FUTURE WORK .....</b>                  | <b>33</b> |
| 5.1. Interdependencies.....                                 | 33        |
| 5.2. Future work .....                                      | 34        |
| <b>6. CONCLUSIONS .....</b>                                 | <b>35</b> |
| <b>ANNEX A. POLICY REGISTRY DATA MODEL.....</b>             | <b>36</b> |
| <b>ANNEX B. GLOSSARY.....</b>                               | <b>53</b> |

## LIST OF FIGURES

|  |    |
|--|----|
| Figure 1: Policy Supporting Services in context.....               | 8  |
| Figure 2: Policy types relevant to the Policy Registry.....        | 10 |
| Figure 3: Functions of Policy Registry in different use cases..... | 14 |
| Figure 4: Policy Registry model .....                              | 19 |
| Figure 5: Policy Registry architecture .....                       | 21 |
| Figure 6: Policy Registry Data model.....                          | 24 |
| Figure 7: Policy Registry functions and process workflow .....     | 27 |
| Figure 8: Use case 1 workflow .....                                | 29 |
| Figure 9: Use case 1 decision flow.....                            | 30 |
| Figure 10: Use case 2 workflow .....                               | 31 |
| Figure 11: Use case 2 decision workflow.....                       | 32 |

## LIST OF TABLES

|   |    |
|---|----|
| Table 1: Policy Registry use cases..... | 13 |
|---|----|

## EXECUTIVE SUMMARY

This report introduces the framework for the EOSCpilot Open Science Policy Registry, a system for EOSC service providers and policy stakeholders to register policies as well as to validate their alignment with EOSC policies. The Policy Registry is connected to and builds on the policy requirements emerging from the EOSCpilot Rules of Participation and Policy Recommendations. Together with the Open Science Monitor and Open Science Policy Toolkit, it complements the policy-supporting services of EOSCpilot WP3 – Policy.

The Policy Registry framework is designed to serve four overarching functions, based on a modular architecture:

- 1) *Registration and submission of policy metadata* in relation to different EOSC policies (i.e. Rules of Participation and Policy Recommendations);
- 2) *Assessing policy alignment and validating policy compliance* of EOSC service providers and general policy stakeholders, based on the organisational and contractual policy metadata which they provide;
- 3) *Storage of submitted policy (meta-)data* in the registry's database;
- 4) *Provision of data* for the Open Science Monitor and other secondary data users such as EOSC governance bodies.

The Policy Registry specifications are closely aligned with the logical structure and contents of different policy types. Furthermore, the specifications relate to the different roles of relevant users in the context of the EOSC. Therefore, this report logically deduces two use cases from the policy propositions included in the EOSC Draft Policy Recommendations and Rules of Participation as well as related aspects such as the EOSC Service Portfolio and Service Management Framework. Use case 1 is centred on EOSC service providers who seek compliance validation to determine whether the policy-provisions for their services are “EOSC compatible” according to the Rules of Participation. Use case 2 focuses on EOSC policy stakeholders (i.e. RPOs, RIs, funders / policy-makers) who seek an assessment of the alignment of their organisational policies with EOSC's Policy Recommendations.

To serve these use cases, an adaptable and extensible system model, architecture, and metadata model are proposed. To enable the automation of policy assessment processes, the proposed initial metadata model uses controlled values where possible. In order to increase its interoperability and in line with contents of the EOSCpilot Rules of Participation (D2.5), the metadata model has been mapped to and reuses controlled vocabularies of the eInfraCentral data schema for services. Based on these uniform components, two action and decision flows are modelled to accommodate the process requirements of both use cases. The advantage of this approach is that it allows the Policy Registry to serve as a single mechanism to assess both policy assessment scenarios in the context of the EOSC.

The Policy Registry is highly interdependent with other components and services of the EOSC. The emerging, rapidly evolving characteristics of the EOSC make the further development of the Policy Registry an ongoing task. This is particularly relevant for the expansion of machine-readable and -actionable service components. Therefore, this report also discusses a short overview of ongoing interdependencies and proposed next steps.

## 1. INTRODUCTION

The EOSC Declaration from October 2017<sup>1</sup> presents the vision for the EOSC “as a data infrastructure commons serving the needs of scientists. It [...] will federate existing resources across national data centres, European e-infrastructures and research infrastructures [...] Resources, components and initiatives of pan-European relevance will be federated on the basis of objective criteria, agreed by stakeholder-driven governance [...]”

To ensure that these ambitious objectives can be met and that the EOSC can realise the added value from EU-wide collaboration, the declaration also highlights that the “EOSC must implement policy hand in hand with technology”. The EC’s staff working document on the EOSC Implementation Roadmap<sup>2</sup> iterates further, that the EOSC needs strong policy guidance to ensure its development path is predictable and coherent.

WP3 of EOSCpilot contributes to the development of this policy guidance. It proposes a set of Policy Recommendations as well as three policy-supporting services which are designed to ensure that the implementation of policies in the context of the EOSC is sustainable and scalable. This deliverable lays out the framework for the Open Science Policy Registry, a service to assess the policy alignment and compliance of EOSC stakeholders and service providers with the EOSC policies and requirements<sup>3</sup>. The Policy Registry aims to bridge the gap between the EOSC policies, i.e. Policy Recommendations<sup>4</sup> and Rules of Participation<sup>5</sup>, and their interpretation and adoption by EOSC stakeholders and service providers. For this, the Policy Registry is designed to assess the implementation of policy requirements, formulated by EOSCpilot, by their addressees (i.e. EOSC stakeholders and service providers). The Policy Registry covers policy aspects of Open Science and, as specified in the EOSCpilot Policy Recommendations, supporting policy domains in ethics, procurement, and data protection.

The Policy Registry captures highly relevant data to monitor the implementation practice and inform the evolution of EOSC policies. Accordingly, the Policy Registry could also provide data which can help to increase the operational clarity and predictability of the EOSC’s policy guidance.

The Open Science Policy Registry presented in this deliverable complements two other policy supporting services:

- The Open Science Monitor (D3.2)<sup>6</sup>, a monitoring framework to gather data based on a set of indicators for the implementation of Open Science practices and policies.
- The Policy Toolkit (D3.5), a best-practice collection of third party tools which EOSC stakeholders can use to inform and improve their Open Science practice.

### 1.1. Overview of the Open Science Policy Registry

Following the Description of Work, the Open Science Policy Registry is designed as an interactive policy validation service which allows the submission and quality control of different policies relevant in the context of the EOSC. The service is based on a policy metadata model, which serves to capture, structure, and represent policy information in the system’s database. The Policy Registry’s functions go beyond the mere registration of policies. They include:

<sup>1</sup> [https://ec.europa.eu/research/openscience/pdf/eosc\\_declaration.pdf](https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf)

<sup>2</sup> [https://ec.europa.eu/research/openscience/pdf/swd\\_2018\\_83\\_f1\\_staff\\_working\\_paper\\_en.pdf](https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf)

<sup>3</sup> EOSC policies are defined as the interconnected set of rules and principles adopted by the EOSC to achieve its mission and objectives. The primary EOSC policies are the Policy Recommendations and Rules of Participation; for this report, only the EOSCpilot Rules of Participation and Policy Recommendations have been considered as these currently provide the clearest basis for metadata and workflow development. The notion of policy requirements refers to the compliance and policy alignment requirements resulting from the EOSCpilot Policy Recommendations (T3.1) and Rules of Participation (T2.3).

<sup>4</sup> [https://eoscipilot.eu/sites/default/files/eoscipilot\\_d3.3\\_final-withannexes-forweb.pdf](https://eoscipilot.eu/sites/default/files/eoscipilot_d3.3_final-withannexes-forweb.pdf)

<sup>5</sup> <https://eoscipilot.eu/sites/default/files/eoscipilot-d2.5.pdf>

<sup>6</sup> <https://eoscipilot.eu/sites/default/files/eoscipilot-d3.2.pdf>

- 1) *Submission and registering of policy metadata*: Users of the Policy Registry can submit policy metadata relevant to different EOSC policies (i.e. Rules of Participation and Policy Recommendations);
- 2) *Assessing and validating* policy compliance in relation to the contractual policies (e.g. terms and conditions) of service providers and organisational policy alignment of EOSC stakeholders, based on the submitted policy metadata;
- 3) *Storing / archiving of submitted policy metadata* in the registry's database;
- 4) *Provision of data* for the Open Science Monitor and EOSC governance bodies.

The external purpose of the system is to assist users in assessing their level of alignment with EOSC Policy Recommendations and compliance with Rules of Participation. As laid out in section 2.2, these users are service providers (who seek compliance with the Rules of Participation) and policy stakeholders (who seek alignment with EOSC Policy Recommendations). Policy stakeholders can be RPOs, RIs, and funders or policy makers. From the perspective of these external users, the Policy Registry provides an access point to submit metadata relating to organisational and contractual policies<sup>7</sup>. For service providers, the assessment of this metadata should determine whether their services meet the EOSC's policy requirements for the service status level "EOSC compliant", according to the Rules of Participation. For other policy stakeholders, the assessment should determine the degree to which a stakeholders' policy adoption conforms with the EOSC's Policy Recommendations. Based on these two use cases, the Policy Registry is an essential service component to realise the vision of a federated service environment based on *common minimum policy standards*. In this view, the Registry operates as an application – or gatekeeper – service ensuring that EOSC service providers and policy stakeholders prove sufficient policy alignment (or compliance) when operating in the federated environment of the EOSC.

Within the EOSC (i.e. internally), the Policy Registry also supports the development of the EOSC's policy and governance framework by capturing information about the implementation of Open Science policies by policy stakeholders and service providers. The policy-specific metadata which is collected as part of the "application process" is not only stored in the system's database, but also aggregated and supplied to the EOSC Open Science Monitor and the EOSC Governance Board. Hence, the Policy Registry is also a mechanism to supply the EOSC with essential data to support informed decision-making on behalf of the EOSC.

## 1.2. The Policy Registry in context

Figure 1 displays the position and functions of the Open Science Policy Registry in relation to the wider policy and governance framework of the EOSC. The policy and governance framework consists of the Policy Recommendations (T3.1) and the Rules of Participation (T2.3). In the mid- and long run, the evolution of the Policy Recommendations and Rules of Participation would be shaped by the EOSC governance bodies.

As described in the Federated Service Management report (D5.3)<sup>8</sup>, it is assumed that eventually a mostly hierarchical relation between the EOSC's Policy Recommendations and Rules of Participation will emerge. This would mean that the EOSC Policy Recommendations define *high-level* principles for policy practice. The Rules of Participation would primarily operationalise these principles, formulating concrete requirements which interested stakeholders and service providers need to comply with, e.g. by embedding them in their terms and conditions as well as other contracts with end-users.<sup>9</sup>

<sup>7</sup> For a definition of different policy types in the context of the Policy Registry, see section 2.1.

<sup>8</sup> <https://eoscipilot.eu/sites/default/files/eoscipilot-d5.3.pdf>

<sup>9</sup> The Policy Registry's data model reflects this assumption in its three-tiered structure, subsuming metadata entities which relate to the Rules of Participation below higher level entities relating to the Policy Recommendations. See section 3.3 for further information.



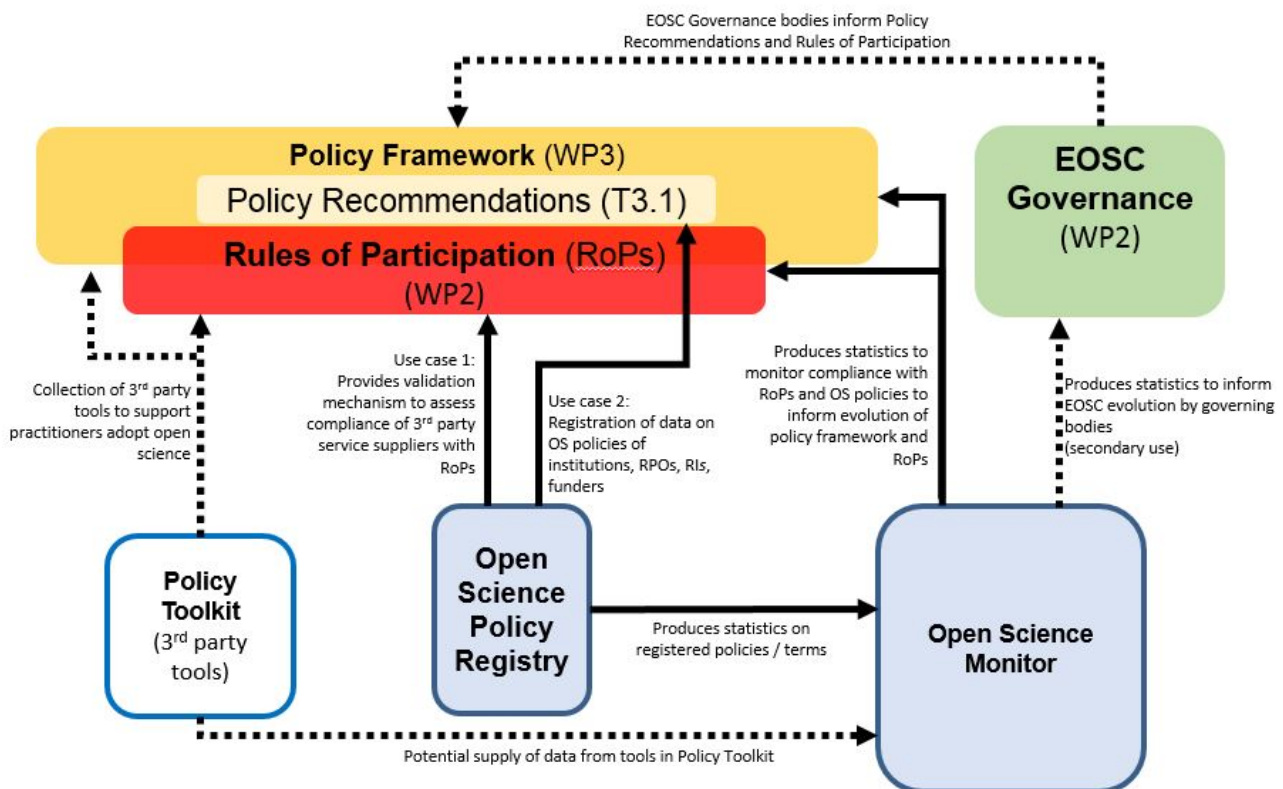


Figure 1: Policy Supporting Services in context

This logical structure implies that in order to meet the requirement of designing a framework which supports the EOsc’s policy and governance framework as *holistically* as possible, two separate use cases for the Policy Registry must be defined. Use case 1 addresses the scenario where of 3<sup>rd</sup> party service suppliers want to validate the compliance of their service policies *with the Rules of Participation* to achieve that the respective service is recognised as “EOsc compatible”. Use case 2 addresses the interest of other stakeholders (such as RPOs, institutions, and funders) to register policy metadata with the Policy Registry to assess their organisational policy alignment *with the Policy Recommendations*. Both use cases, will be explained and explored in more detail in section 2.2.

Within the proposed suite of the policy supporting services, the Open Science Policy Registry is closely related to the Open Science Monitor (D3.2), which monitors and gathers metrics on the implementation of Open Science based on 10 high-level monitoring targets. To further enhance this service, the Policy Registry can supply the Open Science Monitor with highly relevant micro-data on how policies are implemented by stakeholders. This data can be mapped e.g. to the OS Monitoring targets of Licences (openness/licences), FAIRness, and Policy Compliance. For the Open Science Monitor, this data provides an important snapshot of ground-level policy practice. It thus supports the OS Monitor’s task of producing metrics that inform the evolution of the EOsc’s policy framework and Rules of Participation as well as the wider evolution of the EOsc (steered by the EOsc governing bodies).

Additional context to the Registry is provided by the EOsc Service Portfolio (D5.2)<sup>10</sup> and the Federated Service Management Framework (D5.3)<sup>11</sup>. The former lays out the operational service framework and service categories for which the Policy Registry provides an assessment mechanism. Additionally, the Federated

<sup>10</sup> <https://eosc-pilot.eu/sites/default/files/eosc-pilot-d5.2.pdf>

<sup>11</sup> <https://eosc-pilot.eu/sites/default/files/eosc-pilot-d5.3.pdf>



Service Management Framework establishes the operational approach to service management which the Policy Registry will need to support adequately.

The Open Science Policy Registry thus shares a number of interdependencies with other EOSC components, which will affect design choices at the implementation stage of the system. Given these uncertainties, this deliverable defines an initial framework for the Policy Registry. However, more specific, operative design choices, e.g. the final database structure, will depend on the final state of other components. Obviously, the EOSC's final Policy Recommendations and Rules of Participation are crucial components which have a fundamental impact on the Policy Registry's data model. The conceptual data model presented in this report might therefore need to be revised once the Registry is implemented. Equally, the Service Portfolio and Federated Service Management framework define essential requirements for the Policy Registry's design. The specifications for the different service status levels and service types are important operational categories of the Service Portfolio, which need to be embedded in the Policy Registry. Similarly, the final choice for the service management federation will need to be reflected by the Registry. As the final outcome of these debates is still unclear, this report assumes that a loosely federated service promotion model is adopted and that the service categories proposed in the EOSCpilot service portfolio persist. However, depending on changes in these two areas, the operational workflows of the Policy Registry might need to be adapted when implementing the service. In this context, *the* crucial question is whether the EOSC will operate as a relatively loose federation of services, which the Policy Registry would support as a best-effort-service; or whether the EOSC will eventually emerge as a more tightly managed, quality-assured federation of services. The latter would necessitate more comprehensive controls on policy compliance and alignment than foreseen in this report. A discussion of interdependencies is provided in section 5.1.

### 1.3. Structure and purpose of this document

The remainder of this document is structured as follows: Section 2 provides an overview of the conceptual foundations of the Policy Registry as well as the system requirements which emerge from these considerations. Section 3 discusses the proposed conceptual model, systems architecture and data model for the Policy Registry. Section 4 models these components into an overall workflow for the Policy Registry and discusses the specific action and decision flows for the two use cases. Section 5 discusses the requirements for future work, including considerations of interdependencies in four areas and next steps. Section 6 concludes with a summary of findings.

## 2. FOUNDATIONS OF THE POLICY REGISTRY

This section lays out the conceptual foundations of the Open Science Policy Registry. It discusses the underlying typology of policy types, users and use cases, and requirements for the Policy Registry.

### 2.1. Policy types

The Open Science Policy Registry should support the implementation of EOSC policies, i.e. the rules and principles which are adopted by the EOSC to promote its mission and activities. These are (currently) specified in the EOSCpilot Policy Recommendations and Rules of Participation. A crucial underlying assumption for the Open Science Policy Registry is that these policies, of the European *Open Science* Cloud, will broadly support objectives of Open Science. Hence, by supporting the implementation of EOSC Policies, the Policy Registry also serves the *Open Science* element of its name<sup>12</sup>.

In functional terms, the Open Science Policy Registry supports the implementation of policies by providing a mechanism for policy stakeholders and service providers to register relevant policy metadata and assess the provided data for policy compliance or alignment. The main challenge of the Policy Registry is thus to provide a framework which translates human-centric policies into an appropriate, where possible machine-readable representation and workflow. However, the types of policies relevant in the context of the EOSC vary widely – which naturally impacts the design of the Policy Registry. Therefore, the most basic question for the design of the Policy Registry is: What policy types are relevant within the context of the Policy Registry?

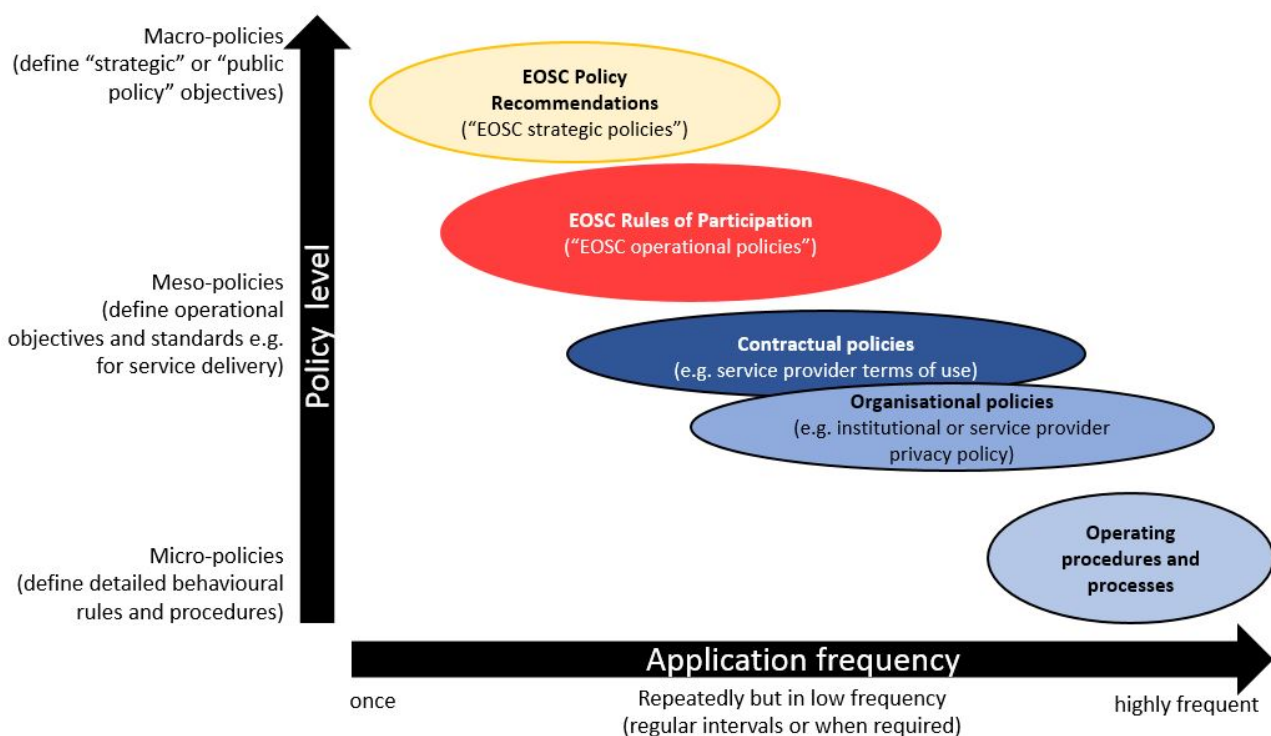


Figure 2: Policy types relevant to the Policy Registry

Figure 2 provides a schema of EOSC policy types. To inform the requirements of the Policy Registry, two dimensions are important:

<sup>12</sup> This underlying assumption is also in line with the definition of *Open Science Policy* as specified in the [Policy Toolkit report D3.5](#): "An Open Science policy is a set of rules and/or principles, usually formulated to advance the realisation of Open Science practices among a dedicated group of target stakeholders."

- **Policy level:** The first dimension consists of a spectrum of policy levels, covering high-level macro-policies, operational meso-policies, and micro-policies with detailed operative instructions. Macro-policies define strategic policy objectives, e.g. in the form of public policies or strategic policy recommendations. In some cases, macro-policies may indicate actions required to achieve stated objectives, e.g. in the form of roadmaps. However, these macro-policies usually do not contain details on *operational* objectives or even exactly describe actions or processes that need to be followed to achieve policy objectives. Meso-policies operationalise such objectives more clearly, e.g. in the form of specific performance indicators or contractual terms. Often, meso-policies also give an indication of the processes that are needed to achieve operational objectives. Nevertheless, only micro-policies describe highly detailed procedures and targets, ideally as step-by-step lists of tasks and actions. This information then enables practitioners to execute policies in a uniform, predictable manner.

Identifying different policy levels is highly important for the Policy Registry because inherent complications and limitations exist for the functions and degrees of automation that can be applied for policy types on different policy levels. In short, the clearer the procedural detail of the policy type which is the basis for a compliance or alignment assessment, the better can this policy type be modelled into a (semi-)automatic validation process for the Policy Registry.

- **Application frequency:** The second dimension consists of a spectrum to describe the application frequency of a policy, i.e. how often a policy needs to be applied. It can be assumed that the application frequency correlates positively with the required validation frequency; i.e. the higher the application frequency of a policy, the more often it needs to be validated. In general, macro-policies have a lower application frequency than meso- and micro-policies. Because micro-policies describe exact procedures to produce certain outputs in line with policy requirements, they need to be applied on a daily – or even second-by-second – basis. Instead, macro-policies often need to be applied only once (or at least infrequently), even though their introduction might be part of larger organisational reorganisations.

The execution frequency of a policy type raises important requirements for the Policy Registry because it defines how often the compliance or alignment of different policy types need to be assessed by the Policy Registry. The potential approaches for assessing policies with a very high execution frequency differ substantially from assessment systems with a lower frequency. However, because the focus of the EOSC's policy and governance framework so far is on policies with a singular or low-frequency execution, the Policy Registry framework is also geared towards these cases.

Along these two dimensions, five different policy types are relevant to the Policy Registry. The first two types are defined by the EOSC and are assumed to have influence on the latter three policy types, which are typically developed by external stakeholders or service providers:

1. **EOSC Policy Recommendations:** The EOSC Policy Recommendations define “high-level” (or “strategic”) policies adopted by the EOSC. They effectively serve as the strategic policy principles that relevant stakeholders, who want to engage with the EOSC, should follow. The expectation is that EOSC Policy Recommendations are reflected and further specified in the Rules of Participation. Where this is not the case, Policy Recommendations should directly seek to impact stakeholder's and service provider's organizational policies.
2. **EOSC Rules of Participation:** The Rules of Participation determine the EOSC's requirements for the participation of third parties in the EOSC (e.g. as service providers). In comparison to the Policy Recommendations, they are “operational” because they lay out operative management requirements which participating third parties need to follow, e.g. to ensure that their services are “EOSC compatible” (e.g. to comply with GDPR, adopt certain privacy-preserving practices, and ensure the FAIRness of data). Nevertheless, the Rules of Participation do not prescribe highly-specific processes or targets (e.g. benchmarks) to achieve these requirements. The impact of the Rules of Participation would be primarily on the formulation of contractual policies (e.g. terms and conditions of service providers). Additionally, the Rules of Participation may also directly impact on organizational policies which are closely connected to contractual policies. Examples for this are

privacy and security policies, which are frequently referenced in contractual terms and conditions, but are usually not part of a contract.

3. **Contractual policies:** These are policies that regulate the relations, define duties, and specify enforcement processes between two or more parties, usually in the form of a contract (e.g. service terms and conditions) or similar arrangements (e.g. service level agreements). Contractual policies are not defined by the EOsc, but by service providers (in the case of standardised terms and conditions) or by service providers and customers together (e.g. in the case of individually negotiated service level agreements). These policies contain detailed information on what the different parties must do to fulfil the aim of the contract. They usually give some information on the sequence of actions and sanctions – i.e. in what order actions must be performed and what follows from non-compliance.
4. **Organisational policies:** Service providers or other stakeholders usually define these policies to prescribe how – or according to which standards - certain actions are performed. Organisational policies thus regulate how a stakeholder or service provider manages its operations to achieve a certain objective. Examples for this are institutional privacy policies, the security policies of cloud storage providers, the procurement policies of funders, or institutional open science policies. Organisational and contractual policies are on a similar policy level because they usually describe and define operational requirements, processes, and duties (but do not specify these step-by-step procedures on a micro-level). However, in contrast to contractual policies, organizational policies do not have multiple parties (they are thus only “internal”) and usually do not detail external enforcement processes.
5. **Operating processes and procedures:** These come traditionally in the form of internal guidance or handbooks, describing how certain actions (e.g. when providing a certain service) are to be performed in exact procedural detail, ideally via step-by-step descriptions. Processes are constituted of multiple procedures to achieve a certain objective. Based on contractual and organizational policies, operating procedures thus define how actions are implemented and performed on a day-to-day (or second-to-second) basis. An example for this would be the internal operations adopted by a service provider to comply with a service level agreement. Operating procedures and processes can also be embedded in the design of technical systems, e.g. by automating access policies in the context of data sharing systems or a federated AAI infrastructure.  
Because operating procedures and processes manage the day-to-day delivery of e.g. products or services, they are likely most relevant to the operative service quality assessments which could be used to class services as “EOsc compliant”. However, the Policy Registry’s purpose is to support the selection of stakeholders and service providers which, based on the compliance or alignment of their policies with the EOsc’s policies, are “compatible” and therefore *generally* eligible for engagement with the EOsc. The purpose of the Policy Registry is *not* to assess on an ongoing basis the delivery and quality of services or other operations to class services as “EOsc compliant”.

According to this differentiation of policy types, the Policy Registry is designed as a solution for assessment processes involving the first four policy types, i.e. excluding operating procedures and processes. Depending on the evolution of the EOsc’s federation model, particularly if this progressed towards a tightly managed, quality-assured federation, it is conceivable that operating procedures and processes would also need to be monitored and assessed on a regular – if not ongoing – basis. This would demand that the Policy Registry’s compliance validation mechanisms integrate closely with the separate mechanism for “EOsc compliant” assessments. In this scenario, the Registry would need to integrate with potential mechanisms to automatically perform policy compliance assessments based on (technical) operating procedures; this could e.g. apply for the AAI protocols of cloud providers, whose technical code execution could be monitored in order to assess their compliance with EOsc access policy requirements. However, at least in its inception phase, the EOsc will likely emerge as a loose “service promotion” federation based on more inclusive standards. The framework presented here, reflects the simpler policy compliance requirements of this case.

Accordingly, the Policy Registry framework presented here is designed as a gatekeeper-solution to support the decision of whether specific stakeholders should *in principle* be admitted to engage with the EOSC. This might be necessary only once, when a stakeholder joins the EOSC federation for the first time, or repeatedly (but not very frequently), if the respective party changes its contractual or organisational policies. In the case of an update, the updated policies would need to be reassessed against the EOSC's policy requirements. This approach is also aligned with the intention to keep the EOSC's governance and participation requirements lightweight.

## 2.2. User groups and use cases

The Policy Registry is designed for two main user groups with specific policy assessment use cases in mind: Service providers and general EOSC policy stakeholders, i.e. RPOs, RIs, and funders / or policy makers. The use cases are summarised in table 1 below.

Table 1: Policy Registry use cases

|                          | <b>Use case 1 – Service providers:<br/>Validation of service policies</b>  | <b>Use case 2 – EOSC policy stakeholders:<br/>Registration of Open Science policy information</b>   |
|--------------------------|--|---|
| <b>Use case summary</b>  | Validation of service policy compliance with EOSC's Rules of Participation for third parties who want to provide their services as part of the EOSC  | Registration of information on Open Science-related policies in order to monitor alignment with EOSC's policy framework and recommendations   |
| <b>Main users</b>        | Third-party service suppliers (i.e. providers who want their services to be recognised as "EOSC compatible")   | Stakeholders which use EOSC services and/or have an interest in being recognised as Open Science policy adopters  |
| <b>What is assessed?</b> | Contractual policies (terms and conditions of service providers) and, if applicable, organisational policies (e.g. privacy policy of a service)  | Organisational policies (e.g. institutional privacy or Open Science policies)   |
| <b>Who benefits?</b>     | <u>Service suppliers</u> : get certified for EOSC service status level ("EOSC compatible")<br><br><u>EOSC</u> : collect detailed data on compliance with RoPs, allowing to manage EOSC service catalogue | <u>Stakeholders (= RPOs, RIs, funders)</u> : assess implementation of policies relating to Open Science<br><br><u>EOSC</u> : receives data on OS policy implementation by relevant stakeholders, facilitating development of EOSC policy framework and governance |

In both use cases, the Policy Registry intermediates between external policies (developed and adopted by service providers or policy stakeholders) and EOSC's own policy principles, expressed through the Rules of Participation and Policy Recommendations. However, how the Policy Registry realises its *policy-supporting* function varies depending on the use case. Figure 3 visualises these differences.

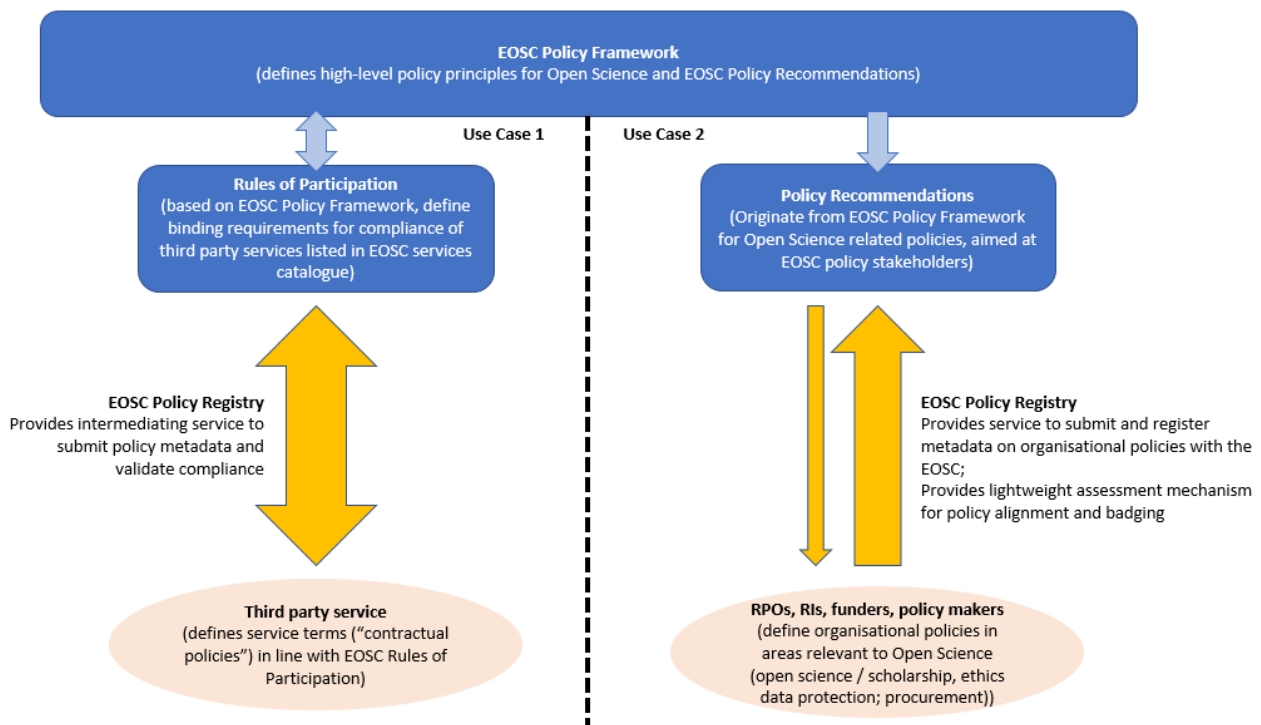


Figure 3: Functions of Policy Registry in different use cases

In use case 1, the Policy Registry intermediates between the Rules of Participation and the service policies of service providers. The objective here is to determine whether a service is “EOSC compatible” and should therefore be included in the EOSC service catalogue, based on the level of compliance with the Rules of Participation. The Policy Registry thus monitors policy developments (by collecting policy metadata from service providers), but also conducts an in-depth assessment to validate whether service providers meet the requirements of the Rules of Participation. Hence, as mentioned previously, the Policy Registry is not just a registry (which captures and stores policy metadata), but also acts as a validation service and gatekeeper for the EOSC Service catalogue.

In use case 2, the Policy Registry intermediates between the EOSC’s Policy Recommendations and relevant organisational policies implemented by RPOs, RIs, funders, and ministries or policy-makers. The primary purpose here is to gather data on policy implementation by stakeholders in order to allow the EOSC to keep track of policy practice. Crucially, this data should allow evaluations of whether – or how – targeted stakeholders implement EOSC policy recommendations. A secondary purpose (indicated by a narrow arrow in figure 3) is to give feedback to stakeholders on their policy implementation efforts through a badging and scoring system which conducts policy assessments based on the EOSC’s Policy Recommendations. Badges are a sign for policy stakeholders with a strong alignment with EOSC’s Policy Recommendations. This badging service is complementary to the badging service of the Open Science Monitor<sup>13</sup>. However, unlike in use case 1, not meeting the policy requirements is not a selection criterion for participation in the EOSC; i.e. RPOs not implementing EOSC’s policy requirements would still be able to use EOSC services. For use case 2, the Policy Registry is thus mainly a registry with an attached lightweight policy assessment service.

### 2.3. Requirements of the Policy Registry

This section summarises the functional and non-functional requirements for the Policy Registry, specifying the functions and qualities the Registry should meet from the perspective of users. Requirements for the

<sup>13</sup> <https://eoscipilot.eu/sites/default/files/eoscipilot-d3.5.pdf>



Policy Registry are defined in three areas: the overall conceptual model, workflow processes (i.e. policy submission/registration, assessment/validation, storage, and production of statistics), and Policy Registry services. These requirements do not differ strongly depending on the user group and use case, therefore no differentiation is made in this respect.

### 2.3.1. Requirements for the Policy Registry model

The purpose of the conceptual model is to serve as an abstract representation of the main components – or concepts – of the Policy Registry. For this, it must capture the relevant entities operating in the system's framework, including users in different roles, different policy types with their functions, and system workflows. The following non-functional requirements are crucial to ensure the Policy Registry can fulfil its functions to a sufficiently high standard:

- **Transparency:** The Policy Registry model needs to operate in a way that makes it easy for third parties to understand which actions the Policy Registry performs why and how. Transparency is a crucial requirement for the Policy Registry as it will shape users' perception of whether the system is fit for purpose. Transparency can be separated into data transparency (i.e. which metadata is used and why?), policy transparency (i.e. why is an action performed and why is certain data needed – which policy does it relate to?), and process transparency (i.e. how is an action performed?). Achieving transparency is a technical task which can be partly embedded in the system design of the Policy Registry, e.g. (in the case of process and policy transparency) by making the Registry's software code open source as well as explaining data processing steps clearly. Additionally, transparency can be achieved through comprehensive documentation and information for users as well as clear explanations why selected policy metadata is required (data transparency). Additionally, transparency is however also a non-technical requirement that must be covered by appropriate governance mechanisms that ensure a transparent translation of EOSC policies into the way these policies are expressed in systems such as the Policy Registry.
- **Auditability:** The model must ensure that the Policy Registry is auditable, i.e. open to systematic examinations which verify the correct function of the system. Adding to transparency, auditability is another crucial requirement to enhance user's trust in the system. In addition to a clear explanation of assessment criteria and open sourcing the Registry's software code, auditability can be ensured by implementing processes which keep versioned files of metadata and document different stages of the metadata submission and assessment process.
- **Coverage:** The model must cover all entities relevant in the full Policy Registry workflow.
- **Expressiveness:** The representation model must be based on abstractions that are sufficiently flexible to express the specific entities relevant to the modelling of relevant policy types. One solution to achieve this is the selection of appropriate metadata entities which define specific restrictions on the metadata level; this avoids the risk of building restrictions into the model's architecture by selecting overly restrictive entities.
- **Extensibility:** The Policy Registry model must be extensible, particularly to make it adaptable to future changes which will likely occur during the evolution of the EOSC's policy and governance framework.
- **Interoperability:** The conceptual entities of the Policy Registry, term vocabularies and code lists to instantiate these must be modelled to facilitate interoperability with related services (i.e. the Open Science Monitor) and external sources, from which the system may draw data (e.g. third-party service catalogues). This can be achieved by reusing appropriate, publicly ontologies as well as cross-mapping metadata at an early stage in the development process.

In addition to the non-functional requirements which establish the quality characteristics of the Policy Registry, a set of functional requirements exist for the model. These ensure an appropriate collection and processing of policy information (captured in metadata) through the system:

- **Actor coverage:** The model's elements must represent all relevant actors who contribute to or are connected to the Policy Registry framework. This includes users that provide input, users that



consume the output, and any indirectly involved actors that provide additional data to the Policy Registry or are in any way affected by the framework.

- **Processes:** The model must define appropriate elements to represent the processes of the Policy Registry. For example, processes of (meta-)data collection, assessing the completeness of data, and perform qualitative evaluation actions must be represented by high-level notions of processing actions.
- **Policy assessment:** The Policy Registry's policy alignment or compliance assessment processes focus on a particular set of policy resources (as defined in section 2.1). The Policy assessment includes the processes of formally checking the completeness of information, assessing its accuracy and whether it meets the policy requirements of the Policy Registry. These Policy Registry model should cover these assessment processes appropriately.
- **Assessment output:** The output of the Policy Registry assessment is a qualitative statement on how well the respective policy under assessment is aligned with the EOSC's own requirements. This evaluation output and its different expressions must be properly represented in the model of the Policy Registry.

### 2.3.2. Workflow and processes of the Policy Registry

The Policy Registry workflow depends on several processes covering the full policy registration and assessment process. As mentioned in section 1, the Policy Registry relies essentially on four macro-functions that define the Registry's process components and together constitute the workflow of the Policy Registry: i) submission and registration of structured policy data; ii) assessing collected policy metadata; iii) storing collected policy data and assessments; iv) producing data for related services such as the Open Science Monitor. From this, a list of functional requirements can be deducted:

- **Process decomposition:** The Policy Registry must enable the separation, specification, and amendment of sub-processes. Particularly the Policy Registry's validation processes might need to be amended, e.g. because EOSC Policies and Rules of Participation change or because monitoring requirements emerging from the Open Science Monitor change. Such changes would likely also have implications for the other processes in the Policy Registry's workflow, such as the collection of policy metadata. This interdependency implies that all process components must be de-composable.
- **Adaptability:** While the requirements for Policy Registry processes might evolve, a process must be able to evolve and change along with the dynamic environment it is defined in. For example, the policy assessment process and criteria in use case 1 might need to be adapted based on revised Rules of Participation. This implies that the metadata submission (and collection) process as well as policy assessment process would need to change.
- **Temporality:** The Registry workflow must include options to define particularly the intervals and time periods for which a registration and assessment process should be conducted. As explained in section 2.1, some policy assessment processes might have to be conducted only once, while others will need to be repeated ad-hoc or in fixed intervals. To satisfy these requirements, the Policy Registry workflow must include temporal qualities.
- **Relating metadata to policies:** The Policy Registry workflow must facilitate a clear relationship between the metadata required by the Policy Registry and the underlying components of the EOSC policy and governance framework. This serves and underpins the adaptability of the Registry workflow, but also enhances the transparency of the Policy Registry model.

### 2.3.3. Policy Registry Services

The Policy Registry will require a set of high-level functionalities, defined as services to allow users to interact with the system, i.e. administer metadata submission and registration, access and manage submitted data, and consume assessment results which are delivered through the Policy Registry's workflow. These services will need to satisfy various non-functional requirements ensuring the overall service quality of the Policy Registry:

- **Availability and reliability:** The Policy Registry services must be highly available and reliable. Availability is defined as the fraction of a time period that an item is, on demand, in a condition to perform its intended function. Availability can thus be expressed as the percentage of a time period that the Registry can execute its processes as requested. Reliability is defined as the ability of a system or its component to function for a specified period of time provided that the conditions which usually ensure its functionality (e.g. energy supply, cooling, etc.) are fulfilled. This means that the Registry services should function with a high degree of reliability as long as the core conditions for the service to operate are fulfilled. Together, availability and reliability are basic quality requirements for services.
- **Accessibility:** Policy Registry services must be web-accessible and lightweight, following an open and transparent approach with regards to the use and documentation of system specifics such as programming languages, frameworks, and the overall architecture. Accessibility helps to maximise the range of service users and can help to ensure trust in them.
- **Multitenancy:** Policy Registry services must be able to host and serve multiple users / tenants, meeting their requirements and specifications. This can be ensured by implementing Policy Registry services according to Software-as-a-Service principles. For this, the Policy Registry services should expose a set of RESTful APIs which cover and connect all relevant service components throughout the Policy Registry workflow.
- **Interoperability:** Interoperability with other policy supporting services (particularly the Open Science Monitor) as well as potential external data sources (e.g. EOSC compliant service catalogues which supply relevant data to the Policy Registry) must be ensured. In addition to the use of RESTful APIs, interoperability can be enabled by using established standards and appropriate web resource synchronization frameworks.
- **Extensibility:** In line with the Policy Registry model and workflow requirements, the Policy Registry services must also be flexible to allow the implementation of extensions or adaptations, particularly if differing requirements emerge from the Policy Registry's policy and governance framework.
- **Scalability:** Since the Policy Registry is designed as a major gatekeeper service to determine the eligibility of third parties to participate in the EOSC, it is expected that the service will need to handle large amounts of requests in a short period of time. Therefore, these services must be implemented in a manner which ensures the maximum scalability of services.
- **Security and data integrity:** Policy Services must be backed by appropriate measures to ensure the system's overall security and data integrity. Such measures should include the definition of a comprehensive security policy, which covers all aspects of the workflow from the authorization and authentication of users to the administration of policy registration processes; responses to external threats such as DDoS attacks must also be considered as part of this. Additionally, data integrity needs to be ensured based on non-repudiation principles, ensuring that proof of record for the origin of any data, proof for the integrity of data, and versioned files of data are kept.

The Policy Registry will need to serve two main users groups, which however pose similar functional requirements: On the one hand, *EOSC stakeholders*, i.e. RPOs, RIs, and funders / policy-makers, who want to assess and prove general policy alignment with the EOSC; on the other hand, *EOSC service providers* who want to validate the eligibility of their services to be included in EOSC service portfolio (with the service status level "EOSC compatible").

The functional service requirements for both user groups are:

- **User registration service:** Users must be able to register with the service, allowing them to submit policy metadata, initiate a policy assessment process, and manage policy metadata on an ongoing basis. The registration service should be geared towards individuals, but connecting them to their respective organisations, ideally based on common email domains.
- **Authentication and authorization service:** The Policy Registry's authentication and authorization service must offer a reliable service for users to authenticate themselves (e.g. as belonging to an organisation) and to prove their authorization to manage the relevant policy metadata.

- **Initiate policy registration and compliance / alignment assessment process:** Probably the core service requirement for the Policy Registry is that it must enable users to register relevant policy metadata and thus to initiate a policy compliance / alignment process. The core element of this function is the submission of relevant policy metadata via an appropriate mechanism, e.g. a metadata form. Since the required policy metadata covers a variety of specific policy information, users should be able to collaborate on submissions and to work through multiple sessions.
- **Save, stop, and delete process:** Once initiated, users must be able to save data which is being prepared for submission. Furthermore, users must be able to control the assessment process by stopping it when appropriate and deleting any related data.
- **Manage submitted data for policy update:** It is projected that policy metadata, once it has been submitted and evaluated, will need to be updated in irregular intervals to reflect for example policy updates on behalf of the submitting user (e.g. when a university changes its privacy policy) or new requirements emerging from updates of EOSC's Policy Recommendations or Rules of Participation. To make sure users do not have to re-enter all policy metadata every time an update is required, they should be able to access policy metadata which they have submitted previously – or transfer metadata from past submissions into the updated metadata scheme (if the update is required because EOSC's policy metadata requirements have changed).
- **Notification on results:** Additionally, users must be able to receive automated notifications on the results of the policy compliance or alignment assessment.
- **Retrieve machine readable data:** Once policy metadata has been submitted, processed and assessed, users must be able to retrieve submitted and processed metadata in open, interoperable, and machine-readable formats, e.g. JSON, CSV and RDF/XML. To achieve the functionality, APIs (or other data retrieval mechanisms) and data conversion services must be in place.

### 3. POLICY REGISTRY MODEL AND ARCHITECTURE

In this section, the conceptual framework for the Policy Registry is presented. This includes the overall model of the Policy Registry, the Registry's architecture, as well as an outline of the proposed data model.

#### 3.1. Conceptual model

The purpose of the Policy Registry conceptual model is to identify and define the conceptual entities relevant for the operations of the Registry. Additionally, the model must also express how the different entities relate to each other. The model is an abstract, static representation of the Policy Registry framework not specifying workflows, but indicating the core components around which workflows need to be designed. The conceptual entities must therefore be flexible enough – and be sufficiently abstract – to accommodate different scenarios and changing implementations of workflows.

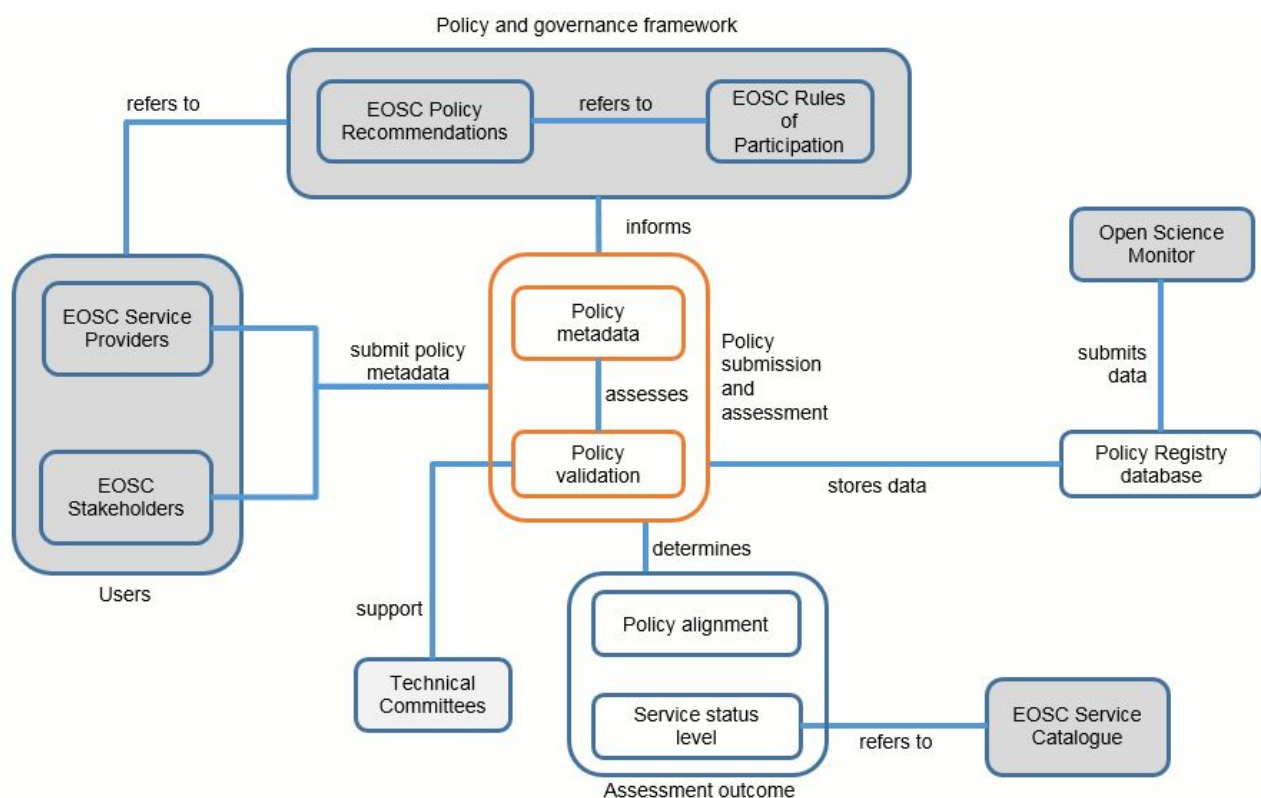


Figure 4: Policy Registry model

Figure 4 shows the Policy Registry model's entities and their relationships. The Policy Registry is interdependent with various other components of the EOSC that are not developed as part of the EOSC. These components, which are relevant, but "external" components of the Registry model, are represented in light grey boxes. The Registry model's entities are:

- **Policy submission and assessment:** Located at the centre of the Policy Registry is the concept of the policy submission and assessment process. The concept denotes the core function of the Policy Registry, i.e. to enable the collection of policy metadata as well as its assessment through a policy validation mechanism. This requires that the component comprehensively covers processes required to compute an assessment outcome, i.e. data collection, parsing, and validation. The collection of policy metadata is facilitated through a structured metadata form, while the policy validation function depends on a semi-automated assessment of the submitted policy metadata against a set of policy criteria, which are informed and defined by the Rules of Participation and Policy Recommendations. To be able to accommodate changing requirements through the Registry, the

policy submission and assessment process is designed to be extensible. Additionally, the component needs to operate based on a clear definition of the underlying data model as well as validation processes to ensure that processes are auditable, and transparent.

- **Policy and governance framework:** The policy and governance framework consists of the EOSC Policy Recommendations and Rules of Participation. Together, these two components provide the set of conditions users must meet if they seek policy alignment with the EOSC Policy Recommendations (in the case of stakeholders) or policy compliance with the Rules of Participation (in the case of service providers). As mentioned previously, the Policy Recommendations and Rules of Participation are partly interdependent, referring to each other. The policy and governance framework is a highly important, but external component for the Policy Registry.
- **Users:** The concept of users refers to the two user groups of the Policy Registry, i.e. service providers and EOSC stakeholders. The underlying assumption is that these users will specify their respective policies in line with the EOSC policy requirements emerging from the Policy Recommendations and Rules of Participation. Once formulated, users submit metadata on their respective policies to the policy submission and assessment component with the expectation to receive an assessment outcome.
- **Assessment outcome:** The assessment outcome component is a component to indicate the result of the policy validation process. Assessment outcomes come in two different forms: An assessment outcome on *policy alignment* is the result of an assessment according to use case 2, i.e. where EOSC stakeholders submit metadata on their organisational policies, which is then assessed in light of the Policy Recommendations. An assessment on the *service status* level “EOSC compatible” is the outcome for an assessment according to use case 1, i.e. where EOSC service providers submit policy metadata in order to assess whether policies are in line with the Rules of Participation.
- **EOSC Service Catalogue:** Like the policy and governance framework, the EOSC Service Catalogue is an external component of the Policy Registry model. The EOSC Service Catalogue references service status levels and specifies different service classes, which need to be considered and incorporated in the Policy Registry. Vice versa, the assessment outcome should provide information on whether a service, based on its policies (e.g. formulated in terms and conditions) can be included in the EOSC service catalogue.
- **Policy Registry database:** The Policy Registry database stores the submitted policy metadata as well as data on the assessment outcome. The concept thus refers to the relational database back-end, which is used to preserve any data which has been submitted through the policy submission and assessment mechanism. The Policy Registry provides its data to the Open Science Monitor, in order to enhance its statistics on policy implementation practice by different open science practitioners.
- **Open Science Monitor:** The Open Science Monitor (D3.2) is a statistical monitoring framework to gather metrics on the implementation of Open Science practices and policies. It is an external component to the Registry and is part of the policy-supporting services suite.
- **Technical Committees:** According to the EOSC’s service portfolio (D5.2) the EOSC technical committees are teams of experts which would be responsible for assigning and monitoring service quality within EOSC. In this role, the EOSC Technical Committees should also support the elements of the policy validation process that cannot be automated, because they require an in-depth, human interpretation of whether EOSC policy standards have been met. Therefore, the Technical Committees are a crucial external component to support the policy validation process.

### 3.2. Architecture

Based on the functional process and service requirements presented in section 2.3, this section explains the proposed architecture of the Policy Registry. The components and modules together present a general framework, however more detailed specifications will need to be developed in the implementation phase of the Policy Registry.



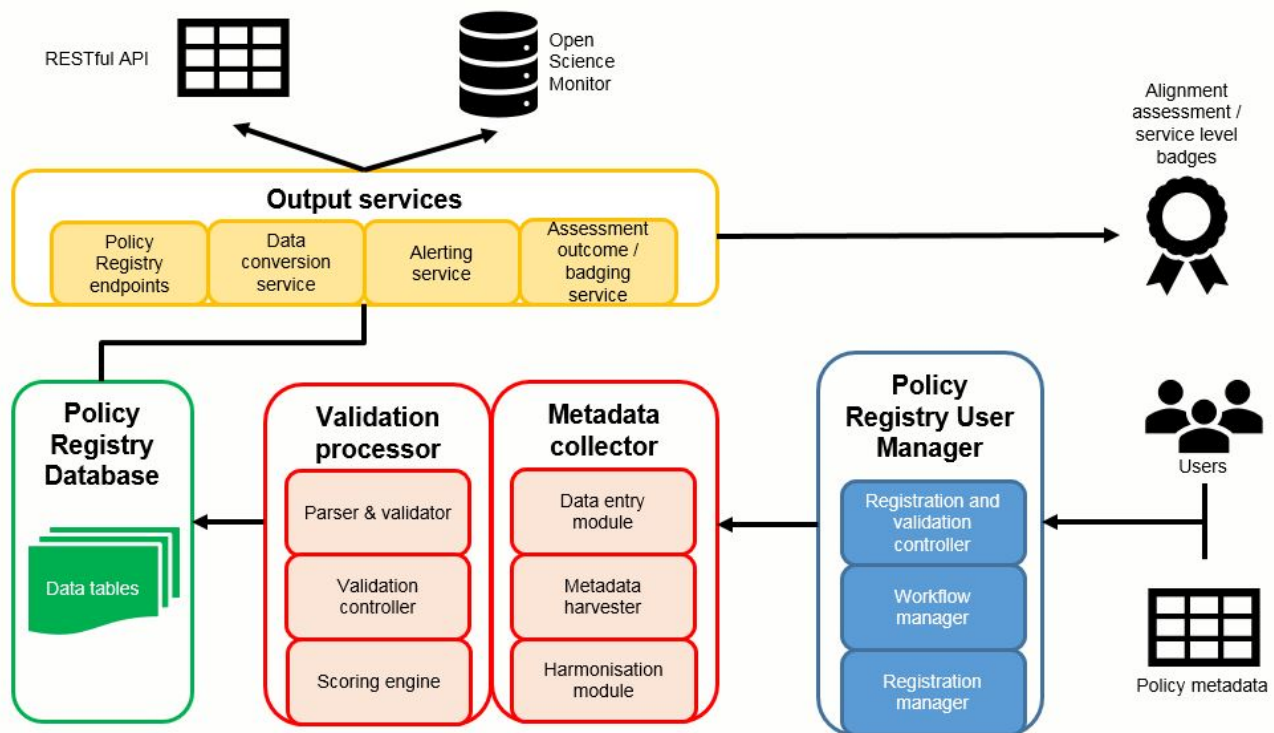


Figure 5: Policy Registry architecture

The Policy Registry is composed of four components that support the generic workflow of the Policy Registry. The three main components are presented at the bottom.

The first component is the **Policy Registry User Manager** component, which provides the user-oriented front-end of the Policy Registry and is thus designed to support users in the policy registration / submission processes. The component builds on three main components:

- **Registration manager:** This component supports the general registration process with the Policy Registry, covering both the functions to register users to be able to submit policies as well as the specific registration process for policies, depending on their use case.
- **Workflow manager:** The workflow manager supports users in managing the overall workflow of the policy registration / submission and validation process. Through the workflow manager, users can initiate, stop, and delete a policy registration process. Additionally, this component should allow users to amend ongoing validation processes with additional information and repeat registration processes based on previously submitted data, e.g. if the organisational policies of a RPO have changed or if changing EOSC policy requirements demand the re-submission of certain policy data. Furthermore, the workflow manager allows users to work collaboratively on the submission of policy metadata, e.g. when multiple co-workers from one organisation collaborate to submit policy metadata.
- **Registration and validation controller:** This component controls the registration process and provides users with feedback on the initial (automatic) assessment of submitted policy metadata, e.g. to control whether mandatory metadata is complete and provided in the correct schema. The Registration and validation controller therefore integrates with the validation processor and thus provides a front-end for pre-final feedback on submitted metadata.

The **Metadata Collector** is the second component, acting as the back-end service to facilitate data collection. It consists of:

- **Data entry module:** The data collector is the back-end to facilitate the submission of manually curated and harvested policy metadata. The data collector is a dynamic metadata form through which users submit policy metadata relevant to their use case. The form is similar to metadata

submission forms e.g. used by ROARMAP and adapts to the specific use case. Interacting with the Data Harvester and Harmonisation Module, metadata can be prefilled with harmonised metadata from other resources.

- **Metadata harvester:** In some instances of use case 1, service providers will already have listed information about their services in EOSC compliant service catalogues. In such cases, the Policy Registry will need to be able to collect relevant policy metadata from external service catalogues (e.g. the eInfraCentral service catalogue<sup>14</sup>). In these instances, the data harvester will provide an automated mechanism (e.g. API queries) to retrieve relevant data from applicable service catalogue endpoints.
- **Harmonisation Module:** Where data is not manually curated as input through the Policy Registry's metadata form, but is automatically harvested from external data sources, limited data cleansing and harmonisation processes will need to be performed. The Harmonization Module is used to handle these tasks ensuring that all ingested data follow a common conceptual model and structure.

The third architectural component of the Policy Registry is the **Validation Processor**. The purpose of the validation processor is to support the data validation and assessment process. The following service modules ensure this overall function:

- **Parser and validator:** For the policy metadata fields based on controlled values or otherwise machine-readable data (e.g. URIs), the data parser and validator conducts the analysis of the submitted data. The activities performed include parsing the data, assessing/confirming the machine-readability of data, and validating that the registered data meets the required structure and schema for the validation of policy alignment (use case 2) or compliance (use case 1). The data parser and validator also integrates directly with the Registration and validation controller to provide immediate feedback on whether submitted data meets the programmatic metadata requirements.
- **Validation controller:** The validation controller provides the function for administrators of the Policy Registry to control the validation process and amend, if required, the assessment outcomes of automated validation processes. Importantly, the validation controller also assists in the assessment of non-machine-readable policy metadata. These metadata fields and any supporting information need to be assessed manually by the EOSC Technical Committees. Through the validation controller, these Committees would also be able to enter metadata on the results of manual policy assessments into the Policy Registry's database.
- **Scoring engine:** Based on the assessment outcomes for individual metadata fields, the Policy Registry's scoring engine conducts the aggregation and computing of data in order to arrive at an overall assessment outcome.

Once an assessment has been conducted, the submitted policy metadata (subject to amendments) together with additional metadata produced through the assessment process to indicate the assessment outcome is stored in the **Policy Registry Database**. This database is the third component of the Policy Registry and stores all policy metadata in versioned data tables.

Lastly, the policy Registry Database interacts with the last component of the EOSC architecture, the **Output Services**, through a RESTful API. The Output Service's purpose is to communicate assessment outcomes to relevant users and other EOSC components, as well as to ensure that data stored in the Registry database can be accessed in a programmatic manner. It is supported by four modules:

- **Assessment outcome and badging service:** Once a policy has been validated and an assessment outcome has been computed, this service provides a summary of the outcome to the submitting users. For this, the assessment outcome and badging service queries the data stored in the database and, based on the outcome data, provides a summary on the achieved level of policy compliance (use case 1) or policy alignment (use case 2). The service also provides badges which users can

<sup>14</sup> <http://einfracentral.eu/basic-page/common-service-catalogue>



incorporate in their organisational websites to indicate their achievement. The assessment outcome and badging service directly interacts with the user registration manager for this process.

- **Alerting service:** The alerting service provides the facilities to alert users once their assessments have been completed.
- **Policy Registry Endpoints:** The Output Services expose a RESTful API which makes the Policy Registry Database accessible to other EOSC services such as the Open Science Monitor or other users who want to access the Policy Registry database. The query processor intermediates data requests from such external users and thus ensures that the correct data tables are provided on request.
- **Data Conversion Service:** Policy metadata stored in the Registry's database should be accessible in different open formats such as JSON, CSV and RDF/XML. The data conversion service ensures that data, once queried, can also be translated into the different formats required by data consumers.

### 3.3. Data Model

The data model is a core component to support the Policy Registry's assessment and validation function. It needs to capture relevant elements of EOSC policies and translate them into metadata entities which can facilitate semi-automated decisions on policy compliance. The metadata entities and properties presented in this section have been deducted from EOSCpilot's Rules of Participation (D2.5) and Draft Policy Recommendations (D3.3). Additionally, in order to facilitate the machine-readability of policy metadata, the data has to model these aspects where possible into a structured vocabulary which is amenable to machine-driven processing and assessment, e.g. in the form of controlled values for individual metadata fields. A second challenge is that while the Policy Registry operates with two separate use cases, the relevant metadata properties and entities should be captured in just *one* data model. Because of the overlap between policy concepts in the EOSC draft Policy Recommendations and Rules of Participation, the same policy metadata can indeed be relevant for the policy assessments in both use cases. Therefore, as already alluded to in the discussion of the expressiveness requirement (in relation to the Registry Model), the main challenge here is to specify the relevant metadata as a meaningful, but flexible common denominator for both use cases. In other words, metadata fields and descriptions should – where possible - be sufficiently broad to accommodate both use cases, but also narrow enough to allow a meaningful assessment of the submitted data for each use case.

Figure 6 shows the metadata properties of the conceptual data model for the Policy Registry. The metadata properties are classed into two categories: core policy metadata (light blue) and complementary metadata (dark blue). The core policy metadata entities retain the material information on various policy aspects in a three-tiered structure. This includes general policy information at the top level, domain-specific policy information on ethics, procurement, data protection, and Open Science on the second layer, as well as further metadata relating to the previous policy domains on the third layer. The complementary entities contain metadata to record information on the resources to which the policy metadata relates. This includes metadata on the related OS resources (software, publications, datasets, services), the submitting stakeholder, the service and data catalogues, and assessment outcome.

By separating policy metadata from complementary, resource-specific metadata, the data model offers the flexibility to register one policy applying to multiple resources. This allows to cover the – likely - frequent scenario where providers of OS resources, particularly service and data providers, have one standard policy which applies to several offerings. In these cases, users of the Policy Registry would have to enter the relevant policy metadata only once, allowing them to just refer to the assigned policy ID in all subsequent registration cases.

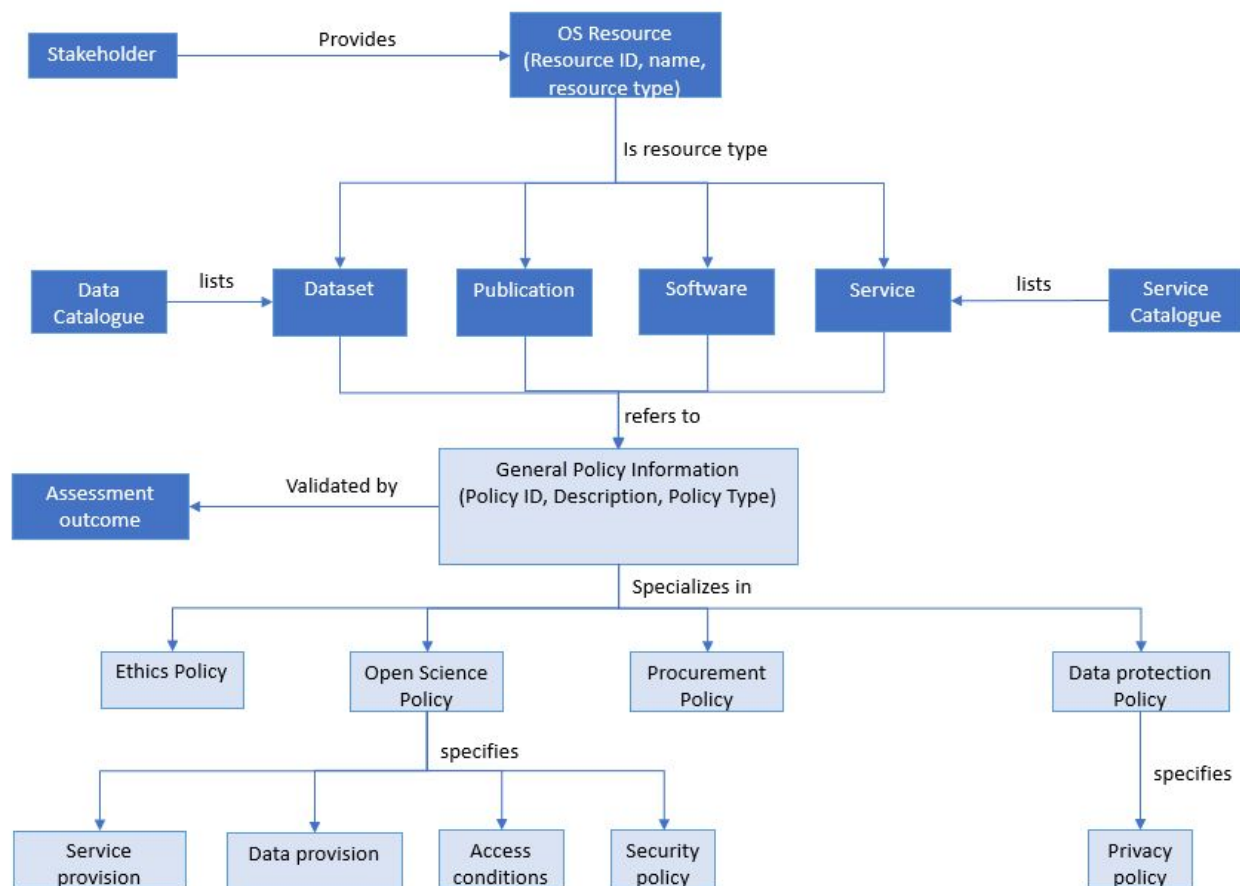


Figure 6: Policy Registry Data model

The metadata entities have been developed in a bottom-up process based on the Rules of Participation and draft Policy Recommendations. The entity properties emerging from the Rules of Participation have been mapped with the Rules of Participation from which they originate<sup>15</sup>. Additionally, the metadata entities use controlled values where possible to facilitate the machine-readability of policy metadata. Where possible, metadata properties and controlled vocabularies have also been cross-mapped to match applicable properties of the eInfraCentral data schema.<sup>16</sup> The motivation for this is twofold: First, the eInfraCentral data schema models metadata on services, including some aspects which are also relevant to the policy validation conducted by the Policy Registry. Second, because the eInfraCentral service catalogue is planned to operate as an EOSC-compatible service catalogue, its metadata schema needs to be cross-mapped to facilitate automated metadata harvesting by the Policy Registry. Both the development of controlled vocabularies and cross-mapping with other data schemas such as eInfraCentral are ongoing work due to the evolving refinements of the Policy Recommendations and Rules of Participation. Hence, concerning its machine-readability and interoperability, this metadata model is a starting point, which however should be revised, particularly at the implementation stage.

With regards to first two layers of the core policy metadata, the following metadata entities apply:

- **General policy information** This metadata entity contains general information that must be submitted for each case. It includes the title and a short description for each policy as well as an automatically generated persistent identifier for each policy. Additionally, it contains a categorisation of the registered policy type. If the registered policy type are terms and conditions (i.e. a contract) or a service level agreement, submitting users are also requested to provide a URL of these as a copy.

<sup>15</sup> For further information on the mapping, please refer to Annex A. The eInfraCentral data schema is described in this report:

<sup>16</sup> For a full documentation of the conceptual data model, included detailed descriptions of entities as well as mappings to the Rules of Participation and eInfraCentral data schema entities, please see Annex A.

- **Open Science, ethics, procurement, and data protection policy:** The second layer of policy metadata contains further information on domain-specific policy aspects which have been derived mainly from the EOSCpilot draft Policy Recommendations. The latter are structured into four policy domains: Open Science, ethics, procurement, and data protection. Accordingly, users are required to provide information on whether they have adopted general policies for Open Science, ethics, procurement, and data protection. For each applicable policy type, users are required to specify whether such a policy exists, when it was adopted, provide a summary of the policy, and copy the URL where the policy can be accessed.

The entities of the third policy metadata layer have been deduced mainly from the EOSCpilot Rules of Participation. As depicted in figure 6, they can be subsumed under the second-layer policy entities which they also elaborate further. In future iterations of the data model, particularly this third entity layer may have to be further refined as more specific policy requirements are specified through the Policy Recommendations and Rules of Participation:

- **Service provision:** This metadata entity contains properties to cover the policy requirement that service providers give sufficient information on *how* a service is provided. This entity requires metadata on the availability of services, provision quality and quality certification, portability of services, as well as APIs, standards, and protocols used. The metadata entity is thus directly relevant to the provision of such documentary information as laid out in the Rules of Participation.
- **Data provision:** The data provision entity contains attributes to describe some of the conditions that define *how* research data is provided. This includes high-level information on the applicable data curation and preservation policy, data portability, and processes applied to ensure that provided data is FAIR. The FAIRness of data and the adoption of data curation and preservation policies feature as policy principles both in the Policy Recommendations and Rules of Participation.
- **Access conditions:** The access conditions entity contains metadata fields to capture information on the description of how access to a service or other open science resource is provided (as described in the EOSC Rules of Participation and Policy Recommendations). This contains structured information on the intended users of a resource, access policy and rights, how access is determined (e.g. free / open, fee-based, excellence-based), capacity limits, fees, and (if applicable) information on peer-review processes to determine excellence-based access. While most of the included metadata properties have been emulated based on the Rules of Participation, they pose more general requirements for any stakeholder to be transparent about who and how open science resources can be accessed.
- **Security policy:** The security policy entity covers the requirement, as mentioned in the EOSCpilot Rules of Participation for service providers, to declare their security policy. The metadata entity requires particularly service providers who follow use case 1 to specify whether they have a security policy in place, as well as to register a summary statement of the policy, its adoption date, and a URL for the policy.
- **Privacy policy:** End-users' privacy is a core concern reflected in the Rules of Participation and Policy Recommendations. The privacy policy entity thus refers to a set of policy requirements for service providers and stakeholders to document their data protection statement, information on GDPR support, and user data that is collected during the provision of a service. These attributes refer to compliance and alignment requirements which are both referenced in the EOSC Policy Recommendations and Rules of Participation. The underlying information on data protection is thus relevant to assessments in use case 1 and 2.

In addition to the core policy metadata, complementary metadata needs to be registered in order to ensure that the Policy Registry can operate as desired. This concerns metadata about the resource to which a policy applies, the submitting stakeholder, and the assessment outcome metadata:

- **OS Resource and resource-specific metadata:** The OS Resource entity contains high-level information on the type of resource for which a policy is being registered. This includes the resource name, type, URL, and provider name. Furthermore, for each resource a persistent identifier needs to

be assigned. Depending on the resource type, further metadata needs to be recorded, including summary descriptions for software, publications, and datasets. Based on the Rules of Participation, more detailed information is required for services, including a short description and categorization of functions, as well as details on the service's maturity and operations.

- **Stakeholder:** The stakeholder entity serves a primarily archival purpose and ensures that each policy metadata submission also contains basic information on the submitting stakeholder. This includes entities to record the stakeholder type, name, URL, and contact information.
- **Service catalogue:** The service catalogue entity also serves an archival function, requiring that for services it is recorded whether the service has been previously listed in a separate, EOSC-compliant service catalogue. This includes information on the service catalogue type, name, and URL. As will be later explored in section 4.2, the EOSC Rules of Participation include a provision that service providers cannot only store their data in the EOSC's own service catalogue, but also in other, compliant service catalogues. Where this case applies, this metadata entity thus serves the purpose to register basic information on where else a service has been listed previously.
- **Data catalogue:** Very similar to the service catalogue entity, the data catalogue entity serves primarily an archival function, too. It requires to record if a dataset has been stored in an external data catalogue, which is compatible with the EOSC's data catalogue. This includes information on the data catalogue type, name, and URL.
- **Assessment outcome:** Based on the previously cited metadata entities, the assessment outcome entity records metadata about the achieved policy compliance or alignment. It therefore records metadata on the achieved service status level (i.e. "EOSC compatible") and policy alignment in the four policy areas of the Policy Recommendations (Open Science, Data Protection, Procurement, and Ethics). Other than the previously cited metadata entities, the assessment outcome entity is not completed by the submitting users, but by the Policy Registry as a result of the policy assessment process. For this assessment process, all previously cited metadata entities – except the service catalogue entity – are taken into consideration.

## 4. POLICY REGISTRY WORKFLOWS AND DECISION FLOWS

To describe the operations of the policy registry, the previously described components need to be assembled into a uniform execution workflow. This section describes the general workflow of the Policy Registry and explains the more detailed workflow sequencing and decision flows for the different use cases.

### 4.1. General workflow

Figure 7 is an abstract representation of the overall Policy Registry workflow. It compiles the previously defined components of the Registry architecture into a sequence of four functional sub-processes: the registration and submission of policy metadata; the assessment and validation of the submitted metadata; the storage and archiving of metadata; and the external provision of metadata.

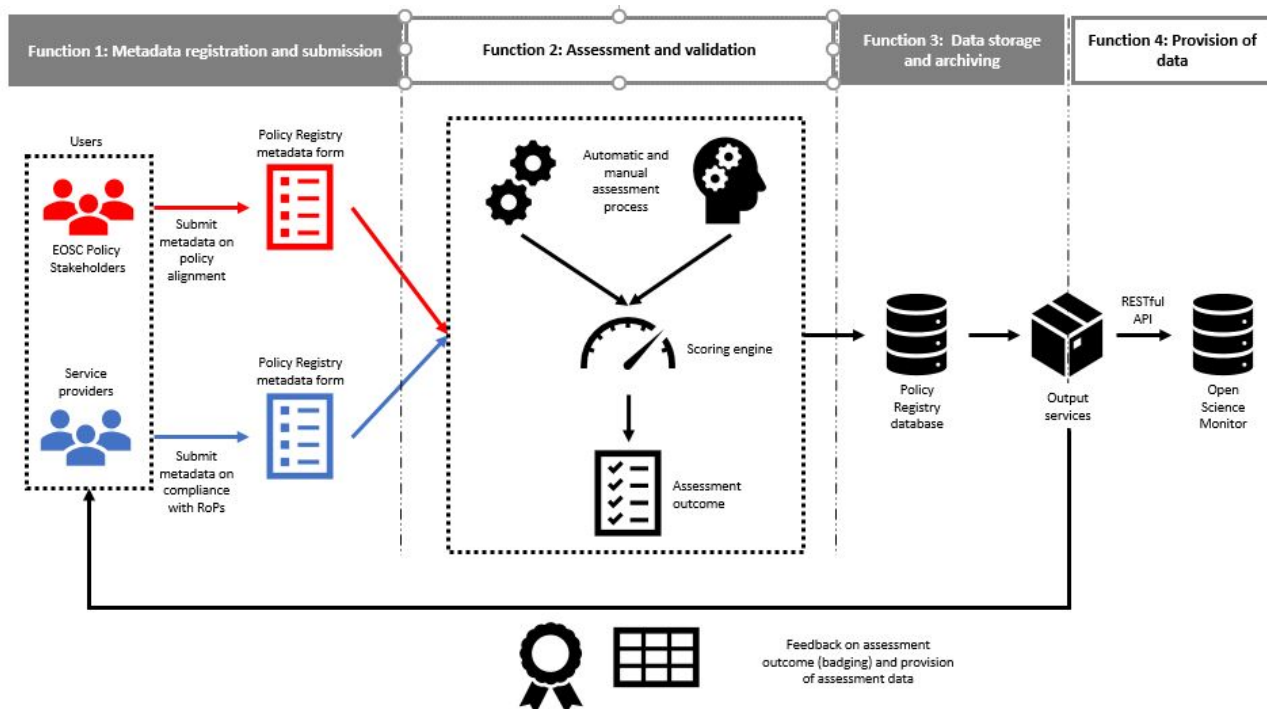


Figure 7: Policy Registry functions and process workflow

The functional sub-processes involve the following characteristics and tasks:

- **Metadata input and submission:** The first step is initiated by the users of the Policy Registry, who provide metadata input relevant to their use case. For this, service providers and other policy stakeholders are presented with two different front-end metadata forms to enter policy-related information relevant to the assessment process. The contents of the metadata-forms reflect the use-case-specific requirements and recommendations discussed in section 3.3. The metadata form is directly linked to the metadata collector component, which provides the back-end of the metadata form, stores the ingested data in a machine-readable, open format, and furthermore has the capacity to automatically harvest metadata from EOsc-compliant service catalogues.
- **Assessment and validation:** The second step is the assessment and validation process, which operates based on the Registry's Validation Processor module. The assessment and validation process consists of four different sub-processes. First, machine-readable metadata fields are checked for compliance with the Registry's controlled vocabulary requirements; in the case of provided URL and PIDs, the validation controller also conducts a cURL-request to check whether the provided resources exist and are live. Second, the EOsc's Technical Committees conduct manual compliance checks to assess and validate information provided in free-text metadata fields (which is thus not machine-readable). Third, based on the automated and manual assessments, the scoring engine will

compute the overall policy alignment or compliance score. Based on this, the assessment outcome metadata is produced as the last step.

- **Data storage and archiving:** In the third step, the resulting metadata entry, including the newly added metadata on assessment outcomes is stored in the relational Policy Registry Database.
- **Provision of output:** The final sub-process of the Policy Registry workflow is to provide output to different addressees. This includes the provision of assessment output data and, if applicable badges, to the users who initiated the assessment process as well as the provision of raw data for further processing to the Open Science Monitor. For this, the Registry's output service module directly interacts with the Registry database and queries data either to forward data in bulk to the Open Science Monitor or to create reports on assessment outcomes for service providers and EOSC policy stakeholders. These reports include both a copy of the assessed policy metadata as well as badges, where applicable.

The principal workflow layout is relatively similar for both use cases from a high-level functional perspective. Nevertheless, the separate use cases require more detailed, slightly deviating workflows to govern the differing assessment and validation processes. The following sections will therefore explore the action and decision flows for both use case. The purpose of the action flow is to define which actions need to be performed in each use case to facilitate the policy assessment process; in other words, what needs to be done by whom? The purpose of the decision flow is to map the questions – or “considerations” – to be taken into account to arrive at an assessment for each use case.

## 4.2. Use case 1 action and decision flow

Figure 8 shows the action flow for a policy compliance assessment under use case 1. In this use case, a service provider seeks an assessment which validates its compliance with the EOSC's Rules of Participation. As introduced in the EOSCpilot reports on the Service Portfolio and Service Management, two service status levels differentiate different levels of policy compliance: EOSC “compatible” services meet the basic policy requirements laid out in Rules of Participation, i.e. they provide all required descriptions of their service terms and conditions and also provide related policy information such as on data protection. EOSC “compliant” services need to meet more stringent requirements on the operational provision of their services, including e.g. the ongoing monitoring of their performance against indicators in service level agreements.

The Policy Registry is not tasked with “EOSC compliant” assessments. However, different EOSC service types have different status level requirements: EOSC core services and EOSC supported services must always be “compliant”, while EOSC service components and End-user services can opt to be either compatible or compliant.<sup>17</sup> Accordingly, for EOSC core and supported services, the compatibility assessment of the Registry

<sup>17</sup> For a discussion of the different EOSC service types, please see the EOSC Service Portfolio report (D5.2): <https://eoscipilot.eu/sites/default/files/eoscipilot-d5.2.pdf>



is only a first step which must be followed up by a separate, ongoing assessment indicate their status as EOSC “compliant”.

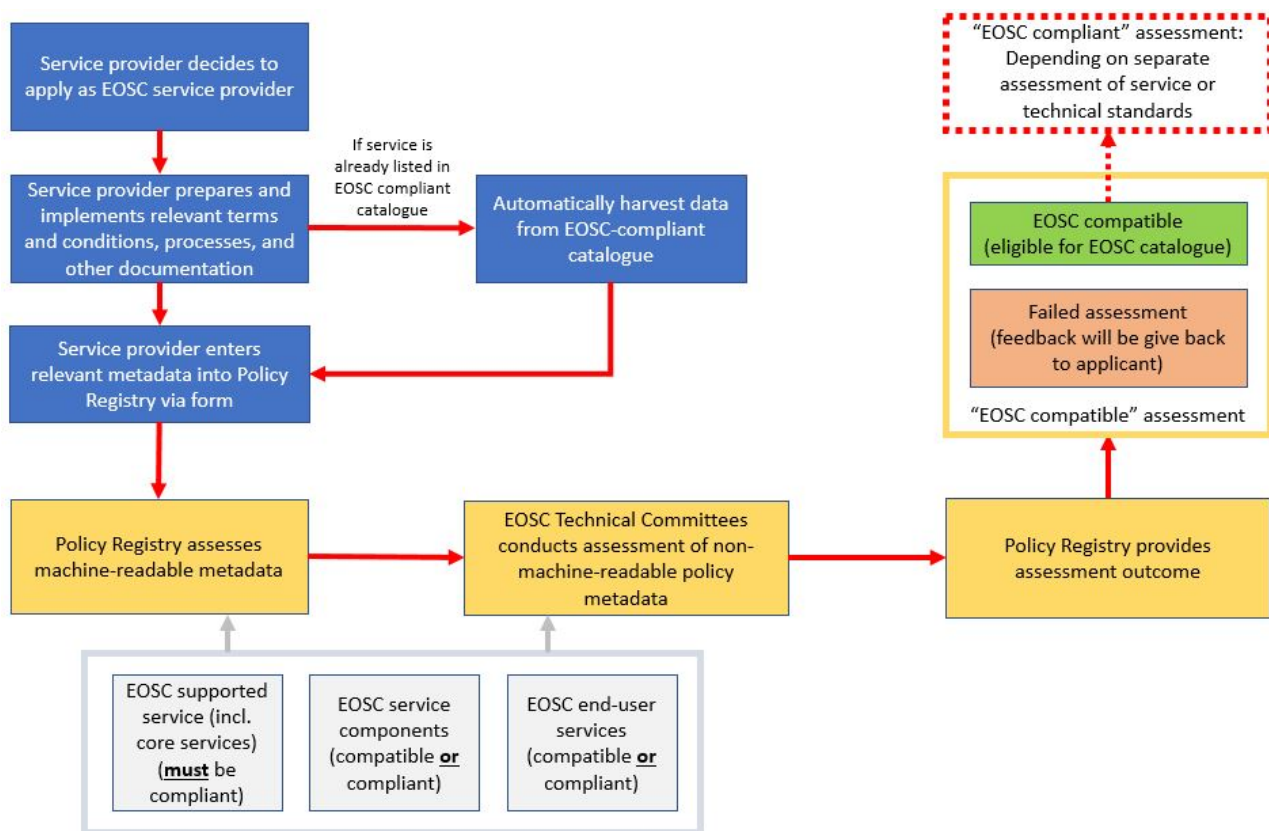


Figure 8: Use case 1 workflow

As depicted in figure 8, the action workflow for use case 1 begins with the decision of a service provider to apply as an EOSC service provider, including a self-assessment on their intended service type<sup>18</sup>. If these do not already exist, the service provider will prepare the relevant policies as well as processes and other documentation to support the service’s compliance with the EOSC Rules of Participation. The service provider then enters relevant policy metadata into the EOSC policy metadata form. If the service is already listed in an EOSC compliant service catalogue, some of the relevant metadata can be harvested from the catalogue to prepopulate the metadata entry.

Based on the provided data, the Policy Registry runs a first round of automatic metadata assessments. This includes the previously described automatic validations of compliance with the Policy Registry’s schematic metadata requirements as well as cURL checks. In a second assessment step, the EOSC Technical Committees conduct a manual policy compliance assessment of the non-machine-readable policy metadata components. Based on both assessments, the Policy Registry provides the assessment outcome “EOSC compatible” or “Failed assessment”. EOSC services seeking mandatory – or voluntary – compliance with the EOSC are then referred further to the separate process to assess whether they are “EOSC compliant”.

Figure 9 depicts the decision flow for use case 1. Based on the structure of the Rules of Participation and the structure of the action flow, the decision flow is relatively simple. The first consideration relevant for the Policy Registry is whether the service has been previously logged in an EOSC compliant service catalogue. If

<sup>18</sup> The suitability of this initial service type indication by the service provider will be assessed and confirmed as part of the policy assessment process.



this is the case, the relevant metadata can be at least partly harvested from this catalogue. The core of the policy assessment process then condenses into just one process step to assess all relevant metadata (top left box) and make the decision to decide whether the required metadata is complete and sufficient (diamond-shaped box).

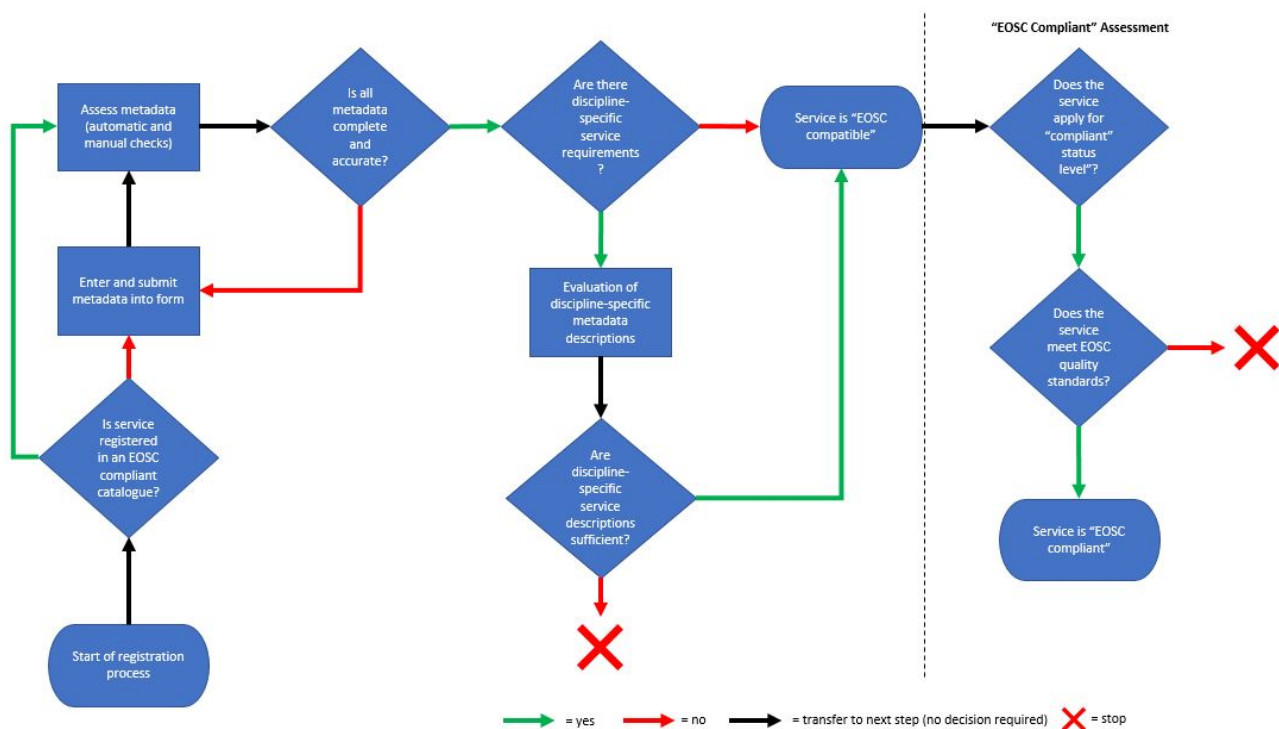


Figure 9: Use case 1 decision flow

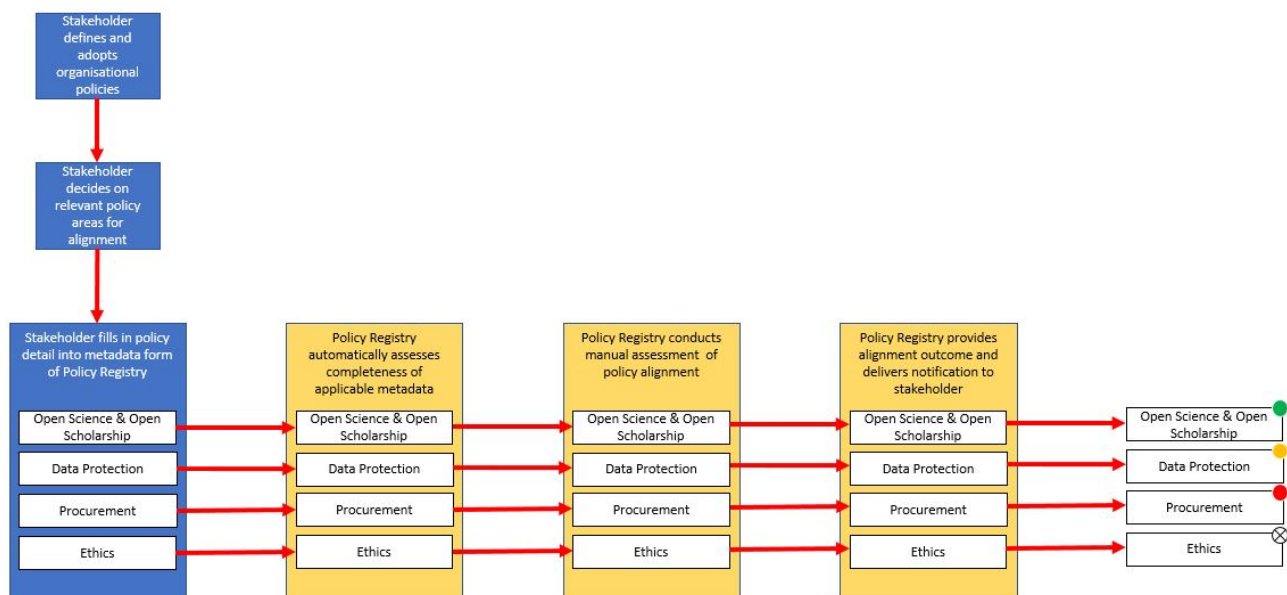
This simplicity in the modelled workflow is a direct result of the logical structure of the initial Rules of Participation, which present all compliance requirements as equally ranking with an equivalent status. Between the different concepts, there is thus no hierarchical or sequential order, which would lead to a more complex decision workflow. The only exception to this is the notion of domain- or discipline-specific service requirements which might apply in some services according to EOSCpilot Rule of Participation 2.<sup>19</sup> As depicted in the workflow, the presence of such requirements triggers a secondary evaluation process to determine whether these discipline-specific requirements have been described sufficiently. However, because the exact specifications for discipline-specific service descriptions are still unclear, this secondary evaluation loop might have to be refined or changed in later iterations.

If the primary evaluation of general policy metadata and the secondary evaluation of domain-specific metadata meet the policy requirements, the service is classed as EOSC compatible. As shown in figure 9, this termination point is at the same time the initiation point for the “EOSC compliant” assessment. This separate assessment process will need to assess whether services under consideration meet specific operational quality and technical standards. However, the exact shape of this process goes beyond of the scope of this report and is therefore represented only schematically in figure 9.

<sup>19</sup> In the health and life sciences, an example for this could be data storage services for clinical trials data. These would likely have more stringent requirements regarding data protection and – possibly – ethics policies. Such requirements would need to be reflected in the service’s policies and submitted policy metadata.

### 4.3. Use case 2 action and decision flow

Figure 10 shows the action workflow for use case 2. In principle, the tasks involved in this workflow are very similar to the action workflow of use case 1. After the stakeholder has decided which policies to submit, the related policy metadata is entered into the Registry's policy metadata form. The action flow of use case 2 then executes the same sub-processes of automatic compliance assessment, manual assessment of the policy alignment, and the provision of the assessment outcome. The main structural difference is that the policy assessment of use case 1 is structured into four parallel workstreams, based on the Policy Recommendation's policy areas: Open Science & Open Scholarship, Data Protection, Procurement, and Ethics. Because these domains exist in parallel, i.e. with no strong hierarchical or logical order between them, the Policy Registry strictly speaking conducts four separate assessment processes, leading to – potentially – separate assessment outcomes in each policy area. Since the Policy Registry is intended to be usable by a range of very diverse users (RPOs, RIs, and funders / policy makers), this parallel structure increases the flexibility of the policy assessment and allows users to submit no information for policy domains not relevant to them. Additionally, the parallel assessment structure also reflects the logical structure of the current draft Policy Recommendations, which do not enforce a hierarchy between policy domains.



**Figure 10: Use case 2 workflow**

The four-streamed assessment workflow is also reflected in the decision diagram for use case 2 (figure 11). The decision flow also shares most of the features of use case 1, with the difference that the Registry will have to consider on which policies metadata has been submitted before conducting the automatic and manual assessment. This assessment is structured into the automated compliance assessment of metadata (i.e. technical compliance assessments of the used vocabulary and cURL tests) and a manual component, which investigates submitted policies in more detail to fully confirm their alignment with EOSC's Policy Recommendations. Based on these assessments, the Policy Registry then provides the assessment outcome, including badges (for integration on websites etc.) to the submitting users. This qualitative measure is structured into three alignment levels: *full*, *partly*, and *no alignment*. In cases where no data has been submitted on a policy domain, or where supplied metadata was insufficient to achieve an assessment, the outcome is *no assessment possible*.

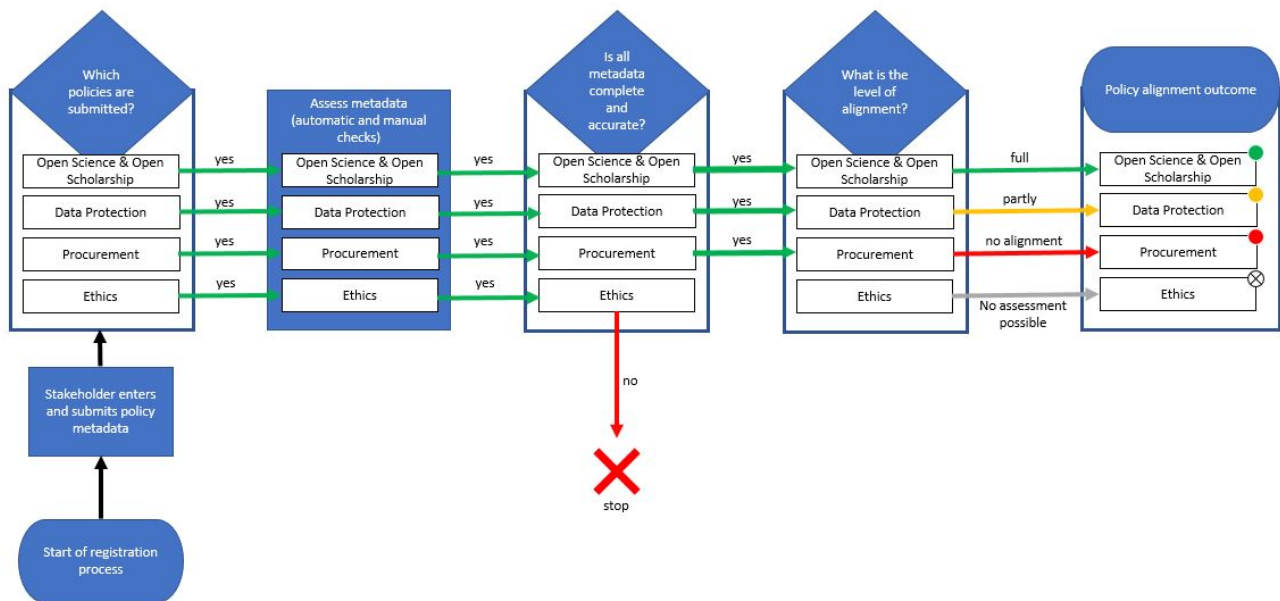


Figure 11: Use case 2 decision workflow

## 5. DISCUSSION AND FUTURE WORK

The previous sections have presented the general framework for the Open Science Policy Registry as a service to (semi-)automate the monitoring and assessment of compliance with EOSC policies. To implement these specifications, various interdependencies must be addressed and resolved in future work. They are discussed in this section.

### 5.1. Interdependencies

To arrive at the implementation of this framework, a series of detailed, interdependent operational issues must be addressed and resolved. Examples for this are the questions of how the policy assessment process would operate in detail, its depth of (manual) policy assessments, and detailed specifications for data base queries. As the Policy Registry and many of its interconnected components are still at the conceptual stage, listing exact implementation questions is currently a speculative endeavour. Nevertheless, the following broad areas of interdependence can be singled out:

**Further development of the metadata framework:** As highlighted in section 3.3, the Policy Registry's data model interlinks closely with EOSCpilot's Rules of Participation and Policy Recommendations. Therefore, the initial conceptual data model presented in this report might have to be revised at the implementation stage to reflect the most recent version of both. Additionally, once they have reached a sufficient maturity, propositions such as the Rules of Participation currently being developed by the EC's High Level Expert Group<sup>20</sup> will need to be taken into consideration.

**Standardisation and automation of policies:** How the Rules of Participation and Policy Recommendations evolve over time will also affect the degree of automation, which can be achieved by the Policy Registry. As was explained in section 2.2, the policy compliance model currently proposed by EOSCpilot gives EOSC stakeholders and service providers a high degree of flexibility in formulating their policies to align – or comply – with EOSC's policy requirements. Particularly at the inception stage, this helps to ensure that the EOSC's policy and governance approach is lightweight and inclusive. Nevertheless, the trade-off of this flexible approach is that it hampers further automation of policy assessments. To achieve this, the Policy Registry relies on a more extensive use of controlled vocabularies in its model. To expand the use of controlled values, however, the underlying policies would need to be further standardised. Appreciating that this is eventually a governance decision, multiple options for policy standardisation could also enable higher policy automation in the context of both the Rules of Participation and Policy Recommendations: In a "full standardisation" scenario, the EOSC would be using a set of uniform, standard policies which EOSC service providers and stakeholders *must* use. In the scenario of "policy modularisation", the EOSC would develop a range of standardised policy modules (each covering some policy aspects) which service providers could use to assemble their "custom" policy. In the "binding commentary" scenario, appropriate EOSC (governance) bodies would create a set of binding, quasi-legal commentaries to interpret the Rules of Participation and Policy Recommendations. These commentaries would define clearly and in detail how the Rules of Participation and Policy Recommendations can – and cannot – be implemented.

**Interaction with other components:** The Policy Registry is a crucial mechanism to support the implementation of EOSC policies and, thus, interacts with a variety of components in- and outside of the EOSC. Externally, this involves for example the harvesting of metadata on services from EOSC-compliant service catalogues. If these are to provide policy metadata which can be directly ingested by the Policy Registry, then their underlying data model will need to be at least mappable with the EOSC Policy Registry. A further question is whether – and how – the Policy Registry will need to interact with the mechanism to assess "EOSC compliant services". If the EOSC were to evolve towards a more tightly managed, quality assured federation of services, then a closer integration of these two mechanisms might be required to provide a rigorous workflow that seamlessly integrates compliance assessments for policies and operational performance requirements.

<sup>20</sup> <http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3353>

**Evolution of the policy and governance framework:** Lastly, interdependencies also exist in relation to the evolution of the policy and governance framework in the mid- and long-term. As a policy-supporting service, the Registry will need to evolve – and remain closely connected – to this framework. From the perspective of this report, this means that the EOSC governance should also ensure that e.g. the EOSC Rules of Participation and Policy Recommendations remain aligned with the metadata models of the Policy Registry and Open Science Monitor. The challenge is that the EOSC governance will need to ensure that mechanisms such as the Policy Registry continue to reflect the conceptual, data, and process requirements emerging from different EOSC policies. Hence, as policies evolve (e.g. through further standardisation) or as new policies are introduced, the EOSC governance will need to ensure that these are translated consistently into technical specifications for their supporting systems. This could for example be achieved through technical and/or community bodies, which decide on and advise on the translation of policies into metadata, process, and systems requirements. Ensuring these linkages, is inherent to achieving that the Policy Registry is transparent, auditable, and fit for purpose.

## 5.2. Future work

As alluded to in the previous section, the context within which the Policy Registry framework is proposed is evolving and changing constantly. During the ongoing work of EOSCpilot, some aspects of the interdependent challenges can be addressed, leading to further refinements – and early stage testing – of the Policy Registry framework. Future work on the Policy Registry framework should focus on three areas:

First, the Policy Registry framework will be refined further as part of the evolving work of WP3 – Policy. The objective of this is to further specify WP3's uniform policy proposition and align the policy supporting services with this where possible. With regards to T3.2 (policy-supporting services), in particular the Policy Registry's metadata model will be reviewed to accommodate potential additional alignment requirements of the metadata model with the updated Open Science Monitor framework. Additionally, the Policy Registry framework will be assessed in line with the emerging final Policy Recommendations, seeking specifically a more refined decision flow and additional (machine-readable) metadata properties which can be used to further elaborate the data model in light of use case 2.

Second, liaison with WP2 – Governance of EOSCpilot will continue in order to ensure that the Policy Registry matches and supports the architecture and functions of the updated governance framework, e.g. in relation to Technical Committees.

Third, ongoing project-external liaison will continue where appropriate. For this purpose, the Policy Registry framework has already been developed in broad consultation with EOSC Hub's WP2 – Governance, which also progresses the work on Rules of Participation further. This consultation will be continued to ensure that the Registry's specifications are further evolved in line with efforts beyond EOSCpilot.

## 6. CONCLUSIONS

This report introduced the framework for the EOSCpilot Open Science Policy Registry, a (semi-)automated system to submit and register, assess, and validate policies of EOSC service providers and stakeholders. The Policy Registry is connected to and builds on the policy requirements emerging from the EOSCpilot Rules of Participation and Policy Recommendations. Together with the Open Science Monitor and Open Science Policy Toolkit, it complements the policy-supporting services of EOSCpilot WP3 – Policy.

The Policy Registry realises four overarching functions, based on a modularised architecture:

- 1) *Registration and submission of policy metadata,*
- 2) *Assessing and validating policy alignment (in relation to Policy Recommendations) or policy compliance (in relation to Rules of Participation),*
- 3) *Storage of submitted policy (meta-)data, and*
- 4) *Provision of policy metadata for related services and secondary data consumers such as the Open Science Monitor and EOSC Governance bodies.*

A major requirement for the Policy Registry is that its specifications are closely aligned with the contents and logical structure of different policy types and policy users in the context of the EOSC. Therefore, the specifications in this report logically deduct two use cases from the current propositions included in the EOSC Policy Recommendations and Rules of Participation as well as related aspects such as the EOSC Service Portfolio and Service Management Framework. Use case 1 concentrates on EOSC service providers who seek an assessment whether their services are “EOSC compatible” according to the Rules of Participation. Use case 2 focuses on EOSC policy stakeholders (i.e. RPOs, RIs, funders / policy-makers) who seek alignment of their organisational policies with EOSC’s Policy Recommendations. While the procedural actions and decision flows for both use cases differ, the proposed modular architecture and uniform metadata model of the Policy Registry framework would allow the realisation of both use cases through just one system.

The Policy Registry is highly interdependent with other components and services of the EOSC. The emerging, rapidly evolving characteristics of the EOSC pose challenges for the development of far-reaching proposals to automate more components of the Policy Registry and make its policy metadata machine-readable. Therefore, the proposed metadata model builds where possible on controlled values and vocabularies, which have also been cross-mapped to – or adopted from – the eInfraCentral data schema. However, further machine-readable extensions of the data model, but also other components, will become possible during the evolution of the EOSC. An update of this work is therefore recommended at an appropriate point in time, in particular to optimise the propositions of this framework in light of the EOSC’s policy and governance framework, service portfolio, and service management framework.

## ANNEX A. POLICY REGISTRY DATA MODEL

The following pages contain a detailed documentation of the Policy Registry's data model. An online version of the data model can also be found under the following URL:

[https://docs.google.com/spreadsheets/d/100VARN94dYj7peiUlrjSuJxPRzp26tN4Yp6\\_cr271mM/edit?usp=sharing](https://docs.google.com/spreadsheets/d/100VARN94dYj7peiUlrjSuJxPRzp26tN4Yp6_cr271mM/edit?usp=sharing)

The following tables list first the core policy metadata entities (marked with light blue header rows), followed by the complementary metadata entities (marked with dark blue header rows).



**Core policy metadata entities:**

| Property                                     | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type  | Rule of Participation crosswalk | eInfraCentral cross-walk        |                      |
|--|--|------------|-------------------------|---|---|---------------------------------|---------------------------------|----------------------|
|  | Use case 1                                     | Use case 2 |                         |   |   |                                 | Property                        | Entity               |
| Entity:<br><b>General policy information</b> |  |            |                         |   |   |                                 |                                 |                      |
| Policy PID                                   | M  | M          | Y                       | Persistent identifier for the policy on which information is submitted.   | Multiple formats possible, e.g. DOI.  | -                               | -                               | -                    |
| Policy type                                  | M  | M          | Y                       | High-level categorisation of the policy type about which information is being submitted, i.e. Terms and Conditions, Service Level Agreement, other organisational policy (e.g. RPO privacy policy), other contractual policy. Organisational policies are policies (including internal guidance documents) which have been adopted to direct the way an organisation organises itself or provides services and other resources (a privacy or security policy are examples for this). Contractual policies regulate the duties and behaviour of several parties (e.g. service provider and user) in relation to each other. They can therefore take any form of a contract which is binding for two or more parties. | Controlled values: Terms and Conditions; Service Level Agreement; other organisational policy; other contractual policy | -                               | -                               | -                    |
| Policy title                                 | M  | M          | N                       | Title of the policy about which information is being submitted.   | Free text / string  | -                               | -                               | -                    |
| Policy Description                           | M  | M          | N                       | Summary statement of the policy, including information on the purpose of the policy and what it aims to achieve.  | Free text / string  | -                               | -                               | -                    |
| Policy URL                                   | M  | M          | Y                       | URL of a dedicated website where a copy of the relevant policy can be found.  | URL   | -                               | -                               | -                    |
| Terms and conditions                         | R  | R          | N                       | A copy of the contractual policy (e.g. terms and conditions) under which a service or open science resource is supplied. These terms regulate the contractual relationship between the supplier of the service or resource and customers.   | Free text / string.   | 1                               | Service Contractual Information | Service Terms Of Use |

| Property                              | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk        |                         |
|---------------------------------------|--|------------|-------------------------|---|--|---------------------------------|---------------------------------|-------------------------|
|                                       | Use case 1                                     | Use case 2 |                         |   |  |                                 | Property                        | Entity                  |
| Terms and conditions URL              | M  | M          | Y                       | URL of dedicated website with information on the contractual policy (e.g. terms and conditions) under which a service or resource is provided.  | URL  | 1                               | Service Contractual Information | Service Terms Of Use    |
| Service Level Agreement               | R  | O          | N                       | A copy of the service level agreement framework which is used to manage the performance level of service provision. In contrast to the "Terms of Use", which define the contractual relationship between supplier and customer, the SLA information should clearly state how performance in relation to the contract is measured. | Free text / string.                                      | 3                               | Service Contractual Information | Service Level Agreement |
| Service Level Agreement URL           | M  | O          | Y                       | URL of dedicated website of service provider with information on service level agreement.   | URL  | 3                               | Service Contractual Information | Service Level Agreement |
| Entity:<br><b>Open Science Policy</b> |  |            |                         |   |  |                                 |                                 |                         |
| Open Science Policy Yes/No            | M  | M          | Y                       | Statement of whether an Open Science Policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).  | Controlled values: yes (y); no (n); not applicable (n_a) | -                               | -                               | -                       |
| Open Science Policy Statement         | R  | R          | N                       | Summary statement of the organisation's Open Science Policy.  | Free text  | -                               | -                               | -                       |
| Open Science Policy adoption date     | R  | M          | Y                       | Adoption date of the Open Science Policy.   | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)                    | -                               | -                               | -                       |
| Open Science Policy URL               | R  | M          | Y                       | URL to dedicated website with copy of the Open Science Policy.  | URL  | -                               | -                               | -                       |
| openness / licenses                   | R  | R          | N                       | A copy of the licence terms under which a service, data, or other open science resources (e.g. literature, software, etc.) is supplied.   | Free text / string                                       | 1                               | Service Contractual Information | Service Terms Of Use    |
| openness / licenses                   | M  | M          | Y                       | URL to dedicated website with information on licence terms.   | URL  | 1                               | Service Contractual Information | Service Terms Of Use    |

| Property                          | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description  | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk |        |  |
|-----------------------------------|--|------------|-------------------------|--|--|---------------------------------|--------------------------|--------|--|
|                                   | Use case 1                                     | Use case 2 |                         |  |  |                                 | Property                 | Entity |  |
| Entity:<br>Ethics Policy          |  |            |                         |  |  |                                 |                          |        |  |
| Ethics Policy Yes/No              | M  | M          | Y                       | Statement of whether an Ethics Policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).           | Controlled values: yes (y); no (n); not applicable (n_a) | -                               |                          |        |  |
| Ethics Policy                     | R  | R          | N                       | Summary statement of the organisation's Ethics Policy.   | Free text  | -                               | -                        | -      |  |
| Ethics Policy adoption date       | M  | M          | Y                       | Adoption date of the Ethics Policy.  | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)                    | -                               | -                        | -      |  |
| Ethics Policy URL                 | M  | M          | Y                       | URL to dedicated website with copy of the Ethics Policy.   | URL  | -                               | -                        | -      |  |
| Entity:<br>Procurement Policy     |  |            |                         |  |  |                                 |                          |        |  |
| Procurement Policy Yes/No         | M  | M          | Y                       | Statement of whether a Procurement Policy for the respective organisation exists or not - or is not applicable (e.g. in case of publications). | Controlled values: yes (y); no (n); not applicable (n_a) | -                               |                          |        |  |
| Procurement Policy Statement      | R  | R          | N                       | Summary statement of the organisation's Procurement Policy.  | Free text  | -                               | -                        | -      |  |
| Procurement Policy adoption date  | M  | M          | Y                       | Adoption date of the Procurement Policy.   | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)                    | -                               | -                        | -      |  |
| Procurement Policy URL            | M  | M          | Y                       | URL to dedicated website with copy of the Procurement Policy.  | URL  | -                               | -                        | -      |  |
| Entity:<br>Data Protection Policy |  |            |                         |  |  |                                 |                          |        |  |
| Data Protection Policy Yes/No     | M  | M          | Y                       | Statement of whether a Data Protection Policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).   | Controlled values: yes (y); no (n); not applicable (n_a) | -                               | -                        | -      |  |
| Data Protection Policy Statement  | R  | R          | N                       | Summary statement of the organisation's Data Protection Policy.  | Free text  | -                               | -                        | -      |  |

| Property                             | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type  | Rule of Participation crosswalk | eInfraCentral cross-walk                          |                      |
|--------------------------------------|--|------------|-------------------------|---|---|---------------------------------|---|----------------------|
|                                      | Use case 1                                     | Use case 2 |                         |   |   |                                 | Property  | Entity               |
| Data Protection Policy adoption date | M  | M          | Y                       | Adoption date of the Data Protection Policy.  | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)             | -                               | -   | -                    |
| Data Protection Policy URL           | M  | M          | Y                       | URL to dedicated website with copy of the Data Protection Policy.   | URL   | -                               | -   | -                    |
| <b>Entity: Service provision</b>     |  |            |                         |   |   |                                 |   |                      |
| Service availability                 | M  | O          | Y                       | Quantitative factor (percentage) describing the fraction of a time period that an item is in a condition to perform its intended function upon demand; described as a percentage.   | percentage (0 - 100%)                             | 1                               | Service level targets and performance information | Service Availability |
| API                                  | R  | R          | N                       | Summary of API documentation.   | Free text   | 2.3                             | -   | -                    |
| API URL                              | R  | R          | Y                       | Link to API documentation for service.  | URL   | 2.3                             | -   | -                    |
| standard                             | R  | R          | N                       | Description of applicable standards used by the service.  | Free text   | 2.3                             | -   | -                    |
| standard URL                         | R  | R          | Y                       | URL of standard description.  | URL   | 2.3                             | -   | -                    |
| protocols                            | R  | R          | N                       | Description of applicable protocols used by the service.  | Free text   | 2.3                             | -   | -                    |
| protocols URL                        | R  | R          | Y                       | URL of standard description on service provider or stakeholder website.   | URL   | 2.3                             | -   | -                    |
| Service portability                  | R  | O          | N                       | Documentation of the steps undertaken by the service provider to ensure the portability of the service (i.e. the option of the service being usable e.g. in different operating systems, cloud computing environments, etc.). | Free text   | 2.4                             | -   | -                    |
| Service portability URL              | M  | O          | Y                       | URL to dedicated website with service portability documentation.  | URL linking to service portability documentation. | 2.4                             | -   | -                    |

| Property                          | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type                               | Rule of Participation crosswalk | eInfraCentral cross-walk                          |   |
|-----------------------------------|--|------------|-------------------------|---|--|---------------------------------|---|---|
|                                   | Use case 1                                     | Use case 2 |                         |   |  |                                 | Property  | Entity  |
| Quality of service                | M  | O          | N                       | Description of the service quality, particularly: service capacity (i.e. how many users or service requests can a service host in a given time); service usage (i.e. at what percentage of the service capacity does the service operate on average); service reliability (i.e. the probability that a service operates without failure in a given amount of time, provided that the service is generally available in this period of time (cf. service availability)). | Free text                                | 2.6                             | Service Level Targets and Performance Information |   |
| Quality certification             | R  | R          | N                       | Description of the quality certification mechanisms which the service complies with, e.g. CoreTrustSeal.  | Free text (e.g. CoreTrustSeal; ISO27001) | 2.6                             | -   | -   |
| Quality certification URL         | R  | R          | Y                       | URL to service provider or stakeholder site where applicable quality certification is documented.   | URL                                      | 2.6                             | -   | -   |
| Usability materials               | R  | O          | N                       | Description of the user documentation and user support channels (e.g. helpdesk contact form or phone line, service manuals, etc.) for a given service.  | Free text                                | 2.7, 1                          | Service Support Information                       | Service User Manual; Service Training Information; Service Helpdesk |
| Usability materials               | M  | O          | Y                       | URL to dedicated website with user support documentation.   | URL                                      | 2.7, 1                          | Service Support Information                       | Service User Manual; Service Training Information; Service Helpdesk |
| Entity:<br><b>Data provision</b>  |  |            |                         |   |  |                                 |   |   |
| Data availability and sharing     | R  | R          | N                       | Explicit statement on the conditions under which data is made available and how it is being shared.   | Free text                                | 2.2                             | -   | -   |
| Data availability and sharing URL | M  | M          | Y                       | URL to dedicated website with information data sharing conditions.  | URL                                      | 2.2                             | -   | -   |

| Property  | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description  | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk |        |
|---|--|------------|-------------------------|--|--|---------------------------------|--------------------------|--------|
|   | Use case 1                                     | Use case 2 |                         |  |  |                                 | Property                 | Entity |
| Data curation and preservation policy Yes/No        | M  | M          | Y                       | Statement of whether an Data Curation and Preservation Policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).   | Controlled values: yes (y); no (n); not applicable (n_a) | -                               | -                        | -      |
| Data curation and preservation policy statement     | R  | R          | N                       | Clear description of the data curation and preservation processes (or applicable policies) which apply for the respective data (1,000 characters).   | Free text  | 2.2                             | -                        | -      |
| Data curation and preservation policy URL           | M  | M          | Y                       | URL to dedicated website where data curation and preservation processes are documented.  | URL  | 2.2                             | -                        | -      |
| Data curation and preservation policy adoption date | M  | M          | Y                       | Adoption date of the Data Curation and Preservation Policy.  | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)                    | -                               | -                        | -      |
| Data portability                                    | R  | R          | N                       | Short documentation of the steps undertaken by the service provider to ensure the portability of the data (i.e. the option of the data which is produced by the service being transferred to other equivalent services in order to avoid service provider lock-ins). | Free text  | 2.4                             | -                        | -      |
| Data portability URL                                | M  | R          | Y                       | URL to dedicated website with information and documentation on data portability.   | URL  | 2.4                             | -                        | -      |
| FAIR processes tools                                | R  | R          | N                       | Description of the procedures and service means which a service provider offers in order to help users of the service ensure that data is FAIR.  | Free text  | 2.7                             | -                        | -      |
| FAIR processes URL                                  | M  | M          | Y                       | URL to dedicated website with information on implemented processes to ensure FAIRness of data.   | URL  | 2.7                             | -                        | -      |



| Property                            | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type  | Rule of Participation crosswalk | eInfraCentral cross-walk                                     |                              |  |
|-------------------------------------|--|------------|-------------------------|---|---|---------------------------------|--|------------------------------|--|
|                                     | Use case 1                                     | Use case 2 |                         |   |   |                                 | Property   | Entity                       |  |
| Entity:<br><b>Access conditions</b> |  |            |                         |   |   |                                 |  |                              |  |
| Access Policy Yes/No                | M  | M          | Y                       | Statement of whether an Access Policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).  | Controlled values: yes (y); no (n); not applicable (n_a)  | -                               | -  | -                            |  |
| Access Policy Statement             | R  | R          | N                       | Explicit declaration of who can access a service or other open science resources(e.g. free for all, fee-based, peer-review), access constraints (e.g. relating to usage intensity), information on how access is managed. | Free text   | 2.2                             | [Basic Service Information; Service Contractual Information] | [Target Users; Terms of Use] |  |
| Access Policy URL                   | M  | M          | Y                       | URL of dedicated website with access policy information.  | URL   | 2.2                             | [Basic Service Information; Service Contractual Information] | [Target Users; Terms of Use] |  |
| Access Policy adoption date         | M  | M          | Y                       | Adoption date of the Access Policy.   | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)   | -                               | -  | -                            |  |
| Allowed users                       | M  | M          | Y                       | Provides a typology of who can access a service or other open science resource, e.g. all, researchers, research managers, specific disciplines (will need to provide information on which discipline).                    | Controlled list of values based on eInfraCentral schema: Research Organisations, Industry, SMEs, Researchers, Scientists, Funders, Policy Makers, Service Providers, Data Providers, others | 2.2                             | Service description  | Target users                 |  |
| Capacity limits                     | M  | M          | N                       | Description of constraints on capacity while maintaining standards of service quality and performance, e.g. "maximum 20,000 concurrent users per day", "10 access requests per second".                                   | Free text   | 2.2                             | Service Level Targets and Performance Information            | Service capacity             |  |
| Access conditions                   | M  | M          | Y                       | Typological description of the conditions under which access is granted to a given service or other open science resources offered by a relevant stakeholder.   | Controlled list of values: open; fee-based; research excellence (e.g. based on assessment by peer reviewers); name-based access; group-based access   | 2.3                             | -  | -                            |  |

| Property                                 | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk        |               |
|--|--|------------|-------------------------|---|--|---------------------------------|---------------------------------|---------------|
|  | Use case 1                                     | Use case 2 |                         |   |  |                                 | Property                        | Entity        |
| Access fee type                          | M  | M          | Y                       | Short description of the type of access fee incurred by users.  | Controlled values, e.g.: free; freemium (free basic service, charges apply for premium service); flat rate (e.g. daily/monthly/annual fees, but no usage caps); unit-based fees (i.e. charges are based on specific units, e.g. hours of service used, number of projects hosted, amount of data hosted, etc.) | 2.5                             | Service Contractual Information | Service Price |
| Access fee statement                     | M  | M          | N                       | Clear explanation of the fees and pricing model which applies to a given service as well as other open science resources offered by a relevant stakeholder.   | Free text  | 2.5                             | Service Contractual Information | Service Price |
| Access fee statement                     | M  | M          | Y                       | URL of dedicated website with information on fees or other access cost.   | URL  | 2.5                             | Service Contractual Information | Service Price |
| Peer review                              | O  | O          | N                       | Provides information or links to resources which inform users about how the peer review process (if applicable) to determine access to the resource is structured as well as which decision-criteria apply. | Free text  | 2.2                             | -                               | -             |
| Peer review                              | O  | O          | Y                       | URL to dedicated website with information on peer review process to determine access to resources.  | URL  | 2.2                             | -                               | -             |
| <i>Entity:</i><br><b>Security policy</b> |  |            |                         |   |  |                                 |                                 |               |
| Security Policy                          | M  | M          | Y                       | Statement of whether a security policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).   | Controlled values: yes (y); no (n); not applicable (n_a)   | 2.2                             | -                               | -             |

| Property                           | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk |        |
|------------------------------------|--|------------|-------------------------|---|--|---------------------------------|--------------------------|--------|
|                                    | Use case 1                                     | Use case 2 |                         |   |  |                                 | Property                 | Entity |
| Security Policy                    | R  | R          | N                       | Summary statement of the organisation's Security Policy which is e.g. used to govern the respective service.  | Controlled values:                                       | 2.2                             | -                        | -      |
| Security Policy adoption date      | R  | R          | Y                       | Adoption date of the Security Policy.   | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)                    | 2.2                             | -                        | -      |
| Security Policy URL                | R  | R          | Y                       | URL to dedicated website with copy of the Security Policy.  | URL  | 2.2                             | -                        | -      |
| <i>Entity: Privacy policy</i>      |  |            |                         |   |  |                                 |                          |        |
| Privacy Policy                     | M  | M          | Y                       | Statement of whether a privacy policy for the provided resource exists or not - or is not applicable (e.g. in case of publications).  | Controlled values: yes (y); no (n); not applicable (n_a) | -                               | -                        | -      |
| Privacy Policy                     | R  | R          | N                       | A copy of the privacy policy, which must include information on how data is protected and how GDPR compliance is ensured.   | Free text / string                                       | 1                               | -                        | -      |
| Privacy Policy                     | R  | R          | Y                       | Adoption date of the privacy policy.  | ISO8601: YYYY-MM-DD (e.g. 2018-12-24)                    | -                               | -                        | -      |
| Privacy Policy URL                 | M  | R          | Y                       | URL to dedicated website with data protection / GDPR compliance statement.  | URL  | 1                               | -                        | -      |
| Information on collected user data | M  | R          | N                       | Clear list of the data which a service collects from its users, including explanation of the purpose of data collection as well as how it contributes to service provision / improvement. | Free text  | 2.7                             | -                        | -      |
| GDPR support                       | R  | R          | N                       | Description of the user support, tools, and (if applicable) service features to ensure compliance with GDPR (e.g. data protection by design and default).                                 | Free text  | 2.7                             | -                        | -      |
| GDPR support URL                   | M  | M          | Y                       | URL of dedicated website with description of the user support, tools, and (if applicable) service features to ensure compliance with GDPR.  | URL  | 2.7                             | -                        | -      |

**Complementary / secondary metadata entities:**

| Property               | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type  | Rule of Participation crosswalk | eInfraCentral cross-walk  |                     |
|------------------------|--|------------|-------------------------|---|---|---------------------------------|---------------------------|---------------------|
|                        | Use case 1                                     | Use case 2 |                         |   |   |                                 | Property                  | Entity              |
| Entity:<br>OS Resource |  |            |                         |   |   |                                 |                           |                     |
| Resource PID           | M  | M          | Y                       | Persistent identifier for the resource about which policy information is being submitted. | Multiple formats possible, e.g. DOI.  | -                               | -                         | -                   |
| Resource type          | M  | M          | Y                       | Short description of resource type, i.e. publications, dataset, software, service, others | Controlled values: software; publication; dataset; service; other   | -                               | -                         | -                   |
| Resource name          | M  | M          | N                       | Name or title of the resource for which policy information is being submitted.            | Free text / string  | -                               | -                         | -                   |
| Resource URL           | M  | M          | Y                       | URL where the resource can be found.  | URL   | -                               | -                         | -                   |
| Resource provider name | M  | M          | N                       | Name of the provider of the resource.   | Free text / string  | -                               | -                         | -                   |
| Entity:<br>Service     |  |            |                         |   |   |                                 |                           |                     |
| Service Function       | M  | O          | N                       | Short description of the main functions which a service provides to users.                | Free text / string; 1000 characters.  | 1                               | Basic Service Information | Service Description |
| Service category       | M  | M          | Y                       | Service categorisation according to eInfraCentral data schema for service category.       | List of controlled values: 1. Networking, 2. Compute, 3. Storage, 4. Data, 5. Software, 6. Application, 7. Security, 8. Analytics, 9. Operations, 10. Training, 11. Consulting, 12. Aggregator, 13. Other | -                               | Basic Service Information | Service Description |

| Property            | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description  | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk  |                     |
|---------------------|--|------------|-------------------------|--|--|---------------------------------|---------------------------|---------------------|
|                     | Use case 1                                     | Use case 2 |                         |  |  |                                 | Property                  | Entity              |
| Service subcategory | M  | M          | Y                       | Service categorisation according to eInfraCentral data schema for service subcategories. | List of controlled values: Direct Connect, Virtual Network, Load Balancer, Application Gateway, VPN Gateway, Exchange, Content Delivery Network, Traffic Manager, API Gateway, Job Execution, Virtual Machine Management, Container Management, Batch Processing, Serverless Applications Repository, Load Balancing, Data, File, Queue, Disk, Archive, Backup, Synchronised, Replicated, Recovery, Mining, Access, Management, Transfer Management, Registration, Persistent Identifiers, Interlinking, Publishing, Discovery Anonymisation, Preservation, Brokering, Annotation, Validation, Platform, Application, Tools, Component, Authentication and Authorisation, Coordination, Certification Authority, Identity ,Attacks protection, Business Analytics, Web Analytics, Learning Analytics, Predictive Analytics, Machine Learning, Accounting, Helpdesk, Monitoring, Analysis, Configuration, Online Courses, Inhouse Courses, Open Registration Courses, Platform, Audit and Assessment of IT Management, Audit and Assessment of Information Security, High Performance Computing, Services, Data, Applications, Software, Publications, Services-Data, Services-Applications, Services-Software, ServicesPublications, Data-Applications, Data-Software, Data-Publications, Applications-Software, Applications-Publications, Software-Publications, Services-Data-Applications, | -                               | Basic Service Information | Service Description |

| Property   | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description   | Value type  | Rule of Participation crosswalk | eInfraCentral cross-walk           |                                       |
|------------|--|------------|-------------------------|---|---|---------------------------------|------------------------------------|---------------------------------------|
|            | Use case 1                                     | Use case 2 |                         |   |   |                                 | Property                           | Entity                                |
|            |  |            |                         |   | Services-Data-Software, Services-Data-Publications, Services-Data-Applications-Software, Services-Data-Applications-Publications, Services-Data-Software-Publications, Services-Data-Software-Publications-Applications, Data-Applications-Software, Data-Applications-Publications, Data-Software-Publications, Services-Data-Applications-Software, Services-Data-Applications-Publications, Services-Data-Software-Publications, Services-Data-Software-Publications-Applications, Data-Software-Publications-Applications, Services-Data-Applications, Software-Publications, Other |                                 |                                    |                                       |
| Maturity   | M  | O          | Y                       | Describes the maturity of the service, referencing TRL-levels; only allows TRL-7 and higher.                    | Controlled values: TRL7, TRL8, TRL9   | 1                               | Service Classification Information | Service TRL, Service Lifecycle Statis |
| Operations | R  | O          | N                       | Describes details relating to the operations of the service, e.g. accounting info or business continuity plans. | Free text / string; 1000 characters   | 1                               | Service Operations Information     |                                       |



| Property                      | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description  | Value type  | Rule of Participation crosswalk | eInfraCentral cross-walk |        |
|-------------------------------|--|------------|-------------------------|--|---|---------------------------------|--------------------------|--------|
|                               | Use case 1                                     | Use case 2 |                         |  |   |                                 | Property                 | Entity |
| EOSC service type             | M  | O          | Y                       | Describes the EOSC service category which applies to a given service. 1) EOSC core services are services which are funded and managed by the EOSC. Core services ensure the core functionality of the EOSC (e.g. EOSC AAI). 2) EOSC supported services are a class of services which are useful for multiple communities but which are not developed by commercial organisations or communities alone. The EOSC therefore finances and manages "supported services". 3) EOSC service components: These are services that can be combined and built-upon to create user facing services. This includes „raw storage“, cloud platforms, VRE generators etc. 4) End-user services are Services that scientists will use to do research. Services will include data repositories, web platforms, VREs (either generic or discipline specific), compute systems, etc. | Controlled list of values: core service; supported service; service component; end-user service | -                               | -                        | -      |
| Entity:<br><b>Dataset</b>     |  |            |                         |  |   |                                 |                          |        |
| Data description and utility  | R  | R          | N                       | Description of the data (what is it?) and its utility (what is the data - potentially - useful for?).  | Free text   | 2.2                             | -                        | -      |
| Entity:<br><b>Publication</b> |  |            |                         |  |   |                                 |                          |        |
| Publication description       | O  | R          | N                       | Short description (e.g. abstract) of the relevant publication.   | Free text / string  | -                               | -                        | -      |

| Property                            | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description  | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk |        |
|-------------------------------------|--|------------|-------------------------|--|--|---------------------------------|--------------------------|--------|
|                                     | Use case 1                                     | Use case 2 |                         |  |  |                                 | Property                 | Entity |
| Entity:<br><b>Software</b>          |  |            |                         |  |  |                                 |                          |        |
| Software description                | O  | R          | N                       | Short description of the relevant software, including the use case for which it was created and what its main function is. | Free text / string   | -                               | -                        | -      |
| Software type                       | O  | R          | N                       | Brief explanation of the type of software, e.g. Python Script.   | Free text / string   | -                               | -                        | -      |
| Entity:<br><b>Stakeholder</b>       |  |            |                         |  |  |                                 |                          |        |
| Stakeholder type                    | M  | M          | Y                       | Type of submitting stakeholder.  | Controlled values: Research Performing Organisation; Funder; Ministry; Research Infrastructure; Service Provider | -                               | -                        | -      |
| Stakeholder name                    | M  | M          | N                       | Name of the submitting stakeholder.  | Free text  | -                               | -                        | -      |
| Stakeholder URL                     | M  | M          | Y                       | URL of website of submitting stakeholder.  | URL  | -                               | -                        | -      |
| Stakeholder contact email           | M  | M          | Y                       | Contact email for submitting stakeholder (domain name must be equivalent to stakeholder URL (stakeholder/url).             | Free text  | -                               | -                        | -      |
| Entity:<br><b>Service catalogue</b> |  |            |                         |  |  |                                 |                          |        |
| Service catalogue type              | M  | O          | Y                       | Describes the service catalogue type in which the service is listed.   | Controlled list of values: EOSC_service_catalogue; Third_party_catalogue   | 1                               | -                        | -      |
| Third-party service catalogue name  | O  | O          | N                       | Name of third party catalogue where service is registered.   | Free text  | 1                               | -                        | -      |
| Third-party service                 | O  | O          | Y                       | URL of third party catalogue where service is listed.  | URL  | 1                               | -                        | -      |

| Property                             | Mandatory (M) / recommended (R) / optional (O) |            | Machine readable? (Y/N) | Description  | Value type   | Rule of Participation crosswalk | eInfraCentral cross-walk |        |
|--------------------------------------|--|------------|-------------------------|--|--|---------------------------------|--------------------------|--------|
|                                      | Use case 1                                     | Use case 2 |                         |  |  |                                 | Property                 | Entity |
| catalogue URL                        |  |            |                         |  |  |                                 |                          |        |
| Entity:<br><b>Data catalogue</b>     |  |            |                         |  |  |                                 |                          |        |
| Data catalogue type                  | M  | O          | Y                       | Describes the data catalogue type in which the dataset is listed.                          | Controlled list of values: EOSC_data_catalogue; Third_party_catalogue                                | 1                               | -                        | -      |
| Third-party data catalogue name      | O  | O          | N                       | Name of third party catalogue where dataset is registered.                                 | Free text  | 1                               | -                        | -      |
| Third-party data catalogue URL       | O  | O          | Y                       | URL of third party catalogue where dataset is listed.                                      | URL  | 1                               | -                        | -      |
| Entity:<br><b>Assessment outcome</b> |  |            |                         |  |  |                                 |                          |        |
| Service status level                 | M  | R          | Y                       | Describes the compliance level of the respective service with EOSC Rules of Participation. | Controlled values: compatible; compliant; failed   | 1                               | -                        | -      |
| Open Science Policy Alignment        | R  | M          | Y                       | Indicator to describe the level of policy alignment in the area of open science.           | Controlled values: full (green), partly (yellow), no alignment (red), no assessment possible (white) | -                               | -                        | -      |
| Data Protection Policy Alignment     | R  | M          | Y                       | Indicator to describe the level of policy alignment in the area of data protection.        | Controlled values: full (green), partly (yellow), no alignment (red), no assessment possible (white) | -                               | -                        | -      |
| Procurement Policy Alignment         | R  | M          | Y                       | Indicator to describe the level of policy alignment in the area of procurement.            | Controlled values: full (green), partly (yellow), no alignment (red), no assessment possible (white) | -                               | -                        | -      |
| Ethics Policy Alignment              | R  | M          | Y                       | Indicator to describe the level of policy alignment in the area of ethics.                 | Controlled values: full (green), partly (yellow), no alignment (red), no assessment possible (white) | -                               | -                        | -      |



## **ANNEX B. GLOSSARY**

EC - European Commission

EU – European Union

EOSC - European Open Science Cloud

OA - Open Access

OS - Open Science

RI - Research Infrastructure

RPO - Research Producing Organisation