D3.2: EOSC Open Science Monitor specifications

Author(s)	George Papastefanatos ATHENA RC Elli Papadopoulou ATHENA RC Marios Meimaris ATHENA RC Leo Mack Jisc Rahul Thorat DANS
Status	Final
Version	v 1.0
Date	13/03/2018

Dissemination Level

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

Abstract:

This document provides the specifications of the Open Science Monitor Framework of the EOSCpilot, for monitoring of Open science trends and impact in EOSC. This deliverable first presents an overview of existing efforts that have so far been introduced for the monitoring of Open Science, then provides the vision of the EOSCpilot for such a framework, and proceeds with the details on the methodology, the modelling concepts, the goals, the metrics and the necessary processes for its implementation.

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 -3-D3.2	
Deliverable lead	ATHENA RC
Related work package	**WP3**
Author(s)	George Papastefanatos, ATHENA RC Elli Papadopoulou, ATHENA RC Marios Meimaris, ATHENA RC Leo Mack, Jisc Rahul Thorat, DANS
Contributor(s)	Natalia Manola, ATHENA RC Elly Dijk, DANS Paolo Manghi, CNR Emma Lazzeri, CNR Dale Robertson, Jisc
Due date	**28/02/2018**
Actual submission date	**19/03/2018**
Reviewed by	**Steven Newhouse, EBI** **Ari Lukkarinen, CSC**
Approved by	**Brian Matthews, STFC**
Start date of Project	**01/01/2017**
Duration	**24 months**

Versioning and contribution history

Version	Date	Authors	Notes
0.1	16/02/2018	All authors and contributors	The draft was circulated for editing and was decided to be sent for review one week after, while a few more comments were to be addressed.
0.2	23/02/2018	Dale Robertson (Jisc)	Draft was sent for review
0.3	24/02/2018	George Papastefanatos, Elli Papadopoulou, Marios Meimaris (ATHENA RC)	Revised version including comments received
0.4	28/02/2018	Elli Papadopoulou (ATHENA RC)	Completed missing parts and executive summary.
0.5	**05/03/2018**	George Papastefanatos, Elli Papadopoulou, Marios Meimaris (ATHENA RC)	Revised version including comments received by the reviewer 1
0.6	**08/03/2018**	George Papastefanatos, Elli Papadopoulou, Marios Meimaris (ATHENA RC)	Revised version including comments received by the reviewer 2
0.8	**13/03/2018**	George Papastefanatos, Elli Papadopoulou, Marios Meimaris (ATHENA RC)	Final Document Sent for approval and submission
1.0	**19/03/2018**	Dale Robertson (Jisc)	Final Document Submitted

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This report is part of the Policy Work Package (WP3) of the EOSCpilot project, which aims to understand the European state of the art of policies contributing to Open Science (OS), then draw recommendations that will eliminate barriers, and foster OS policy development and uptake by smart utilisation of the relevant enablers. Particularly for the EOSC Open Science Monitoring Framework Specifications, it is one of the three sub-tasks that support a policy driven approach in monitoring OS results. The Monitor is primarily envisioned to become a dynamic tool or service which assesses OS aspects such as openness and FAIRness of research results while keeping track of the impact these have on science, economy and society. Complementary to that, the OS policy registry (D3.4) and the Policy Toolkit (D3.5) will act towards ensuring that information about macro- and micro-policies, laws and regulations is well communicated to the Monitor in order for compliance with stakeholders to be met.

Thus, this report presents the first phase of an EOSC Monitor structure, delivering a Framework of Measurable Targets and specifications that EOSC should have. Moreover, it provides a review of the landscape of existing efforts in monitoring open access to research outputs. The review and categorisation of the approaches will enable organizations and initiatives, interested in implementing an OS monitoring mechanism, to identify similar efforts, then build upon and extend these tools in order to adjust them to their specific monitoring goals. It also provides a methodology with steps to be followed for the implementation of the OS monitoring framework by the EOSC and its adaptation by other interested organizations as well as the framework specifications. The specifications provide the key modelling concepts, architectural considerations, standards and processes that an OS monitor framework must support, as well as an overview of the added value services that the framework must offer to end users.

Particularly for the structure of the document, Section One introduces the scope and main components of the deliverable and highlights correlations with other Work Packages (WPs) of the EOSCpilot project. Section Two provides an overview of existing monitoring efforts and perceptions that are then mapped resulting to the monitoring targets of the EOSC OS Monitor, presented and described in Section Three. The report continues with Section Four, presenting the technical specifications, i.e., model, architecture design, processes and services that the OS Monitoring framework should offer. Lastly, Section Five concludes this first phase of the EOSC OSM with remarks and information on the further actions that will be taken alongside the work of D3.4 and D3.5.

1. MONITORING OF OPEN SCIENCE

Open Science (OS) or Science 2.0 is a major shift that has taken place in recent years in the way the scientific process is being performed by making use of the recent advancements in the information and network technologies, such as the Web 2.0. It mainly refers to new paradigms and ongoing changes in the way research is conducted, based on open digital access to research artefacts and data-driven science and affecting all phases of the research lifecycle, from the conceptualization of an idea, the collaboration of researchers, the production of scientific results, the evaluation and validation of the research output. Open science is recognized as one of the key drivers for promoting wider accessibility, transparency and integrity, trusted collaboration for research excellence and citizen participation in the scientific process. This is most evident by a number of policy recommendations and directives that have set the realization of Open Science as a first priority for Research and Innovation in EU [EU EOSC Declaration (2017), OECD Making Open Science a Reality (2015), Mallorca Declaration on Open Science (2016),] and have indicated the principles and the necessary actions for the establishment and development of the European Open Science Cloud [Realising the European Open Science Cloud,2016], as the underlying infrastructure to support Open Science in Europe.

In recent years, Open Access (OA) and Open Data policies have been developed by many research funders and policy bodies to encourage the free flow and continued preservation of data and information between researchers to accelerate research and foster innovation. However, today in Europe, thousands of researchers across scientific disciplines and other potential users are still unable to easily access data and other outputs arising from publicly funded research. The EOSCpilot project aims to address the main challenges and problems related to the realization of the EOSC, including the lack of a governance model and the lack of interoperability between data infrastructures, and making an important step towards building a uniform open innovation environment for fostering Open Research in EU, through the provision of clear incentives and rewards for the sharing of data and resources.

The realization of such an environment is a continuous process, whose basic requirements include a principled approach for monitoring and measuring the uptake and impact of Open Science trends and practices, across a clear set of high level monitoring targets, such as the openness, findability and accessibility to open science elements.

OS Monitoring is a set of services for supporting different stakeholders, such as Research Performing Organisations (RPOs), Research Funding Organisations (RFOs) and Government Bodies to measure among others:

- levels of compliance with the European Union's laws, regulations and policies regarding research and research results dissemination;
- Open Science Resources' (i.e. research artefacts, educational resources, research collaboration, citizen science) levels of openness, trustworthiness and FAIRness in each stage of the research lifecycle;
- impact of open science on society and economy.

Examples of existing services belonging to this class are JISC Monitor, consisting of Monitor UK[1] which focuses in reporting Article Processing Charges (APCs) and Monitor local[2] for compliance, and OpenAIRE Gold[3] for FP7 activities. OS Monitoring will enable the assessment of these goals, the identification of pain

[1] https://jisc.ac.uk/monitor-uk
[2] https://monitor.jisc.ac.uk/local/about/
[3] https://www.openaire.eu/postgrantoapilot

or strong points in the implementation of the EOSC through clear measurable indicators, thus facilitating evidence-based policy making.

1.1. Challenges in the context of EOSC

It is yet unclear how the uptake and impact of OS practice ought to be monitored and measured, on research in general but also on society. Several studies and research efforts have been introduced in the last two years targeting this need, which are presented in detail in Section 2. First efforts (RAND survey, SPARC Howopenisit, JISC Monitor, etc) attempted to setup a set of measurable indicators, collect data about them and measure aspects of OS related to the Openness of research results, such as open access to publications and open research data. Recent works address the need for measuring not only Openness of research results but also the FAIRness of datasets being published in OA digital repositories, thus metrics associated with the FAIRness of the data are introduced. Finally, there are recent works that consider other aspects of the OA lifecycle, introducing concepts and metrics for measuring research collaboration and validation (e.g., open peer review, open reproducibility methods for the validation of research results, etc.), citizens' engagement and citizen-produced science (e.g. activities where citizens are involved in data collection, analysis and reporting of scientific results) as well as other output beyond research publications and datasets, such as open educational material and open source software.

In general, OS monitoring is an emerging topic and one of the most active domains in the overall OS ecosystem, denoting its importance and role towards the development and long-term sustainability of the EOSC. On the other hand, the overall landscape is quite fragmented, with the various approaches and studies usually capturing isolated aspects of the Open Science movement, without however subsuming the proposed indicators and their findings into a principled monitoring framework that can be used as a tool for the continuous monitoring, assessment and validation of OS efforts.

In this respect, the main objective of this deliverable is to review, categorise and build on the existing efforts and then proceed to identify the requirements, design the specifications and provide the methodology and the key components for the implementation of a comprehensive OS monitoring framework for the EOSC. The design of such a framework goes beyond the mere definition of indicators that measure aspects of the OS or the elaboration of a single study that measures OS characteristics and provides insights in a specific point in time, and entails (Fig 1):

- The definition of a concrete methodology for deriving clear and ***measurable targets*** (called high-level monitoring targets in this context) which drive the monitoring process. In this respect, the methodology follows a ***policy-driven approach*** for identifying high-level monitoring targets; i.e. compliance to policies (i.e. funders mandates) will drive the definition of appropriate indicators.
- The identification of the key ***resources*** and trends that constitute the backbone of the Open science movement, such as the open access to publications, FAIR research data, open source software, citizen science, etc. The monitoring targets and their associated indicators aim at measuring aspects of these OS resources, e.g. a monitoring target may refer to and measure the degree of accessibility of (open) research data across data repositories in EU.
- The definition of the ***relevant stakeholders*** in the EOSC and most importantly their interest and requirements for monitoring targets and OS resources. EOSC stakeholders, mainly Research Performing Organisations (RPOs), Research Funding Organisations (RFOs) and Government Bodies, but also researchers and citizens, pose different requirements on the monitoring goals and the OS resources to be measured.
- Finally, the definition of the appropriate ***metrics and indicators*** which will quantify each monitoring

target. These indicators can be quantitative or qualitative, scalar or binary (e.g. a certain practice adheres to a policy or not) and can be combined to provide an overall score for a quality indicator. Moreover, each indicator will be associated with a ***data source*** from which monitoring data are collected, as well specific ***processes*** for further processing, validation and aggregation.

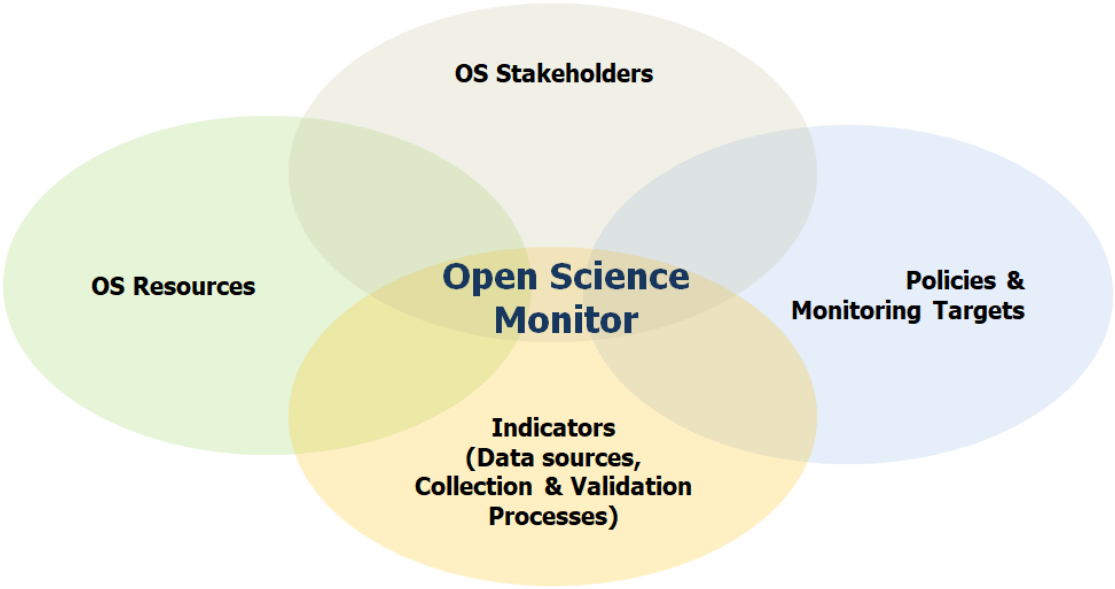

Figure 1: Building blocks of the EOSC Open Science Monitor framework

The above elements provide the building components of the EOSC OS monitor. In essence, the main goal of the OS Monitor activities (in the context of Task 3.2) in EOSCpilot is to provide the specifications for a coherent framework, an integral part of the OS lifecycle that offers a continuous monitoring and validation process of the various aspects, uptake and trends in the OS ecosystem. Such a framework should:

- Follow transparent and open monitoring processes for the representation of indicators, the identification of data sources, the collection and processing of data, the evaluation and scoring of monitoring targets and finally the presentation and visualisation of monitoring results and insights.
- Be dynamic and extensible to new OS practices and trends, allowing for the accommodation of new OS resources, the readjustment to new priorities and monitoring targets as well as the monitoring and collection of new metrics and indicators.
- Address diverse stakeholder needs (RPO, RFO, Government Bodies, Researchers, etc.) and offer added value services on top of the collected data, such as rich visualisations, navigation and search capabilities, scoring and badge systems, etc. for enabling the implementation and customization of the framework by different stakeholders. For example, a national funding body may target the monitoring of application or conformance to OS policies, whereas a Data institution may focus on the monitoring of the Fairness of the deposited data.

1.2. Scope of the deliverable and outline

The scope of this deliverable is multifold. It first aims at providing a thorough review and categorisation of the existing efforts and tools proposed so far for the monitoring of Open science resources at national, regional, European or international level. The review and categorisation of the approaches will enable organizations and initiatives, interested in implementing an OS monitoring mechanism, to identify similar efforts, build upon and extend these tools to adjust them to their specific monitoring goals.

It then proceeds with setting the requirements and proposing a methodology with the steps to be followed for the implementation of the OS monitoring framework by the EOSC and its adaptation by other interested organizations. More specifically, two possible paths are foreseen for the deployment and operation of the OS Monitor in EOSC: The first considers that OS Monitor is one of the service in the overall EOSC system (as described in D5.1 The European Open Science Cloud Architecture: Anatomy and Physiology), i.e., a monitor-as-a service tool as part of the EOSC software stack, collecting indicators and monitoring data by the organizations participating in EOSC, and offering to the EOSC stakeholders the functionality for monitoring, visualizing and gaining insights about OS trends and impact. The second enables the implementation, customisation and deployment of the OS monitor framework to individual organizations, which in turn use the framework for monitoring and presenting domain- or organization-specific targets and indicators. The latter results can also be published to the EOSC portal. Thus, the specifications section provides the key modelling concepts, architectural considerations, standards and processes that an OS monitor framework must support, as well as an overview of the added value services that the framework must offer to end users.

Although a central part of the monitoring process is the employment of well-defined measurable targets and indicators, it is noted, that it is out of scope of this deliverable to define and propose a general-purpose and closed set of these concepts and proceed with their collection and evaluation. This process is part of the specific implementation of such a framework, which needs to be carefully designed and customised to the domain-specific needs and requirements of the monitoring application.

Finally, this deliverable is closely connected with the following deliverables of the EOSCpilot project. First, it drives its requirement regarding the policy-driven monitoring of Open science from D3.1, where a thorough policy landscape review is presented, focusing on the areas of infrastructures, services, data and skills. The policies and their categorisation in D3.1 can indicate high levels goals and D3.3 derives policy recommendations that could enhance indicators for the framework's plurality. It is also related to D3.4 and D3.5 regarding the policy toolkit and the policy registry, respectively. The policy registry will specify the requirements and a service framework for machine-readable policies, which will enable the standardisation of the collection and validation of policy-conforming indicators from data sources. An additional output of this service will be the production of metrics to measure compliance with the EOSC's Principles of Engagement (D2.5). The policy toolkit will provide a pool of third party tools that can be used by external stakeholders to optimise and align their open science practice with state of the art solutions. Next, the specifications take into consideration the architecture concepts and assumptions of the EOSC services, as presented in D5.1. Following similar strategy as D3.2, D7.2 considers skills framework and provides preliminary specifications for data stewardship in EOSC. A common question, which both documents are tackling, is whether to offer such a solution within EOSC and/or focus on already available frameworks which can be loosely integrated in EOSC. Finally, the D8.2. provided the categorization of the stakeholders and their different perspective from their participation in the EOSC; this will be used in the OS monitor to identify the requirements, the different monitoring targets and added value services that should be offered to each of them.

It should be noted that the whole landscape as regards the OS Monitoring efforts and initiatives is quite dynamic, with new indicators and processes being introduced under different domains and contexts in the EOSC ecosystem. As such, this deliverable provides a first review of the current landscape and a first set of specifications based on the current development phase of the EOSC. In addition, it is foreseen that the activities performed in Task 3.2 regarding the Policy Registry (D3.4) together with D3.3 regarding Policy

Recommendations in Open Science, data protection, procurement and ethics will provide better insights of the policy-conformance indicators that should be measured in the OS Monitor. The goal is to review and enrich this deliverable based on the output of these activities.

The outline of the deliverable is as follows. Section 2 provides a thorough literature review and presentation of all existing efforts introduced so far for monitoring Open Science. Section 3 introduces the EOSCpilot approach for the design of the OSM, including the methodology for deriving the main OSM elements. It also suggests an initial categorization of the proposed metrics and indicators of Section 2 and their employment for the assessment of Monitoring Targets in the OSM. Section 4 presents in more details the requirements and the architectural / functional specifications of the framework, and its interconnection with the overall EOSC system, and conclusions are drawn in Section 5.

2. OPEN SCIENCE MONITORING FRAMEWORKS – OVERVIEW OF EXISTING EFFORTS

The recent emergence of OS initiatives and practices has highlighted the need for monitoring and assessing its impact in a principled way; a multitude of efforts have been introduced so far addressing aspects of OS monitoring, focusing on different phases and elements of OS development. Still, the overall landscape is in its infancy and quite fragmented, with the various approaches and studies usually capturing only certain quality aspects of the Open science movement, such as the degree of Openness in journal publications or the FAIRness of data and data repositories, respectively.

The first initiatives that have paved the way in OS monitoring activities towards the establishment of an OS monitoring framework are the Open Digital Science approach with the development of OS indicators and metrics and the study of RAND for monitoring OS trends and efforts up to 2016. Open Digital Science[4] has introduced indicators considering the research lifecycle steps of "Conceptualisation & data gathering/creation", "Analysis", "Diffusion of results", "Review and evaluation", as well as measurements of other Open Science elements, including drivers and constraints, namely/those being "Reputation system, recognition of contributions, trust", "OS skills & awareness", "Science with society".

The Open Science Monitor Framework by RAND[5] that was released in 2016 is the first attempt in putting together different aspects of Open Science and assessing them with different stakeholders to provide the Open Science community with quantitative and qualitative measurements and analytics. Their study refers to OS characteristics that involve research artefacts (mainly open access publications and research data), Scholarly communication activities (altmetrics, peer reviews etc), citizen science and public engagement; however it is a study, rather than a monitoring framework, which captures and assesses various OS characteristics in a specific point in time, without providing more concepts and solutions on how these characteristics can be continuously monitored or how results are updated to provide the overall EOSC community with up-to-date feedback.

EOSC OSM recognises these successful contributions and tries to incorporate parts of their work within its methodology and overall context while at the same time ensuring that major deviations between them are avoided. However, main developments of EOSC OSM rely on policy aspects and OS trends that previous approaches didn't touch upon or that were only briefly expressed by them, such as FAIR data. Other differences lie in the implementation of these two efforts, with EOSC OSM offering provisions for dynamic implementation as opposed to more static approaches.

This section provides a review of research efforts, projects and initiatives relevant to the monitoring of Open Science practices and trends in EU and worldwide. It presents monitoring and measurement frameworks that have been proposed for evaluating key Open science elements, such as research and educational resources (e.g., publications, data and software artefacts, educational material, etc), open science processes and practices. To the best of our knowledge, the review presented in this section is the first work attempting to study and categorise all recent approaches and tools according to the OS aspects they cover and their level of maturity.

Out of a variety of monitoring approaches and tools with disparate targets and capabilities, the more mature efforts and those of most interest to the stakeholders seem to be those related to the monitoring of OS research results, i.e., OA to publications, open science data and research software; hence these OS elements are covered primarily in this section. In addition, educational resources are included in this review

[4] https://ec.europa.eu/digital-single-market/en/news/open-digital-science-final-study-report

[5] http://ec.europa.eu/research/openscience/pdf/monitor/open_science_monitor_methodological_note.pdf#view=fit&pagemode=none

to understand their state of openness and use across countries as well as the impact that they have in Education. Regarding communication activities and processes that enhance scholarly and research collaboration, peer review practices and data citations are included while in measuring open science impact this work has also been inspired by monitoring achievements in open data. Lastly, with reference to the policy section, current state of the framework is limited to measuring first phase of policy readiness and adoption.

2.1. Monitoring Open Science Policies

A foundation - and incentive - for the adaptation and progress of Open Science in practice is the existence of Open Science policy mandates. Since the late 1990s, Open Science policies have been gradually introduced by research organisations (e.g. universities and research labs), funders (public, government, or private funding bodies), and publishers. While Open Access to publications remains the most common area for which mandates and policies are created, other Open Science policies, e.g. requiring the publication and long-term archiving of research data, are increasingly adopted too.

A recent catalyst for EU Member States to accelerate the development and implementation of such policies was the 2012 Communication on access to, and preservation of scientific information[6], which invited them to take action on OA policy-making. In mid-2016, this was followed by the Amsterdam Call for Action on Open Science[7], which identified and proposed 12 action items towards achieving OS. Although not a formal or binding decision, many member states have already been advocating ways to introduce and align their practices to these mandates, such as LIBER's 5 Open Access Principles for Negotiations with Publishers.

The increasing proliferation of policies however adds intricate complexity to scholarly communications - and those who want to facilitate it. Not only do Open Access publishing policies become more common, but the expansion of other Open Science policies, particularly regarding data archiving, add additional dimensions to the work of researchers, research managers, and librarians. Structuring this evolving environment is a highly important task to ensure not just that workloads remain manageable, but that new policies are taken up.

Accordingly, different working groups of the international Research Data Alliance (RDA) are therefore seeking to standardise particularly policies for the management of research data: in 2015, The RDA's Practical Policy Working Group[8] proposed a framework for 11 machine-actionable policy templates which could be used to automate policy compliance and enforcement in various data management systems[9]. Drawing from a practice survey across 10 data management systems, used by 30 institutions, the proposal identifies standardised policies and components which are most relevant for majority of institutional users. The effort aims to limit and standardise the scope of frequently used policies. It thus contributes to the emergence of a common policy framework for research data management systems and could have practical impact, e.g. if it becomes referenced frequently in public procurement tenders. Notably, the RDA's 11 templates on practical policy are endorsed in the European Commission's ICT specifications for referencing in public procurement[10].

Additionally, the RDA Group on Data Policy Standardisation and Implementation[11] has defined a common classification for the data publication policies of publishers[12]. In practice, journals can apply a number of

[6] https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

[7] http://openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf

[8] https://www.rd-alliance.org/groups/practical-policy-wg.html

[9] https://www.rd-alliance.org/practicalpolicyoutcomespolicytemplates-v2.html

[10] http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1500883656509&uri=CELEX:32017D1358

[11] https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation

[12] https://www.biorxiv.org/content/biorxiv/early/2017/04/04/122929.full.pdf

gradually different policies to require whether - and how - the data which underpins articles should be made accessible. The goal of the proposed four policy classes is to help journals identify a suitable policy class, implement standardised policy texts and define easy-to-follow processes for the sharing or publishing of research data. Accordingly, the four policy classes are located along a spectrum of differing requirements for the sharing of research data (see figure below).

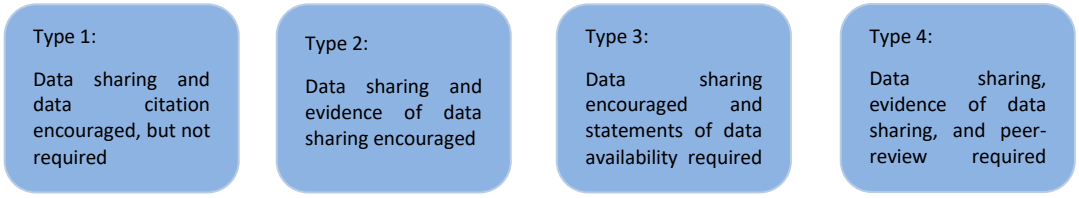

Figure 2: Standardised policy types for research data[13]

In sum, this should help to create easily understandable and recognisable policies, which authors, researchers, and peer-reviewers can refer to in their work. Leading contributors to this development have been Jisc and the publishers Wiley and Springer. The four policy types are also currently used by the more than 2,500 Springer journals.

However, despite these notable initiatives, the challenge of making the open science space manageable and navigable requires multiple answers. Efforts to actively manage the proliferation and rapid evolution of open science-related policies through coordination and standardisation need to be supplemented by a more reactive approach: Therefore, various stakeholders and initiatives have implemented solutions to monitor open science policies. Broadly speaking, these monitoring solutions involve "classical" monitoring services, producing aggregate statistics on the introduction of policies, and "user-centred" services which are designed to help users actively navigate the complex landscape from their micro-view. The remainder of this section therefore provides an overview, comparison, and trend analysis of the main approaches and tools.

The Registry of Open Access Repository Mandates and Policies (ROARMAP)[14] is arguably the classical archetype: Established in 2005, it provides a searchable international registry which tracks the growth of open access mandates and policies adopted by universities, research institutions and research funders. Providing a macro-perspective of developments, ROARMAP collects information on policies which require or request researchers to provide open access to their peer-reviewed research article output by depositing it in an open access repository. Until February 2018, 898 institutions had registered their policies with ROARMAP[15]. The deposit process requires them to also provide information such as adoption dates, policy scope (including applicable artefacts), embargo and publishing options. Based on these submissions and subsequent reviews, statistics are aggregated on a quarterly basis, structured by organisational types. In sum, ROARMAP thus provides a reliable service to get a macro-view of the growth of Open Access policies and mandates.

Like ROARMAP's backend, Fairsharing offers a refined database with detailed information on currently 106 policies of journals, funders, organisations, and individual projects[16]. Other than ROARMAP, the Fairsharing database is primarily curated by a dedicated project team. Individual stakeholders can contribute by claiming policies or add missing ones, however they are not the primary source of knowledge for the project. While it does not publish aggregate statistics, the database allows extensive filtering, leading to more detailed insights into the emerging policy practices of different domains: In addition to basic information on policies (e.g. description, publication date, publisher type), records also specify which

[13] Figure recreated from https://www.biorxiv.org/content/biorxiv/early/2017/04/04/122929.full.pdf
[14] http://roarmap.eprints.org/
[15] This consists of 83 funders, 56 funder and research organisations, 11 "multiple research organisations", 674 research organisations (e.g. universities or research institutions, and 74 sub-units of research organisations.
[16] https://fairsharing.org/policies/?q=

databases and standards the policy mentions as well as which domain it is used in. The categorisation of "domains" used by Fairsharing is not limited to disciplines of science, but also covers practically relevant domains for the application of open science policies, including data sharing, metadata standardisation, and publication activities. By linking policies to one or multiple domains, the database enables the monitoring of divergence. The Dutch Royal Academy for Arts and Sciences recently highlighted that a crucial success factor for the adoption of Open Science will be the extent to which its implementation allows precisely such domain-specific divergence, instead of attempting to create one-size-fits-all-solutions[17]. Fitting into this broader push for differentiation are initiatives such as ScienceEurope[18] and the EU-funded project Parthenos[19] which are developing domain specific research management protocols[20] and a policy wizard[2122] to help scientists navigate through the plethora of domain-specific and general policies. To support and monitor these developments, the Fairsharing database is an important source of knowledge: It helps to create a detailed repository to track how policies are connected across domains, databases, and standards.

From the practitioner perspective of librarians, research managers, and researchers, the previously mentioned initiatives however have one crucial deficit: They do not make the complex policy environment more navigable from their user view. Jisc's Sherpa services therefore take a distinctively different approach: An intermediating service which actively helps users, particularly librarians, researchers, repository and research managers, to navigate the increasingly complex open access policy landscape. Rather than aggregate statistics, Sherpa services provide primarily an interface where users can check how they can comply with various open access mandates. The strong demand for this use case has led to the development of four Sherpa services:

1) The use case of **Sherpa RoMEO[23]** is to allow researchers and librarians to easily see publishers' conditions for open access archiving on a journal-by-journal basis. For this, Sherpa RoMEO provides summaries of self-archiving permissions and conditions of rights given to authors. The service aggregates and analyses publisher open access policies from 99 countries, covering more than 22,000 peer-reviewed journals and serials from 2,515 publishers[24]. The summaries provided by Sherpa are based on the copyright transfer agreements and open access policies of publishers. Additional sources of information can be other publisher documents (which are available online) and personal communications with the publisher. Journal titles are gathered from publishers' websites and supplemented with additional information from feeds including the British Library's Zetoc service, DOAJ, and Entrez. The back-end of the service is based on a database which contains standardised policy information, using Sherpa terms, to facilitate the comparison between publishers[25]. Sherpa RoMEO highlights the basic conditions of publishers' archiving conditions through a four-colour coded scheme[2627]. More detailed information is available via the individual

[17] https://knaw.nl/en/topics/openscience/blog-jose-van-dijck-on-open-science/open-science-major-differences-between-disciplines

[18] http://www.scienceeurope.org

[19] http://www.parthenos-project.eu

[20] http://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf

[21] http://data.d4science.org/ek5XeHhJR0ptMXNjeVJqL0x3cDgxZ2thOXZpc0lJOFNHbWJQNStIS0N6Yz0

[22] https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b1eb4e98&appId=PPGMS

[23] http://sherpa.ac.uk/romeo/index.php?la=en&fIDnum=|&mode=simple

[24] http://sherpa.ac.uk/romeo/statistics.php?la=en&fIDnum=|&mode=simple

[25] Sherpa RoMEO also publishes summary statistics on the policies included in its database. However, other than in the case of Roarmap, these do not form the core service offer: http://sherpa.ac.uk/romeo/statistics.php?la=en&fIDnum=|&mode=simple

[26] http://sherpa.ac.uk/romeo/definitions.php?la=en&fIDnum=|&mode=simple&version=#colours

entries for journals: This includes archiving options and applicable embargo periods for the author's pre- and post-prints as well as the publisher's version; furthermore, general archiving information, links to copyright policies, and compliant funder policies are stated[28]. All Sherpa RoMEO data is openly licensed, allowing re-use for non-commercial purposes; data can also be accessed via an API[29].

2) **Sherpa Juliet**[30] enables researchers and librarians to see funders' conditions for open access publication. Similarly, and complementary to Sherpa RoMEO, Juliet is a searchable database, summarising up-to-date information about funders' policies and their requirements on open access, publication and data archiving. With policies being primarily analysed and curated by the Sherpa team, Juliet is also very similar to RoMEO in operational terms. Data is also made accessible via an API[31] and summary statistics for the 139 funders and 144 policies are provided[32]. For each individual funder, users can also find more detailed, mostly standardised information on funder's general open access policy, as well as requirements for open access archiving, publishing, and data archiving[33].

3) **Sherpa FACT**[34] combines the facilities of RoMEO and Juliet as well as some data from the Directory of Open Access Journals and Europe PMC/PubMed Central. The tool provides guidance to researchers on whether a journal complies with the open access policies of Research Councils UK (RCUK), Wellcome Trust and Charity Open Access Fund (COAF). Sherpa FACT also offers easily understandable advice on available options[35]. From a user perspective, Sherpa FACT thus merges the facilities of Sherpa RoMEO and Juliet, allowing practitioners to quickly check whether a journal's publishing and archiving policy is compliant with the requirements laid out in the open access policy (or policies) of one or multiple funders. Data is also available programmatically through a beta API[36].

4) **Sherpa REF**[37], launched in March 2016 and currently in beta version, is the latest addition to the Sherpa suite. Its use case is similar to Sherpa FACT: Helping authors and institutions decide whether a journal allows them to comply with the OA REF policy of the Higher Education Funding Council for England (HEFCE)[38]. The underlying technical and operational functionality is highly similar to Sherpa FACT; data is also available via an API[39] and can be reused for non-commercial purposes[40].

A crucial challenge for the development of both "classical" and "user-centred" monitoring services is the extent to which these service offers can be combined - or at least made interoperable - to exploit synergies. Curating and standardising the underlying data through for either service is a laborious task. Therefore, a common metadata framework to monitor open science policies can harness substantial synergies: First

[27] Colour codes include green (can archive pre-print and post-print or publisher's version/PDF), blue (can archive post-print (ie final draft post-refereeing) or publisher's version/PDF), yellow (an archive pre-print (ie pre-refereeing)), and white (archiving not formally supported).
[28] http://sherpa.ac.uk/romeo/issn/2226-6933/
[29] http://sherpa.ac.uk/romeo/apimanual.php?la=en&fIDnum=|&mode=simple
[30] http://v2.sherpa.ac.uk/juliet/
[31] http://v2.sherpa.ac.uk/juliet/api.html
[32] http://v2.sherpa.ac.uk/view/funder_visualisations/1.html
[33] http://v2.sherpa.ac.uk/id/funder/695
[34] http://sherpa.ac.uk/fact/index.php
[35] http://sherpa.ac.uk/fact/about.php?juliet_id=&funderlist=
[36] http://sherpa.ac.uk/fact/apimanual.php?juliet_id=&funderlist=
[37] https://ref.sherpa.ac.uk/
[38] http://www.hefce.ac.uk/rsrch/oa/Policy/
[39] https://ref.sherpa.ac.uk/api
[40] https://ref.sherpa.ac.uk/reuse

steps in this direction have been proposed by the PASTEUR4OA project, which proposed a CERIF-based description for Open Access policies[41].

Finally, the activities of T3.1 of EOSCPilot have also provided a first categorization of policies. Policy areas with importance to the first phase of EOSC are indicated in the deliverable D3.1 Policy Landscape Review (WP3-EOSCpilot). They concentrate on the main elements of EOSC, these being **infrastructures** and **services, data**, **skills** and procurement, as follows

- **Policies for Infrastructures and Services:** mainly focusing in achieving interoperability, standardising procedures and minimising transaction costs for both the private and public sector.
- **Policies fostering the free flow of data:** mainly with respect to ethics, data re-use and overcoming issues such as costs to access authors' work, copyrights (including copyright exemptions) of personal and non-personal data in the private and public sector.
- **Policies for improving skills and supporting the development of open educational resources:** mainly focusing in developing open educational resources and in delivering data related skills to support research lifecycle needs targeting not only researchers but everyone interested in Open Science, by simultaneously supporting their research portability and mobility.
- **Policies regarding public procurement:** mainly referring to infrastructures and services, and their procurement at the Member State level to then identify related issues within EOSC.

In addition, the work of D3.3 on Policy Recommendations is expected to provide more concrete indicators regarding application and compliance to certain policies.

2.2. Monitoring Open Science Resources

This section provides an overview of monitoring efforts as well as recommendations/suggestions made by organisations, projects and working groups of experts regarding the monitoring of Open Science Resources, i.e., OS research artefacts and results. Ongoing studies and recent work are included to highlight the most current state of such exercises stipulating more progressed areas yet also potential gaps in the literature. Most of these efforts have addressed elements of research output such as (open access to) publications, FAIR (research) data, open source software, open educational resources, and scholarly communication/research collaboration activities. These are presented in the following sections.

2.2.1. Open Access to publications

Open Access movement dates few decades now, forming one of the first attempts to normalise publishing and access costs by creating new business models that accelerate knowledge production, information dissemination and innovative research achievements implementation. It also aimed to identify Intellectual Property Rights and copyright patterns permitting re-use as well as to curate and preserve published materials by promoting archiving in institutional repositories. As a result, today there are three core routes of OA[42] that are being followed according to:

[41] https://ac.els-cdn.com/S1877050917303022/1-s2.0-S1877050917303022-main.pdf?_tid=196f7748-125c-11e8-ae4d-00000aab0f6c&acdnat=1518704920_952e91dbc370013a177cc1398574534d

[42] Other variations such as Diamond OA or Bronze OA are derivatives of one of the core routes that choose a specific/specialised focus and requirements to meet.

- *Green route or self-archiving* referring to free of processing and access costs publications, both pre-prints and post-prints, that authors deposit in institutional repositories. In the Horizon 2020 mandate, the European Commission suggests the Green route for peer-reviewed research outputs to be followed even on the occasion that submission to a journal is expected in the long run, provided that publishers' policies/requirements don't contradict with authors' copyrights.

- *Gold route or OA publications* concerning peer-reviewed publications in OA journals. Like the green route, they are freely accessible to the public, however their publication is subject to article processing costs (APCs) which are later reimbursed by research funders as legal publication costs.

- *Hybrid OA* declaring double-dipping in the publication process. That translates to payment of peer-reviewed articles both for their processing by peers and for their access through subscriptions to journals or other synergies/memberships with publishers.

Indicators for measuring OA to publications. One of the efforts to measure OA in journals is the "HowOpenIsIt?: Open Access Spectrum"[43] which translates the core elements of Open Access identified in journal policies to a classification that shows levels of openness in journals, ranking them from closed access to more "open" approaches until fully open access is achieved. Like the work done for journals, SPARC together with the Open Research Funders Group (ORFG) developed a guide to assess openness in funders policies related to the OS Resources of publications, data, code and software[44].

Other attempts, such as **Nichols & Twidale**[45], have examined the possibility of developing an h-index contributing to the wider picture of researchers OA practice. Particularly for their approach, it focuses on papers and books to provide metrics in the form of an individual's index that informs about, but is not limited to, access, re-use and preservation of research outputs while also allowing for a comparative view with similar closed access values in practices.

Recommendations on monitoring OA. A new study by Jeroen Bosman and Bianca Kramer[46], examines levels of openness of articles determined by the OA journal that they are published in, in correlation to the license that has been given to them.

Discussions for the Open Impacts framework in the Open Scholarship Initiative[47], in 2016, revolved around three areas: *Measuring openness*, Utilization measures, Understanding economic impacts of open. Particularly for the needs of the first area, an "openness score" was envisaged to be composed by metrics expressing capacities of license, availability, permanence and format measures.

The recommendations expressed at the Knowledge Exchange Workshop in 2016[48] were driven by existing efforts such as JISC's Monitor Local and the Netherlands' national open access monitoring workflow. The

[43] SPARC's how open is it? Methodology https://sparcopen.org/our-work/howopenisit/
[44] HowOpenIsIt for Funders http://www.orfg.org/resources/
[45] http://onlinelibrary.wiley.com/doi/10.1002/asi.23741/full
[46] https://peerj.com/preprints/3520/
[47] OSI2016 *"openness score"* http://osinitiative.org/osi-reports/osi2016-reports/report-from-the-open-impacts-workgroup/
[48] http://repository.jisc.ac.uk/6648/1/KER0001_KE_workshop_Monitor_OA_Publications_and_cost_data_MAR17_v4_(2).pdf

naming of the working groups that made these suggestions reveal the areas that they also focused on: *Data (collection), Workflows, Standards* and *Policy*.

- The group on *Data collection for monitoring OA publications* made recommendations of collection activities with respect to differences pertaining to the OA routes (Green, Gold, Hybrid), mostly concerning standardised procedures that would facilitate integration with CRIS's (metadata, standardised data formats, etc).

- Aggregation issues occupied the discussions of the *group on Workflows for monitoring OA publications*, too. Among their recommendations were the inclusion of license statements in Crossref (for offsetting agreements) and the use of ORCIDs; CRIS's to integrate different categories of OA and monitors to be performed at the end of embargoes was also suggested.

- The use of standards was discussed within the *Standards group for monitoring OA publications.* Libraries were recognised as key players to ensure that all stakeholders follow standardised procedures for articles deposit. Publishers were advised to deliver standardised information as well as to be CERIF compliant. The benefits of introducing a field for APCs in the OAI-PMH was also discussed. Regarding issues relevant to CRIS's, it was agreed that they are ideal for monitoring OA which to do so efficiently, they should use PIDs to interconnect and cross-check information and OpenAIRE guidelines to ensure interoperability among them and repositories.

- Lastly, recommendations of the *Policy group on monitoring OA publications* revolved around the scope of monitoring and steps in ensuring its effectiveness, while it also identified gaps and weaknesses that could be strengthened if addressed within/ incorporated in policies (e.g. OA data should have an API).

Tools for monitoring OA. JISC has developed a series of services to meet OA needs. JISC Monitor[49] produces funders, institutions and publishers reports about OA expenditure to inform Higher Education Institutes in the UK about the range of publishing costs (Articles Processing Charges - APCs) per stakeholder when following the Gold or the Green route of OA respectively. SHERPA/FACT[50] on the other hand, helps in identifying and overcoming compliance issues between funders requirements' and publishers' OA policies.

OpenAIRE GoldPilot for FP7 captures OA funding and costs in preparing and publishing open science within H2020 projects to then develop a framework where strengths and weaknesses are identified and communicated to the EC.

Lastly, according to responses gathered from a quick survey on the National Open Access Desks of the OpenAIRE network, regarding national monitoring mechanisms and attempts to measure Open Access stands Dutch national monitoring portal, NARCIS[51] , United Kingdom's JISC services and Danish National OA Indicator[52]. More information is provided in the Appendix II.

[49] https://www.jisc.ac.uk/monitor-uk
[50] http://www.sherpa.ac.uk/fact/about.php?juliet_id=&funderlist=
[51] https://www.narcis.nl/
[52] https://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access/Publications/open-access-barometer

2.2.2. Monitoring Open Science data – FAIR approaches

The success of EOSC depends on the ability of its users to work with a large-scale, dynamic infrastructure that spans multiple scientific domains. Users benefit with new approaches to data science and contribute to the open science provided they follow FAIR principles, i.e., Findable, Accessible, Interoperable and Reusable principles[53]. FAIR principles were initially developed to measure data for its findability, accessibility, interoperability and reusability. Since its introduction in 2014, the FAIR principles are widely discussed and modified to apply to services, trainings, infrastructures and repositories. Various updates (e.g. FORCE11[54]) add to the long-term and sustainable data curation across multiple data lifecycles. This makes data stewardship (and with it, also the open science data) a collective endeavour, involving at least the individual researcher, colleagues in the study, their host organisation and the associated research domains, and potentially other communities that care about the data[55]. This section summarises various attempts made by FAIR communities/expert groups to relate FAIRness of data, trainings and services to their openness.

Relation between FAIRness and Openness. Many of the FAIRness principles for data, services or trainings are prerequisites for their openness, however a service or data being FAIR does not directly imply that it is also open. In this regard there is also some discussion about the "levels" of openness which are subjected to ownership, intellectual property rights, sensitivity issues, licensing etc. Barend Mons[56] explains it very adeptly: "FAIR is not equal to Open…. The FAIR principles are directed more towards technical aspects than towards moral and ethical aspects of data or services." FAIR principles do not imply data being "open" or "free", ..., but require clarity and transparency around the conditions governing access and reuse." "FAIR principles stress upon provisions to make data available for reuse under clearly-defined conditions and licenses, available through a well-defined process, and with proper and complete acknowledgement and citation. This will allow much wider participation of players."[57] On the other hand, Open does not directly mean FAIR. Many of the datasets, trainings and services are open. However, without being FAIR, e.g., without proper metadata or software to access, they are useless to their intended users. This is especially true for fields such as medical data where patient health history matters, or humanities where working on already available data makes a large part of the research. In summary, for OPENness, ethical and moral aspects should be added to FAIRness of data/services/infrastructures.

[53] https://www.nature.com/articles/ng.3544

[54] https://www.force11.org/fairprinciples

[55] EOSCpilot D7.1 Skills landscape analysis and competence model, pp. 41

[56] Barend Mons, Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud (Data as digital objects approach)

[57] https://www.universiteitleiden.nl/en/news/2016/03/the-fair-principles-herald-more-open-transparent-and-reusable-scientific-data

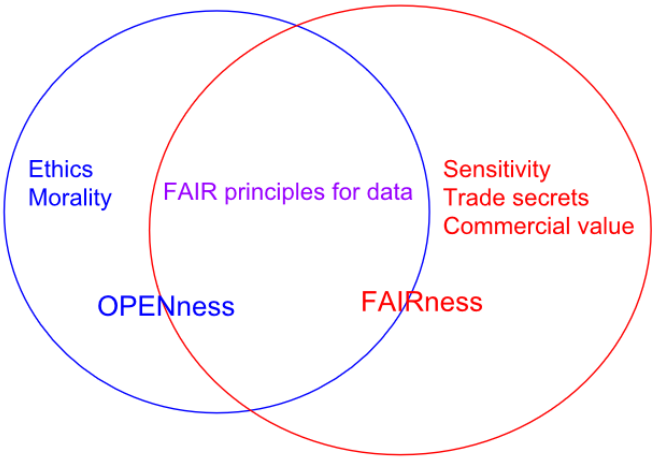

Figure 3: Connection between Openness and FAIRness

Metrics, tools, services and certifications. Several FAIR **metrics** have been proposed by the communities/groups to define framework within which the FAIRness of the trainings, data, and services can be assessed[58]. To measure the FAIRness of the data, services, trainings, several attempts are being made in the form of frameworks/criteria:

- fair metrics group[59]: a group of experts whose founding members include two of the FAIR guiding principles' initiators. Their scope is to create metrics that will measure different communities' data objects' FAIRness by assessing and evaluating each one of the F, A, I, R principles. First draft of their effort is the fairmetrics form[60] and a GitHub page[61] that attempts to complete this form for each principle inviting stakeholders to contribute to this activity as well. FAIR Metrics[62] gives "core set of semi quantitative metrics having universal applicability for the evaluation of FAIRness, and a rubric within which additional metrics can be generated by the community." This is an ongoing activity which at this stage has proposed a set of metrics also available in machine readable formats[63], each one explaining how it should be used on the research artefact that is expected to be used when measuring its FAIRness.

- NIH Commons Framework Working group on data FAIRness Metrics

[58] Dumontier's webinar, Are we FAIR yet?, 31st January, slide on Metrics.
[59] http://fairmetrics.org/
[60] http://fairmetrics.org/fairmetricform.html
[61] https://github.com/FAIRMetrics/Metrics
[62] A design framework and exemplar metrics for FAIRness
Mark D Wilkinson, Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos, Michel Dumontier
bioRxiv 225490; doi: https://doi.org/10.1101/225490
[63] https://github.com/FAIRMetrics/Metrics

- RDA\WG Data Fitness for Use Criteria[64]: a group that tries to match "data objects, access services, and data management processes such as the level of annotation, curation, peer review, and citability or machine readability of datasets" with a corresponding metric.

- RDA IG on Data Discovery Paradigms[65]: a group inspired by FAIR principles trying to improve data search and solve data discovery issues (F). The goal is to identify concrete deliverables such as a registry of data search engines, common test datasets, usage metrics, and a collection of data search use cases and competency questions.

FAIR Tools. In addition, a variety of tools have been developed which have their technical and conceptual background in these metrics.

- Open Data Certificate[66] is a free online tool developed and maintained by the Open Data Institute, to assess and recognise the sustainable publication of quality open data. It assesses the legal, practical, technical and social aspects of publishing open data using best practice guidance.[67]

- DANS FAIR badges via FAIRdat [68](FAIR data assessment tool): a tool to assess FAIRness of datasets. The prototype tool was tested during various workshops [69]and meetings on the datasets of internal and external repositories. Currently running on a survey monkey but will be migrated to a database to accommodate needs of enriched metadata such as PIDs, contributors, etc. For the calculations, the tool takes into consideration the principles Findability, Accessibility and Interoperability by outlining 5 criterion levels. The average result of these principles leads to the last principle: Reusability. All the principles are translated to FAIR badges that follow levels of compliance.

- FAIRsharing[70] as the data catalogue that brings together policies, standards and databases, accommodating in such way FAIR needs

- FAIR Accessor - Interoperability and FAIRness through a novel combination of Web technologies[71]: includes Container Resource, MetaRecord, Triple Descriptors and FAIR Projectors, each one to have been assessed according to FAIR principles

- The DTL FAIR Data team has developed tools that compose the so-called 'Data FAIRport':
    - FAIRifier and Metadata Editor (to create)
    - FAIR Data Point (to publish)
    - FAIR Search Engine (to find)

[64] https://www.rd-alliance.org/groups/assessment-data-fitness-use
[65] https://www.rd-alliance.org/groups/data-discovery-paradigms-ig
[66] https://certificates.theodi.org/en/about/badgelevels
[67] certificates.theodi.org
[68] http://blog.ukdataservice.ac.uk/fair-data-assessment-tool/
[69] For example at the Open Science FAIR, Athens, September 2017: http://www.opensciencefair.eu/training/parallel-day-3-2/fair-metrics-starring-your-data-sets
[70] https://fairsharing.org/
[71] https://peerj.com/articles/cs-110/

- ORKA (to annotate)

These tools are currently for demonstration purposes only. They are still under development and are deployed at SURFsara.

- Finally, Tim Berners-Lee's 5 star Open Data[72] rating is a generic assessment approach to the publishing of Linked Open data, which can be applied to the assessment of Open Research Data.

Monitoring Services built on top of these tools. The tools are used to monitor and assess services provided by repositories, archives, libraries, e.g. data deposition, curating, trainings etc.

- JISC FAIR in practice[73]: An assessment of about 16 institutions in the UK that have described (publicly) their research data infrastructures (or plans for internal systems and services).
- The reusable data project[74] focuses on reproducibility of research permitted by licenses. It provides the user (initially designed for data providers) with a taxonomy of licenses and terms of use and conditions to show consequences in replicability of research results and artefacts that is based on licenses selection.
- FAIRDOM[75] is a joint action of ERA-Net ERASysAPP and European Research Infrastructure ISBE to establish a data and model management service facility for Systems Biology.

Certification of repositories. Various repository certifications are available with a system of approval seals[76] and using badges[77]. The *European framework of certification levels* for repositories consists of three layers :

- **Basic Certification** is granted to repositories which obtain CoreTrustSeal certification
- **Extended Certification** is granted to Basic Certification repositories which *in addition* perform a structured, externally reviewed and publicly available self-audit based on DIN 31644/nestorSeal
- **Formal Certification** is granted to repositories which *in addition to* Basic Certification obtain full external audit and certification based on ISO 16363.

CoreTrustSeal and NestorSeal are acquired by reviewing their requirements, ISO 16363 by external auditing of the criteria. There are various websites where repositories are registered along with their seal status (certification, open access etc.), e.g., registry of research data repositories Re3data.org, the directory of open access repositories opendoar.org, and registry of open access repositories roar.eprint.org.

The testing of datasets for FAIR principles by DANS has led to the conclusion that most of the FAIR principles (if not all) can be applied on the level of the repository. Therefore, repositories have to be evaluated, namely on how Findable, Accessible, Interoperable and Reusable they enable their data holdings to be rendered. This requires extensive mapping of available certification requirements for repositories to FAIR principles of data. The example of mapping of repository requirements to FAIR principles of data for

[72] http://5stardata.info/en/
[73] https://researchdata.jiscinvolve.org/wp/2017/10/04/5-highlights-rdss-market-research/
[74] http://reusabledata.org/#who-we-are
[75] https://fair-dom.org/
[76] https://www.coretrustseal.org
[77] https://www.re3data.org

CoreTrustSeal[78] is given in Appendix: Mapping of CoreTrustSeal repository requirements to FAIR principles for data. Recently, FORCE11 has developed levels of compliance to help researchers choose a FAIR data repository in the form of FAIR Data Decision Tree[7980].

2.2.3. Software artefacts – Open (source) Software

Data and software are intrinsically connected. In fact, "data are completely inaccessible without software, unless the data takes the form of printed matter."[81]. Below some efforts related to the sustainability, openness and FAIRness of software produced by research activities are presented:

- Within the community that it has established, Software sustainability institute (UK)[82] tackles issues relevant to skills and training, recognition and rewards, career paths and reproducible research, making research software outputs and derivatives easier to be used and understood.
- FAIR Software? How can we make it easier to find, access, deposit and reuse software?[83] forms an exertion of applying the FAIR principles to the more technological area of research software outputs. Accordingly, usage of PIDs, proper metadata schemas and DOIs as well as deposition in a digital repository and license attribution, are some of the inevitable yet designated steps in achieving software FAIRness.
- Choose a license[84] was created by GitHub to assist developers in selecting an open source license to appoint their code with, ensuring proper re-use by others. Three main categories are suggested: the MIT license (permits use with attribution to the creator), the Apache license 2.0 (permits use with attribution to the creator securing patent rights) and the GNU GPLv3based (copyleft-share alike securing patent right). They were chosen from a vast range of other licenses that exist, because they are highly preferred and used by users. A similar effort for software license attribution is the Free/Libre/Open Source license selection wizard by John Cowan[85].

A new way of measuring openness: the Open Governance Index[86] goes beyond the open source licensing schemas to address governance issues related to openness of the open source project. More specifically, thirteen (13) metrics are proposed related to areas of Access, Development, Derivatives, Community Structure, all of them ensuring success in the long term.

[78] CoreTrustSeal is a newly founded certification seal. It is a collaboration of two formerly existing seals: Data Seal of Approval (DSA) and ICSU World Data System (WDS).
[79] https://www.force11.org/group/wp3decision-trees/decision-tree-prototype-fair-data-1-v02
[80] https://docs.google.com/document/d/100Yv2hd9yfrlOscrMiGLqwqW_VWLruxRGoMr68YstNg/edit
[81] A Conceptual Approach to Data Stewardship and Software Sustainability: Scientists in charge, with a little help from their friends. Patrick Aerts (NLeSC) and Peter Doorn (DANS). The Hague, 2017. https://dans.knaw.nl/nl/over/organisatie-beleid/informatiemateriaal/AConceptualApproachtoDataStewardshipandSoftwareSustainability_DEF.pdf
[82] https://www.software.ac.uk/
[83] https://figshare.com/articles/FAIR_Software_How_can_we_make_easier_to_find_access_deposit_and_reuse_software_/5620690
[84] https://choosealicense.com/
[85] http://vrici.lojban.org/~cowan/floss/
[86] https://timreview.ca/article/512

2.2.4. Monitoring Open Education

Open Education is widely known as the form of education which is typically offered online and where barriers such as admission or certification fees as well as academic corpora encumbrances are eliminated. Open Educational Resources and Open Textbooks have become very popular due to their benefits to students and the wider society/community. Initiatives have tried to capture some of their elements and implementation procedures to understand their value and get a better view of current state and evolution of practice.

In this section, resources show transparency and openness issues in OER while realising that training is part of education procedures/methods and one of the most valuable source of achieving OS expertise.

2.2.4.1. Open Educational Resources

Open Educational Resources is of these initiatives where monitoring elements are already represented in developed tools like the OER World Map and the OER impact Map. The prior, i.e. OER World Map[87], uses as main variables organisations, services, projects, people, events, stories, publications, tools to then measure "Entries by type", "Entries by secondary education sector", "Projects by funders", "Top 5 Service Languages", "Top 5 Countries", "Services by License", "Services by Topic", "Top 5 keywords", "Projects by participant's countries", "Services by Audience", "Entries by primary education sector". Furthermore, OER impact map[88] consists of hypotheses that have been selected in such way to capture perceptions on how OER may be influencing issues related for example to political change or democracy in education, but also to outline the idiosyncrasy of users when selecting an open educational resource. In providing statistics, the map takes into consideration the educational sectors of School-K12, College, Higher Education and Informal (meaning individuals not necessarily falling in a category or enrolled in a programme, such as enthusiasts and individual lifelong learners).

Other tools that assist open courses creation, dissemination and re-use, are OER commons website. It falls under the William and Flora Hewlett Foundation's worldwide OER initiative which has created a platform intended to be used for the needs of both learners and educators. It also aims to create more collaborations and partnerships that will lead to innovative research and advancements around the subject. OER commons website[89] provides a space to build and take up open courses which are then categorised by subject area, grade level, material type. Information about the Conditions of Use, Content source (provider), primary user, media format, educational use and language are included. Among other websites accommodating open education needs is the Educational Platform of STEM4Youth[90], a Horizon 2020 funded project with open licensed materials and an upper goal of familiarising students with science and technology.

[87] https://oerworldmap.org/
[88] http://oermap.org/
[89] oercommons.org
[90] https://olcms.stem4youth.pl/discipline

2.2.4.2. Open Educational Resources

In addition to ethical and moral aspects, three topics are important to turn FAIR data principles to reality; namely training, workflows and metadata as discussed during the FAIR data community call in summer[91].

Trainings: To teach FAIR skills to EOSC users, the trainings themselves need to be FAIR.[92] Many attempts are being made to address the demand in the community[93] for a broader standard which applies to all major aspects of training and training materials. To date, several projects exist to create standards for individual aspects of training: In the scope of EOSC, the GO FAIR Initiative aims at creating the cultural and technical requirements for both building and sharing training materials. Wp7 EOSCpilot (Skills and Training) in its deliverable D7.1[94] describes possible competence models (e.g. EDISON Data Science Competence Framework) to fill the data stewardship gap. In their deliverable D7.2, training materials focused on the required skills to use EOSC infrastructure.

Furthermore, in the current landscape, part of few Open Science trainings that offer licenses to enable re-use are courses located in the materials section of the ELIXIR website[95].

2.2.5. Monitoring Research Collaboration/ Scholarly Communication

Open peer review

Peer review was broadly known as the process that validates and promotes researchers' success by giving them the opportunity to feature their work in scientific books and journals. That perception started to change with Green Open Access that made it possible for everyone to deposit at least a pre-print of their research irrespective of confirmation/ publication status. However, it is worth noting that open peer review goes beyond open access to publications to enable **transparency** in the process itself. Up to recent years, peer review was performed behind publishers' closed doors and researchers couldn't acquire much information about the procedure. Introduction of open peer review made it possible to identify and document those steps, spread motivation to the academic community which has enthusiastically responded positive to calls by accelerating in such way collaborations and new achievements. Nowadays, there is even the option of transferring information about a review from a journal where submission wasn't accepted to another journal.

Recognising that this is a very popular subject now which generates new ideas and innovative approaches every day, what is provided here is a brief overview of existing efforts showing open elements found in that context.

Open peer review platforms, services and tools

PeerJ[96], Publons[97] and F1000[98] are of the very well-known and widely used platforms offering open peer review services. Part of PeerJ's capabilities are post-publication Review History[99] with reviews state of

[91] European commision High level expert group : FAIR Data community call, 26 July 2017
[92] D7.1 WP7 EOSCpilot
[93] Five priorities for GO-TRAIN implementation Network (IN), draft of the outcome of the workshop hosted by CODATA at the international Council for Science in Paris on Friday 3 February 2017
[94] https://eoscpilot.eu/sites/default/files/eoscpilot-d7.1.pdf
[95] https://tess.elixir-europe.org/materials
[96] https://peerj.com/
[97] https://publons.com/home/

openness lying to the reviewers that have signed the report however the more decisive action of report's publication is taken by authors themselves. Otherwise, Publons offers the possibility of setting up a Publons account to deposit or link all peer-review activity in order to keep track of it. It allows tracking of citations and altmetrics for papers reviewed by individuals as well as enabling them to show interest to journals that they wish to perform reviews for. F1000 openly publishes peer reviews alongside research outputs and appoints CC-BY licenses and DOIs to all referee reports[100] making them reproducible and citable.
Royal Society[101] also attributes CC-BY licenses to referee reports and has developed scenarios according to authors', referees' and review reports' openness.

Moreover, a module for Open Peer Review[102] (OPR) was developed with the intention of it being integrated in institutional repositories to facilitate peer review practices for the long-tail of science.

Peer review in the mainstream

OPR protocol[103] renders a framework for the implementation of open peer review where requirements permit authors, among others, to submit their article to OPR protocol compliant repositories and reviewers to publish open licensed reports that are citable via the OAI-PMH. Towards achieving peer reviews citability, Crossref has introduced a metadata schema for peer review[104], to answer the community's popular demands while also making peer reviews discoverable and creditable.

Components of Open Peer Review

The systematic review performed by Tony Ross-Hellauer showed differences in perceptions of what constitutes open peer review, the normalisation of which form a useful resource as it led to identification of OPR traits.[105] Those are "Open identities" (authors and reviewers know one another's identities), "Open reports" (referee reports published together with other research outputs), "Open participation" (not exclusively performed by editors but rather it's open to comments from the public), "Open interaction" (communication between reviewers and authors), "Open pre-review manuscripts" (pre-prints), "Open final version commenting" (comments on final versions of publications/post-publication review), "Open platforms (decoupled review)".

Todd Carpenter[106] on the other hand, examined peer review guidelines to provide a set of elements constituting the different phases of data peer review (including data papers). Four (4) are the main areas concerning those phases: "Editorial review" checking quality and relevance values of data with respect to journals scope, "Metadata review" relevant to metadata quality, completeness, conformance to standards and datasets DOIs, "Data review" according to data re-use, units of measure in the datasets, data formats' consistency, and "Methodology behind creation of datasets" to tackle data collection methods'

[98] https://f1000.com/
[99] https://peerj.com/benefits/review-history-and-peer-review/
[100] https://f1000research.com/about
[101] http://rsos.royalsocietypublishing.org/open-peer-review
[102] http://www.openscholar.org.uk/institutional-repositories-start-to-offer-peer-review-services/
[103] http://www.openscholar.org.uk/wp-content/uploads/2013/07/OPR-protocol.pdf
[104] https://www.crossref.org/blog/making-peer-reviews-citable-discoverable-and-creditable/
[105] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437951/
[106] https://arxiv.org/ftp/arxiv/papers/1704/1704.02236.pdf

inaccuracies and replicability issues. Other Review criteria are "link to public repository", "descriptions of how to access data", "citations to other relevant materials", "ethics of experimentations", etc.

Data and software citations

Citation data are important because among others they are primary sources containing information about provenance, but their structure does not always allow machine-readability especially if they are to be combined with bibliographic data. For many years, reference data were provided by Crossref but not made open by the stakeholders. The initiatives that follow have been exploring ways to achieve those data becoming machine-readable, open and independent from the research results that they are tied with, ensuring availability even if research results themselves are published under closed access.

The Joint Declaration of Data Citation Principles[107] provides a set of recommendations that interested parties could use and extend to develop tools that support data citation activities similar to publications, namely "Importance", "Credit and Attribution", "Evidence", "Unique identification", "Access", "Persistence", "Specificity and Verifiability", "Interoperability and Flexibility". Open Citation corpus (OCC)[108] captures some of these principles since it is an open repository containing citations of scholarly publications that are made open and available for re-use after they have been appointed a public domain license (CC0). For license distribution purposes, OCC is also used by the Initiative for Open Citations (I4OC) by both open access and subscription-based publishers. The ultimate goal of the I4OC is citation data availability, visibility of research outputs and interlinking. Similar efforts have been made available for software citations, too. Principles[109] were modified by FORCE11 and DataCite recently released their updated schema[110] supporting software citation needs.

2.2.6. Monitoring Citizen Science

"Citizen Science refers to the general public engagement in scientific research activities when citizens actively contribute to science either with their intellectual effort or surrounding knowledge or with their tools and resources." (Socientize Project, 2013)

Citizen Science, otherwise known as crowd-sourced science, activities vary from granting parts of one's own property to the state for scientific purposes to contributing code and knowledge to research projects such as OpenStreetMap[111] or the British Library's Labs[112].

Citizen science projects and citizens' engagement[113]

More specifically, there are three dimensions showing the intersection of citizen science and policy: level of geography (local community, city, regional, country, continent), policy application areas (which differentiate between public policy and policy that facilitates citizen science), level of engagement and the type of citizen science activity. Particularly for the latter, the scenarios identified according to the means

[107] https://www.force11.org/datacitationprinciples
[108] http://opencitations.net/
[109] https://www.force11.org/software-citation-principles
[110] https://blog.datacite.org/metadata-schema-4-1/
[111] https://www.openstreetmap.org/#map=6/38.359/23.810
[112] http://labs.bl.uk/Engaging+the+Public+in+Linking+British+Library+Data+via+Games+with+a+Purpose
[113] https://www.wilsoncenter.org/sites/default/files/Citizen_Science_Policy_European_Perspective_Haklay.pdf

and reflected participation in conducting citizen science are: *passive sensing* where individuals' participation doesn't require any effort (e.g. sensors in smartphones), *volunteer computing, e.g. memory and power grants*, *volunteer thinking in collaboration with scientists*, *full-scale environmental and ecological observations, participatory sensing, and civic/community science with active contribution and participation to developing scientific tools and methods*.

There are several approaches to assess quality of citizen science projects. *Crowdsourcing* ensures quality of results by taking into consideration the amount of responses received per specific target while, in addition to that, *social approach* requires more information of the participants involved in the project so that more experienced individuals can supervise less experienced ones. *Geographic approach* endorses practices based on pre-existing verified data of a certain geographical area and *domain approach* builds on top of that by applying domain specific knowledge. *Instrumental observation* factors into/weighs the tools utilised to make an assumption or to produce results and, lastly, *process-oriented* practices are being performed by citizens that have undergone specialised for the occasion training.

2.2.7. Research Impact

Research impact is multi-dimensional. It concerns sectors which research uptake has influenced, like impact on society, on economy or political impact while it also involves dissemination activities that have been undertaken beyond the traditional environment of academia (e.g. altmetrics). This section refers to a collection of practices and indicators characterising research impact in the aforementioned forms. Additionally, section 2.2.8 Monitoring Open Government Data, which reflects the more mature area of Open Data Monitoring, has also inspired some of the selected impact indicators of the EOSC OSM.

How to track the impact of research data with metrics[114] forms a very useful guide with information on impact metrics and tools assisting their implementation. It is intended to be used by stakeholders who would like to assess their institutions' data impact.

Metrics Toolkit[115] is a very recent initiative which combines indicators with specific application processes concluding to a set of proposed research impact metrics. Metrics are categorised by type of research artefact, these being book chapters, books, datasets, journals articles and software/code/scripts. Examples of relevant metrics are blog mentions, journal acceptance rate and policy mentions.

However, Dimensions.ai[116] provides metrics that intend to measure research impact based solely on publications citations. These metrics are: "Citation Counts" for researchers' publications that have been cited by others, "Field Citation Ratio" and "Relative Citation Ratio" for citations comparison, "Citation Recency" showing uptake of research in two years span, "Highly Cited Indicator" for trending citations, "H-index" used to capture researchers performance and academic impact, "Annual Citation Rate" for journal papers citations collection yearly and "Altmetric Attention Score" mainly focusing in mentions received for research related outputs.

[114] http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics
[115] http://www.metrics-toolkit.org/
[116] https://figshare.com/s/c80c8f63d71cf89c0a90

2.2.8. Monitoring Open Government Data

This section provides a brief overview of existing and mature efforts for monitoring open data. These efforts were the first to tackle the problem of principled monitoring of open science and can thus be used as evaluable reference models, methodologies and implementations that can guide, and possibly complement, the specification and implementation of the EOSC OS Monitor.

Open Data Monitor[117]. The Open Data Monitor platform is designed to offer information to visitors about open datasets. It provides an overview of existing and available open data resources, and enables analysis and visualization of data catalogues. It works by harvesting and harmonizing data from external open dataset sources to offer visual and analytical insights about the composition of open data repositories on different regional levels (e.g., regional, national, European). It aims to provide analytic functionality for tasks such as assessing the quality of metadata, comparing different sources, providing licensing information, scoring the openness dimension and providing the data in various open formats, among others.

Open Data Watch[118]. Open Data Watch is an initiative that aims to provide data and statistics for the facilitation of planning, monitoring and evaluating the results of economic, social, demographic and environmental programs through the use of open public data. This allows monitoring of the progress over time for developing countries regarding their open data quality and openness.

Open data barometer[119]. The Open Data Barometer (OBD) aims to assess the impact and prevalence of open data initiatives in 115 jurisdictions around the world. With a strong regional focus, it is becoming a global policy making tool that evaluates governments and their open data initiatives based on three aspects, namely (i) readiness, (ii) implementation, and (iii) impact. Following a well-defined methodology, the Open Data Barometer issues the OBD global report, a summary of the most important and insightful findings.

Global Open Data Index[120]. The Global Open Data Index (GODI) is provided and maintained by the Open Knowledge Network. It is an annual global benchmark for publication of open government data, that uses a crowdsourced survey-based method for assessing measures and indicators for the openness of government data. GODI aims to create valuable information and insights to drive self-assessment over open government data publishers, as well as inform the general public.

European Data Portal[121]. The European Data Portal collects metadata of Public Sector Information that is readily available on public data portals across countries in Europe. It uses a series of indicators and metrics that measure the maturity of open data portals, covering the assessment of national policies, quality of available open data features, and the impact of the open data initiatives. It provides intuitive visual information, country rankings and overviews on open data readiness, policy use, impact and maturity on the national level. It includes measures such as usability and reusability of the data, policy presence, as well as social, political and economic impact, among others.

[117] https://opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex
[118] http://opendatawatch.com/monitoring/
[119] http://opendatabarometer.org
[120] https://index.okfn.org
[121] https://www.europeandataportal.eu/en/dashboard

OECD OUR Data Index[122]. The OECD OURdata Index assesses and evaluates the efforts and actions performed by governments towards implementing open data. It addresses three critical areas, namely (i) openness, (ii) usefulness and (iii) reusability of open government data. The employed methodology harvests data from sources such as public business registers, patent information, public tender sources, social data and legal data from member countries and focuses on government efforts to ensure availability and accessibility of public sector data. The methodology of the OECD OUR Data Index is based on the guidelines of the G8 Open Government Data (OGD) Charter.

[122] http://www.oecd.org/gov/digital-government/open-government-data.htm

3. PUTTING OPEN SCIENCE INTO A MONITORING CONTEXT

3.1. Approach and methodology

Monitoring and evaluating the advancements, trends and impact of OS in Europe is recognised as one of the most important steps towards the realisation of the EOSC vision. The **EOSCpilot Open Science Monitor Framework** (**EOSCpilot OSMF**) aims to build a model and initial high-level specifications for providing useful analytics to researchers as well as enabling research performing and funding organizations within EOSC to monitor and gain insights about the OS movement, regardless of their service management systems and the technology behind them. Moreover, it aspires to become a dynamic tool in the future, with benefits both to the organisations using it to measure the OA levels of implementation and impact to their community and for the High Level Stakeholders which are the European Union's Legal Entities and Bodies, including the Member States and their respective Units. Gaps, implications and new ways of performing OS are among the elements that could be identified through EOSC OSM which could then be easily incorporated within the stakeholders' scope and strategic planning for OS.

A prerequisite for the design of the framework is the definition of its objectives - goals and the identification of its core elements, i.e., what should be monitored and how. In this section, the methodology followed for deriving these concepts is presented and an overview is shown in Figure 3. The proposed approach is a methodological path to be used during the implementation of the framework, which can guide the initial design as well as any further refinements and extensions of the framework. It provides the steps that should be followed by an organization for deriving and choosing high level monitoring targets, identifying the OS elements to be measured, mapping monitoring targets to specific indicators and processes for data collection. In what follows, the core steps of the methodology are presented:

Step 1. Identification of the Open Science Activities: The first step addresses the need to identify which parts of the OS lifecycle are of interest in the monitoring process. These can include the conceptualization of a research task, the data and literature collection, the analysis and development of the research output, the publication, the review and evaluation of the research result as well as the reuse and reproducibility of results by the scientific community. These phases entail different open access practices and elements which are being considered by the monitor, with a special focus on the policies that these elements fall under.

Step 2. Policy-driven derivation of monitoring targets: The proposed monitoring framework adopts a policy-driven approach for deriving high level objectives, i.e., target dimensions to be measured in the monitoring process. Policies on Open Access at different levels, such as the international, national and regional levels, as well as micro policies are considered as primary sources for deriving more concrete measurable targets (e.g., *Openness*, *FAIRness*, etc.) that should be monitored in the framework. Furthermore, monitoring targets can be organised into more specialised sub-targets, to measure more specific aspects of OS. For example, a policy recommendation stating that research data repositories should follow a data archiving plan indicates the monitoring target for long-term preservation of OS artefacts monitored by a set of indicators, such as whether an organization applies such a plan, or the period (e.g., months, years) for which preservation is guaranteed.

Step 3. Identification of the main Open Science Resources and Indicators: In the next step, the monitoring targets are being mapped to OS elements they apply to, as well as to indicators that quantify these targets. OS elements are well-defined artefacts of OS practices, such as publication in open access journals, research data made available in open access repositories, open source software, open educational

material, etc. In the context of this framework, these elements are called OS **resources**[123] to state the importance of their contribution for the development of the "Open Science World".

Step 4. Design of monitoring processes, tasks and workflows. Each indicator must be associated with a set of processes, which are employed for the collection of data, the validation and scoring of metrics (e.g., combination and aggregation of metrics for deriving an accumulated score for a target dimension), the visualization of the results, and so on. These processes must be well documented in the form of workflows and tasks, to be performed for the collection and quantification of the indicators.

Step 5. *Modelling and implementation of the framework*. The next step involves the detailed design, implementation, and customization of the framework, which includes the design of the functionality at its whole, as well as all added value services offered by the framework.

Step 6. *Continuous validation of the monitoring targets*. The last step follows the operation of the OS monitoring framework and refers to the continuous validation and refinement of the monitoring methodology (i.e., targets and indicators) and results in EOSC. A monitoring process, to be effective, must adapt to new OS practices and new policies, validating and readjusting its target goals as well as the indicators for their evaluation.

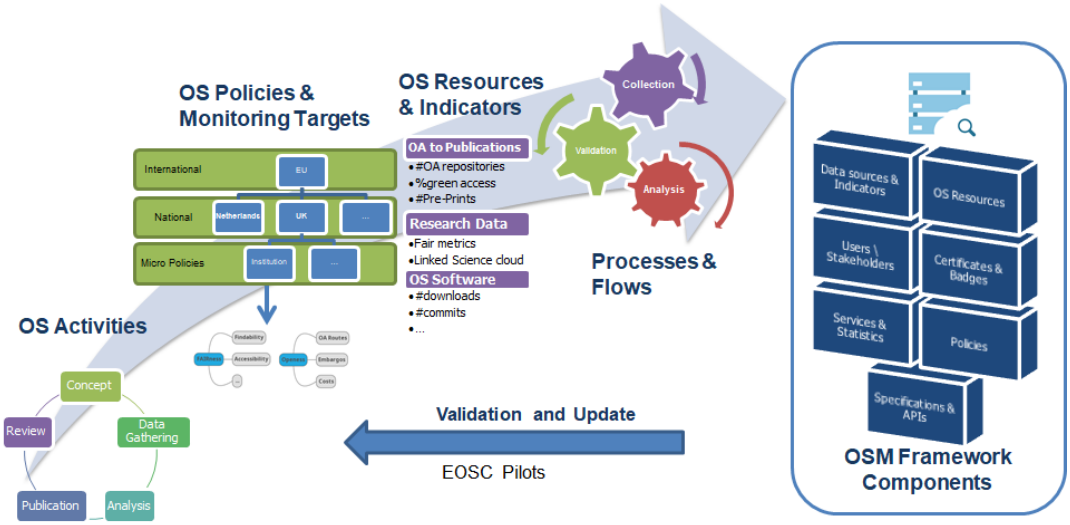

Figure 4: OS Monitor Methodological approach

The next sections describe the application of the methodology for the potential implementation of the OS Monitor in EOSC. Namely steps 1-3 are applied in Section 3.2, 3.3 for deriving an initial set of *Monitoring Targets* and *Monitoring Indicators*, i.e., a first categorization of the targets - sub-targets and their mappings to indicators. The indicators are selected from the related efforts and tools presented in Section 2. Steps 4 and 5 are analysed in Section 4 regarding the processes, the initial design of the specifications and the offered functionality of the framework.

3.2. Monitoring Targets and Dimensions of the EOSC OSM Framework

Monitoring targets capture high level aspects of the OS practices and trends and thus they are the main measurable dimensions of the OSM. Many of the efforts presented in Section 2 consider such high-level

[123] Also see definition in glossary.

targets and propose indicators that quantify them. Based on this literature review, the current section identifies commonalities and differences across these approaches and incorporates the various targets, indicators and associated stakeholders into a coherent monitoring framework. Following the methodology, the output of this activity, i.e., an essential part of the OSM, is a taxonomy of targets and sub-targets, each of them addressing a different stakeholder group. Each target / sub-target is mapped to a set of indicators that provide quantifiable results (e.g., a score).

Aligned with most existing efforts, this categorization primarily considers aspects of openness, FAIRness, trustworthiness as well as the impact of OS in:

- research artefacts: mainly publications, research data and software;
- educational resources;
- research collaboration;
- citizen science practices.

On this activity, vital information came also from policies both from *D3.1 Policy Landscape Review* and from model micro-policies (RECODE, PASTEUR4OA, etc) which was then matched to the OSM Monitoring Targets (MTs).

Putting together the “building blocks” containing elements that the EOSC Open Science Monitor Framework is comprised of was not a trivial task, as each block represents a Monitoring Target and/or sub-targets, encountering specific policies (mandates) and regulation issues, which are most of the time by definition expressed and implemented differently from stakeholders. Furthermore, as already mentioned in Section 2.2 “Monitoring Open Science data - FAIR approaches” there are many correlations between Open (Access) and FAIR, with one of the main distinctions between them being that FAIR is designed to ensure machine readability and reproducibility, while Open Access concentrates on costs, eligibility and ethical issues to make something open. Overlaps were evident or became apparent while studying existing monitoring tools and approaches that support Openness and FAIRness measurements. In this respect, a first review enabled the identification of contextual and semantic differences and overlaps, and allowed for a first categorization of the targets, which was further refined based on the policies described in D3.1 Policy Landscape Review. The outcome of this procedure was a first taxonomy of the EOSC OSM Monitoring Targets, sub-targets and the associated dimensions that facilitate them. It should be noted that this categorization can be further extended and refined with more specific goals during the implementation of the framework addressing different stakeholder needs.

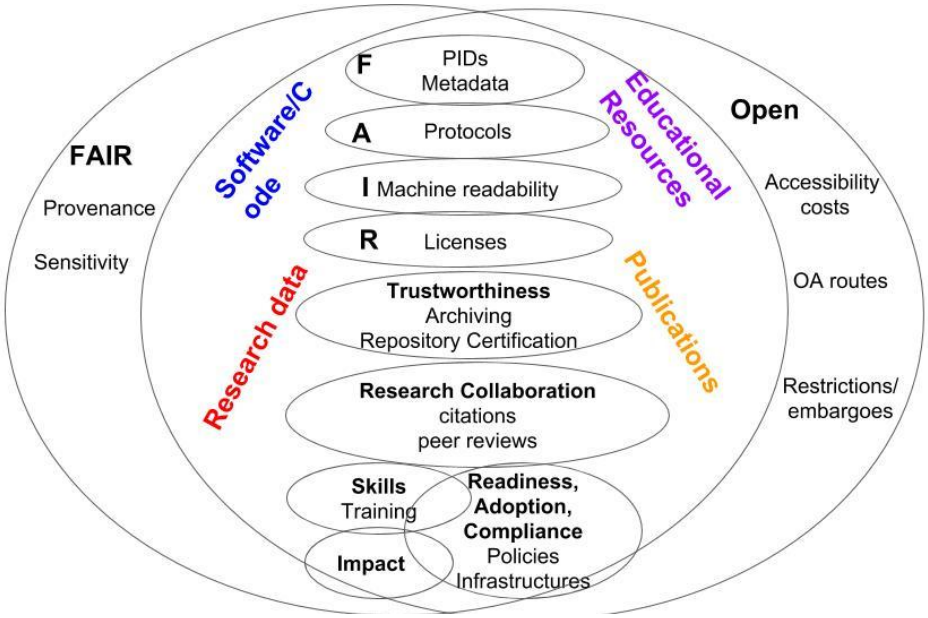

Figure 5: Monitoring Targets considered in OSM

Figure 4 consists of two circles, one containing basic elements/dimensions that characterise FAIR principles and the other concerning aspects of Open Access. Additional elements that are included in the figure deal with complementary issues, such as practices (e.g., peer review) and policies (e.g., skills). The middle part of the figure, the intersection of the two circles, highlights more prominent elements that accommodate common targets in both FAIR and OA spectrums. These common targets are displayed in bold and in the EOSC OSM Framework are the primary **Monitoring Targets (MTs)**; more specifically these are: **FAIR**, **Open (Access)**, **Research Collaboration/Scholarly Communication, Trustworthiness, Skills** and **Readiness**. The fact that, in the common circle, FAIR is separated by/per principle and is not represented as a whole, was purposely chosen to better reflect interconnection between dimensions (findability, accessibility, interoperability, reusability) and elements (PIDs, licenses etc) of FAIR and Open Access. Unique dimensions are kept outside of the common circle acknowledging disparities. The following subsections 3.2.1 - 3.2.4 explain the monitoring targets and sub-targets in detail.

3.2.1. FAIR Principles – Measuring FAIRness

FAIR by definition demonstrates **Findability**, **Accessibility**, **Interoperability** and **Reusability** of research artefacts. Particularly Accessibility and Reusability correspond to the open access principles concerning access and reuse, though the sub-principles that they highlight make more prominent commonalities and differences with Open Access. To name a few:

- **Findability** (FAIR) similarly to Open Access, concerns PIDs, metadata, metadata descriptors and open directories presupposing that information is indexed and searchable. This concerns also locus of deposit as *visibility* is increased when data are included in repositories that themselves belong in registries such as re3data.

- **Accessibility** in the FAIR context has to do mainly with technical provisions like mechanisms or **protocols** ensuring access to data but even with human-centric factors like that of scope and data **understandability by users**. In the context of Open Access, Accessibility depends more on the openness of these technical

provisions which translates to requirements not broadly for protocols but for *open* protocols. Moreover, Access on this occasion is closely associated with **location of deposit and/or publication** (full open access journal, repository etc) as well as with **access costs** characterising/underpinning those locations and with **embargo periods** that may apply to specific disciplines.

- **Interoperability** is a vital principle that puts together different elements in order to further define them to set the requirements that make them understandable; technically speaking, interoperability is otherwise expressed as **machine readability**. Interoperability concerns among others metadata completeness, formats of research artefacts, as well as standards that need to be in place for integration purposes.

- Equally, **Reusability** in Open Access is interwoven with *open* **licenses** as well as copyright restrictions or exceptions to mine text and data (terms and conditions also considered) while in FAIR, machine readability of standard licenses and inclusion of **provenance** as a prerequisite/ requirement complement the aforementioned.

3.2.2. Open Access – Measuring Openness

In measuring openness of Open Science Resources, there are two main facilitators: **OA costs** and **licenses**. Embargoes causing delays in making research outputs available to the public as well as restrictions in mining content are also considered here, similarly to Accessibility from FAIR. Cost measurements revolve around expenditure to make research results and artefacts open (includes routes of OA - especially for Gold and Hybrid OA where APCs and membership fees are taken into account), expenditure to gain access to research results and artefacts, eligibility and type of costs that will be reimbursed by funders as well as costs ensuring infrastructures or service development and maintenance. IPRs and copyrights, as expressed in licenses as well as transparency of procedures and activities such as *peer review,* are additional factors that accommodate openness needs. More on transparency in research procedures is examined in section 3.2.4 Research collaboration/ scholarly communication.

3.2.3. Long Term Preservation – Measuring Trustworthiness

There was a debate as to whether trustworthiness could be claimed as a Monitoring Target, since it seems to have a less strong hue. However, it was decided to be included to give credence to such criteria in services (e.g., repositories). In addition to the above, this monitoring target combined with FAIRness of research artefacts is a valuable source in exploring long term preservation in the context of FAIR. Thus, what the EOSC OSM measures here is:

Archiving - to identify platforms (locus of deposit) that researchers use to archive their data (Institutional, subject repositories, libraries) and others that have provisions and mechanisms for long term preservation (back-ups, sensitive data etc).

Certification - to check solidity according to relevant certifications, such as the recently formulated CoreTrustSeal and ISO 16363.

3.2.4. Policy – Measuring Policy Dimensions

This section focuses on policies, from the early stages of their development to adoption by stakeholders concerning even later phases of compliance. Therefore, the Monitoring Targets to address such issues were formed as follows:

Readiness - Measuring preparedness. This Monitoring Target gathers aspects that could be used to assess preparedness to welcome Open Science policies developments as well as to accommodate technical needs with infrastructures and services deployment.

Policy Adoption. The first phase of EOSC OSM touches upon issues that could more accurately be described as application measurements. Currently, this Monitoring Target aims to discover the level of policy that the stakeholder is subject to (e.g., European, national, institutional) along with information on the type of policies (data sharing, open access, research data management, etc.) that have been developed by the stakeholders and level of commitment in following them (e.g., mandates, strategically anchored).

Policy Compliance. Compliance measures are expected to be deployed during the work of the Open Science Policy Registry and technically facilitated by Open Science Toolkit respectively. This is an additional step of the OSM activity, an inevitable extension for monitoring completeness, to be earmarked in parallel with and complementary to D3.4, D3.5 work which will be incorporated and promoted in one of these deliverables.

3.2.5. Research Collaboration/ Scholarly Communication – Measuring transparency and dissemination

EOSC OSM takes under consideration different phases and activities that compose the process of peer review and focuses mainly on their ***transparency***. Therefore, openness of these entities is examined (e.g., open identities characterising open peer review versus more closed, traditional approaches like these of single or double blind review). Regarding **citations**, the type of research resource is recognised (code, data, bibliographic etc.) and open initiatives like the i4oc for open data citations allow comparisons with respective closed ones. **Communication and collaboration activities** among stakeholders and initiatives organised to promote data related work are also represented under this Monitoring Target.

3.2.6. Open Education – Measuring Impact and Open Educational Resources (OER) uptake

EOSC OSM focuses on teaching materials or virtual teaching environments and tools, like MOOCs, to evaluate openness while attempting to measure their FAIRness as well. Identification of topics, publications and curricula about OS are included in measurements about **OER** along with information that leads to assessing teaching and learning experience.

3.2.6.1. Skills – Measuring Expertise and Uptake in training

One of the four policy layers of *D3.1 Policy Landscape Review* is skills. It is an important component that, combined with metadata and workflows, is considered to make openness a reality. Here the types of skills necessary to perform OS are understood (web technologies, data science, legal aspect etc), and **training** is

examined regarding literacy programs and people's participation showing expertise and uptake in these activities.

3.2.7. Citizen Science – Measuring citizen engagement

This Monitoring Target concerns specific aspects of citizen science engagement, but the actual performance and implementation of OS by citizens is not included in measurements. Instead what is represented is transparency in sharing research beyond Academia, demographics of citizens participating in research related projects or enabling parts of them.

3.2.8. Impact – Measuring research impact on society and economy

The endeavour of the EOSC OSM in measuring the impact is limited to providing an overview of specific aspects of the **science/research, economic and public/community sector** as well as capturing elements of **research excellence** (mostly indicators on rewards and incentives). Part of this section's measurements are driven by more mature monitoring mechanisms, those used in the "Open Data world". It is worth noting granularity in such attempt, since to draw a concrete yet rather representative conclusion on that matter requires a span of years of OS exploration and implementation.

3.3. From Monitoring Targets to Indicators – What the EOSC OSM captures

This section describes how the EOSC Monitoring Targets can be measured in the context of EOSC, categorised per type of research artefact that they concern and enriched with interested stakeholders. As indicated in the previous section 3.2, these targets are limited to Openness and FAIRness of research artefacts, repositories' trustworthiness, transparency and dissemination of research collaboration/scholarly communication activities, open educational resources uptake, expertise and skills progression regarding open science, citizen engagement and research impact. Policies target preparedness, adoption and compliance of stakeholders over them, however for the latter, more structured information and concrete measurements will be derived by the work of the upcoming deliverables, i.e. the Open Science Policy Registry and the Policy Toolkit.

Monitoring Targets exhibit hierarchical structure; that is, an individual MT can be further specialized in one or more sub-targets (children) or generalized to more abstract super-targets (parents). In this sense, MTs are conceptually structured in the form of a tree, where each tree node represents an MT, and each edge represents a parent-child (or target-subtarget) relationship between two MTs. This can be seen in Figure 5, where the nodes colored in grey represent the MTs defined in this deliverable. Furthermore, in the lowermost layer of the tree, the leaves, colored in orange, represent potential indicators for their parent MTs. In the figure, several indicators are indicatively depicted on the leaf level.

Appendix A contains the detailed mappings between targets and sub-targets, stakeholders and indicators in tabular form.

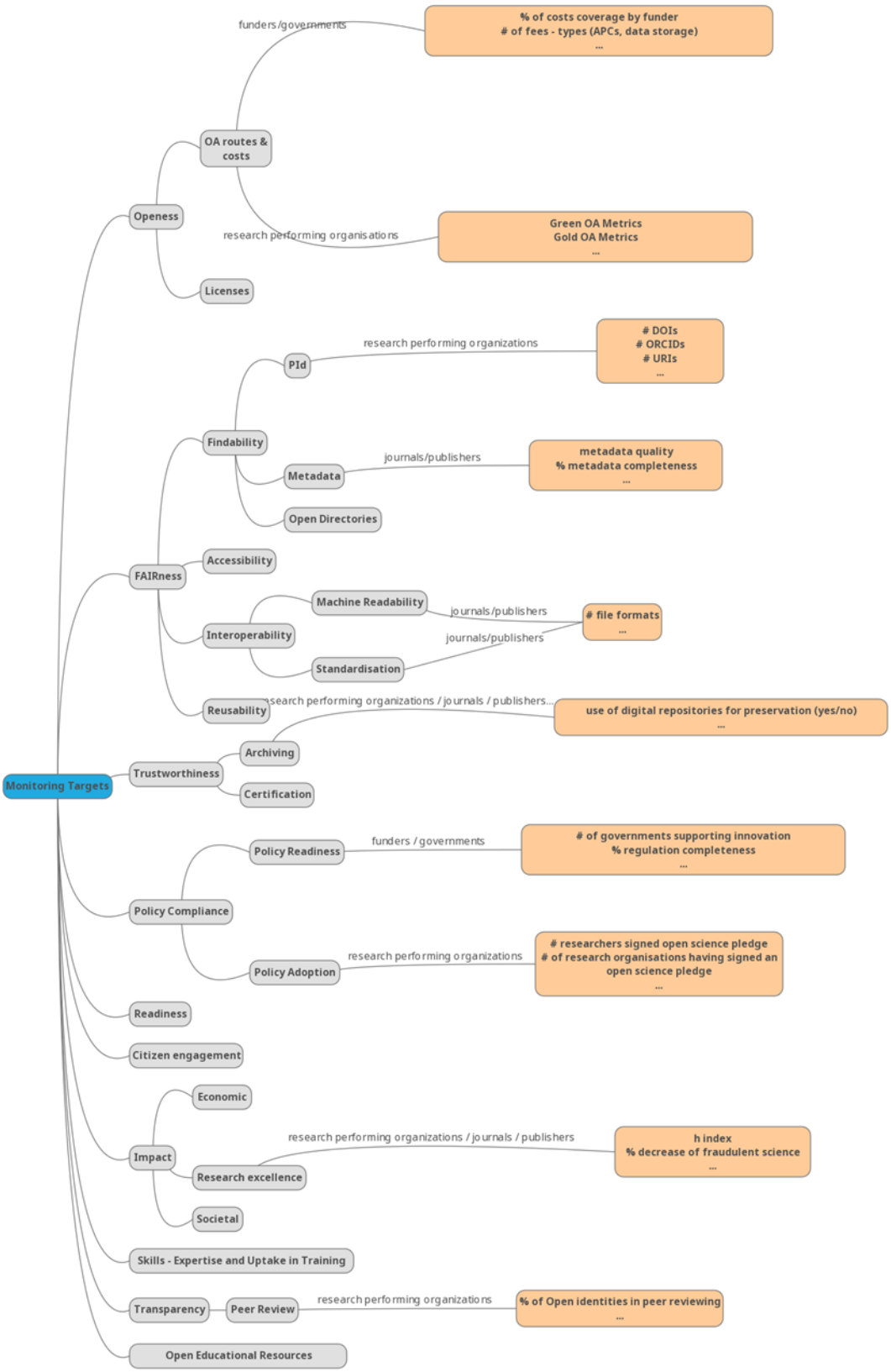

Figure 6: Categorization of monitoring targets, sub-targets and indicative indicators

4. THE EOSC OS MONITOR SPECIFICATION

In this section, the specifications of the OS Monitor are presented, comprising three core elements. The first element is the conceptual model for the OS Monitor, which includes all the core entities and their interrelationships, such as actors, research resources, monitoring targets and indicators. The second element is the OS Monitor processes and workflows, which offers the core functionality required to perform and orchestrate monitoring tasks. Finally, the third element is the OS Monitor services, which provides the offered functionality to the end users.

To define the scope and specifications of these three core elements, a brief requirements analysis is first presented, identifying functional and non-functional requirements for each of the three. Then, an overall framework architecture is provided, outlining the core components and modules, as well as the interactions between them. Furthermore, the basic flow of tasks for a monitoring process is given, and finally, in the remainder of the section, the specifications for the model, processes and services are presented and discussed.

4.1. Requirements of the OS Monitor

To design the specifications for the OS Monitor, a set of requirements first needs to be devised. These requirements cover both the functional and non-functional aspects of the OS Monitor, paving the way for its implementation, as they will ultimately form the basis for the basic technical elements of the monitor, namely the *OS Monitor Model*, the *Monitoring Processes and Workflows*, and the *Monitor Services*, i.e., the admin and end-user functionality that will eventually be exposed as a set of user actions on top of the monitoring framework. In what follows, the requirements for these elements are presented.

4.1.1. OS Monitor Model Requirements

The OS Monitor Model forms the representational foundation for the entities that interact within the OSM. In this sense, it must provide appropriate abstractions for all the components that are relative to the monitoring framework, spanning from users and their roles (e.g., admins, end-users etc.) to monitoring targets, indicators and governing policies. As a representation model, it should support several non-functional characteristics, that will allow it to be complete and in accordance with current requirements, but also extensible and interoperable with respect to external sources. These non-functional requirements for the model are listed as follows:

1. **Coverage**. The model must consist of all the well-defined entities that participate in a generic monitoring workflow, from the beginning stages, including conceptualization of a monitoring flow, definition of the context and purpose of the monitoring flow, to the final stages, including the definition of the expected outcomes, results and form of the output.
2. **Extensibility**. The monitor model and its constituent elements must be extensible and adaptable to future changes. This way, they become future-proof, and can be redefined and interchanged during different life-cycles.
3. **Expressiveness**. The representation model must provide appropriate abstractions that do not impose restrictions when instantiating its entities. This can be achieved by providing appropriate metadata and characteristics that can be used to define restrictions, rather than impose them. For

instance, limiting a measurable goal to having exactly one metric is an unnecessary restriction at the model level. However, providing attributes that allow defining the number of indicators per target provides more expressivity for the modeller.

4. **Interoperability.** The conceptual entities and their links, as well as the term vocabularies and code lists that instantiate these, must be modelled in a way that allows interoperability between the monitor model and external sources. This can be achieved by reusing appropriate ontologies and vocabularies that already exist and are publicly available, wherever possible.

Apart from this list of non-functional requirements, which is necessary to ensure that the OSM model is governed by these characteristics, there exist a set of rigid functional requirements that it should exhibit. These are concerned with ensuring that the appropriate level of knowledge capturing takes place, covering all necessary conceptual entities that must be represented, and is summarized as follows;

1. **Users and Actors.** The model must define elements that represent all relevant users that contribute to the monitoring framework, both actively and passively. These include users that provide input, users that consume the output, and intermediary actors that provide OS resources or are in any way affected by the monitoring framework.
2. **Processes and Tasks.** The model must define appropriate elements to represent the processes that take place within monitoring. For example, processes that collect data, compute indicators, and perform scoring actions must be represented by high-level notions of processing actions.
3. **Resources under monitoring.** A monitoring process is focused around particular artefacts, or resources, that are currently under evaluation, assessment or scoring and thus these should be covered appropriately by the OSM model.
4. **Indicators and Scores.** The output of monitoring is in the form of a quantified evaluation measure, either low-level indicators, or high-level scores. As these are extensible, flexible and customizable, as discussed in Sections 2 and 3, these should be properly represented with appropriate high-level entities in the OSM model.

4.1.2. Monitoring Process and Workflow Requirements

The OS Monitor depends on *Processes* and *Workflows* as structural elements of monitoring resources. A monitoring process is broadly defined as a series of sub-tasks/sub-processes that cover the monitoring life-cycle of a particular resource, from input collection to metric assessment and evaluation. Thus, for instance, the steps of (i) identifying a specific open science resource (e.g., an open data repository), (ii) retrieving its associated metadata, (iii) defining how to evaluate it as an open science resource, and (iv) measuring it, comprise a monitoring process. Combining several of such processes together and/or defining services with added functionality on top of these processes leads to the creation of workflows. In this context, in the following list, the *functional* requirements for OS Monitor processes and workflows are identified:

1. **Process decomposition.** The OS Monitor should be enable specification of sub-task components of a monitoring process. This includes defining the input, specifying data cleaning and harmonization, defining the generic dimensions (i.e., monitoring targets) as well as the exact measures (i.e., indicators) to be computed, calculating the indicators, and combining the results to form the output.
2. **Temporality.** This includes the ability to define time periods and intervals that a monitoring process

will operate. For example, a user might be interested in one-off monitoring of a specific resource, to include the results in a static report that is to be published once. On the other hand, the user might need to specify periodic executions of a process to create a time-series of monitoring results. This, a monitoring process must include temporal characteristics.

3. **Dimensionality.** A monitoring process must have specific monitoring targets, or dimensions, to form the basis of the monitoring assessment. When creating a monitoring process, the OS Monitor must provide the ability to attach monitoring targets. These can either be selected from an existing list or added ad-hoc.
4. **Relating targets to indicators.** Specification of a monitoring target is not enough, because to quantify the resource's conformance to the monitoring target, one or more concrete indicators must be assigned to the target. For example, to quantify the accessibility (monitoring target) of an open dataset, the process might measure the dataset's online availability via a URI (indicator), which can be true or false. Thus, when defining a monitoring process, it is required to assign indicators to monitoring targets.
5. **Adaptability.** The requirements for a specific monitoring process might change and evolve over time, while the scope and context of the process remains the same. In this sense, a process must be able to evolve and change along with the dynamic environment it is defined in. For example, a process monitoring the discoverability of a resource might require new and more refined indicators a few months after its initial implementation. Thus, it is required for a process to be able to change and adapt to dynamic contexts.
6. **Ability to combine monitoring processes.** As part of a larger effort to collectively monitor sets of resources and exploit the results, monitoring processes must be able to be combined into a single context. A set of monitoring processes can have its own descriptive metadata, and can be executed sequentially, in parallel, or in an indifferent temporal order.

4.1.3. OS Monitor Services Requirements

Services are defined as high-level functionality on top of the OS Monitor, that can be used to administer, customize, configure and consume (e.g. visualize, export etc.) monitoring resources, processes, workflows as well as the input and the output.

In this context, the set of **non-functional requirements** for the OS Monitor services are:

1. **Multi-tenant SaaS Architecture.** The OS Monitor should expose a set of appropriate APIs that cover the functionality for all components of the monitoring life-cycle, such as creation and configuration of processes and workflows, definitions of appropriate visualizations and export flows for the resulting output, subscriptions to monitoring processes, social engagement (e.g., sharing of results in social media) and so on. This should be covered in a SaaS manner, covering the specifications and requirements of different users (tenants), i.e. research administrators and front-end users.

2. **Accessibility and Interoperability.** The services of the OS Monitor should be interoperable, open and web-accessible in a light-weight manner that remains transparent with respect to implementation specifics (e.g., programming languages and frameworks etc.). An architecture that consists of open, RESTful APIs accomplishes these targets and is therefore a requirement for the OS Monitor. In addition, the OS monitor should be easily configurable to harvest information from various data sources. Interoperability with other monitoring tools and frameworks (as those presented in Section 2) for the exchange of information regarding indicators and results should be

enabled via the use of established standards and synchronization web frameworks (e.g., ResourceSync[124]).

3. **Reliability.** The services must exhibit a reliable online presence, without down-time or undocumented changes.
4. **Extensibility.** The services must be extensible to account for dynamically changing application contexts.
5. **Scalability.** With wide adoption, the OS Monitor will handle very large amounts of requests, as well as simultaneous execution of tasks and processes such as monitoring, visualizing, exporting and so on. Therefore, the framework must be designed in a way that enables scalability.
6. **Security.** The OS Monitor should provide appropriate security measures by defining and employing an appropriate security policy, which will span from authorization and authentication of research administrator actions, data protection and safe-guarding of the integrity of the data, as well as non-repudiation, to ensure that the monitoring processes are properly and safely instantiated and configured and prevent DDoS attacks and other malevolent practices.

The functionality of the OSM services targets different types of users. Specifically, two main user roles are targeted, namely (i) the **research administrator** responsible for defining, configuring, initiating and orchestrating monitoring processes, and (ii) the **end-user/consumer** (e.g., a funder, a researcher, etc), interested in retrieving and consuming the results and output of the monitoring processes.

For the **research administrator**, the functional requirements of the OS Monitor services are:

1. **Create monitoring processes.** The research admin must be able to create new processes and ascribe appropriate functional and non-functional metadata. Functional metadata include characteristics that are required for the execution of the process, and include defining the following:
    a. Resource(s) to be monitored.
    b. Monitoring targets and indicators to be measured.
    c. High level scoring functions for indicator combination/aggregation.
    d. Sources for data collection.

Furthermore, non-functional metadata include descriptive characteristics, such as tags, textual descriptions, labels and so on.

2. **Control execution of monitoring processes.** The research admin must be able to define appropriate conditions and definitions for controlling the execution of monitoring processes. These include the following:
    a. Initiate a monitoring process.
    b. Stop/Cancel a monitoring process.
    c. Set time interval and scheduling for execution of a monitoring process (e.g., run every three months).

[124] http://www.openarchives.org/rs/1.1/resourcesync

3. **Configure and administer monitoring processes.** The research admin must be able to edit the characteristics, definitions and metadata of a monitoring process at any point after its creation.
4. **Delete monitoring processes.** The research admin must be able to remove a monitoring process from the OS Monitor.
5. **Administer collaboration on existing processes.** The research admin must be able to invite other research admins to collaborate on shared monitoring processes, by enabling them to perform the same set (or possibly a subset, depending on rights and roles) of functionalities that he/she is entitled to as creator.

On the other hand, the set of services that the OS Monitor should employ for end-users is mainly concerned with exploration and consumption of the monitoring results. Specifically, the following functional requirements should be facilitated for **end-users**:

1. **Viewing monitoring results.** The end-user must be able to consume the output and results of a monitoring process in context-dependent, meaningful and interesting ways. To this end, the OS Monitor should provide a set of services for delivering content from the results of monitoring processes, to address human-readable scenarios of usage. These include:
    a. Textual reports, including plain HTML output.
    b. Visual reports, including graphs, charts and other meaningful visualizations.

Furthermore, the output must be retrievable in different levels of specialization/generalization, providing users the ability to see monitoring overviews, as well as more detailed reports. Finally, the OS Monitor must enable end-users to share content on social media and other relevant channels.

2. **Machine-readable accessing of monitoring results.** Aside from users being able to view and read monitoring results, it is often the case that further processing (e.g., programmatically) is required in order to exploit the output. To this end, a set of APIs and services should be provided that offer machine-readable access to the resulting data. To ensure machine-readability, appropriate services must be in place to convert the data to widely used, open and interoperable formats, such as JSON, CSV and RDF/XML.
3. **Combining and comparing output from different processes.** The OS Monitor should enable comparisons and combinations (e.g., aggregations) of results that stem from different monitoring processes. The resulting reports must be consumable in the same way individual reports are consumable, based on point (1).
4. **Personalization of experience.** The end-user must be able to experience personalized usage of the OS Monitor by being able to highlight specific monitoring processes as favorites, to follow their progress and evolution across time. Furthermore, the OS Monitor should provide appropriate services that enable end-users to “star”, subscribe, and allow alerts for news and cumulative digests of the monitoring results of their resources of interest.
5. **Searching and exploring OS Monitor elements.** The end-user must be able to search, browse and explore different types of OS Monitor elements, such as monitoring processes, stakeholders, research resources, monitoring targets and so on. Services include:
    a. Keyword search.

b. Filtered (advanced) search.

c. Faceted search.

d. Exploratory browsing, including trending items, recent feeds etc.

In what follows, the aim and scope of the described requirements will be further clarified by introducing an appropriate architecture of the OS Monitor, which puts all required functionality in place.

4.2. OS Monitor Architecture

In this section, an overview of the overall architecture envisioned for the OS Monitor is presented. The components and modules are defined based on the functional requirements presented in the previous section for the OS Monitor's processes and services. It should be noted that this is an envisioned, high-level architecture for the OS Monitor. Detailed architecture and specifications will be defined during the concrete design at the implementation phase of the OS Monitor.

In Figure 6, the overall architecture of the EOSC OS Monitor can be seen, along with the monitor's various components and modules. In the lower part of the figure, marked with orange, lies the **core functionality layer** of the OS Monitor.

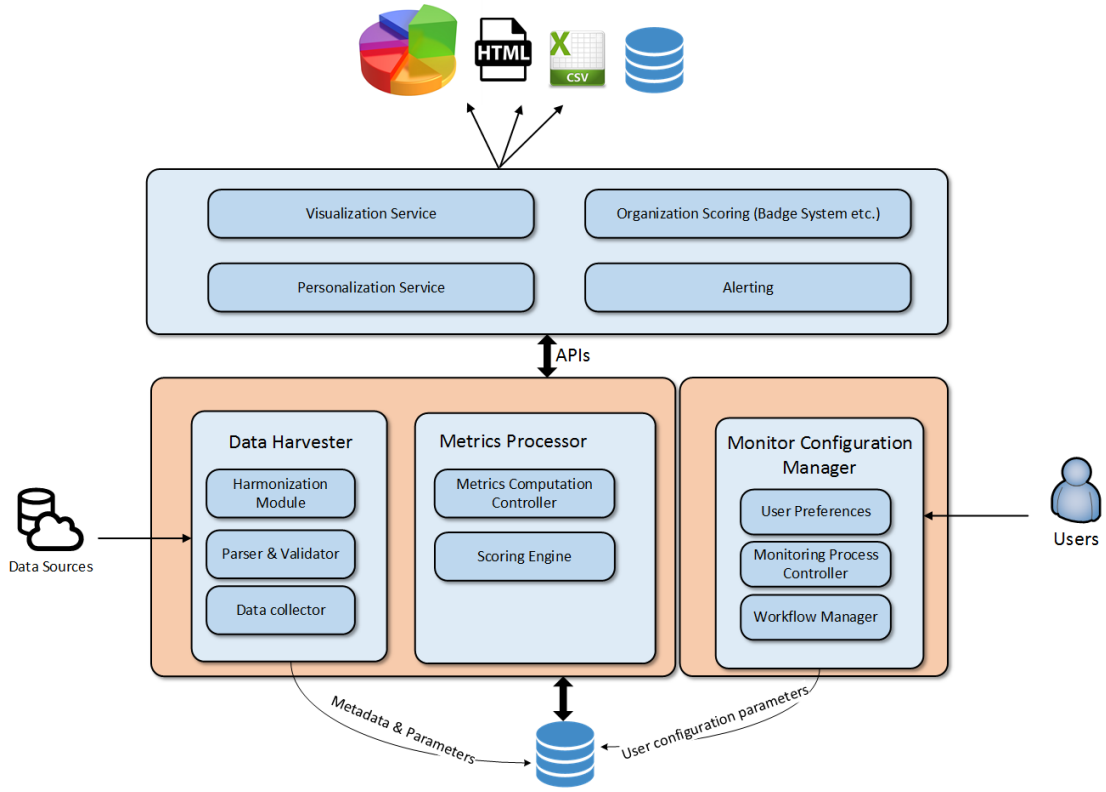

Figure 7: EOSCPilot OSM Architecture

This is divided into the following components:

Data Harvester. This component is responsible for data collection from external sources through the *Data Collector* module, as well as parsing and validation of the input through the *Parser & Validator* module and homogenization and harmonization of the validated input through the *Harmonization* module.

The input of the Data Harvester comes from external data sources, which can be seen at the left of the figure, and is handled by the Data Collector module, which includes the appropriate core functionality for ingesting data and metadata from external data sources using open APIs for handshaking and retrieval. OSM will make use of existing standards and APIs in order to be interoperable with established monitoring tools and platforms in the OS ecosystem for harvesting and processing data about indicators. For example, acquisition of data related to citations of open research publications could be retrieved from sources such as I4OC[125], oaDOI[126] and DataCite[127]. Data needed to assess OA repositories will be collected from sources such as OpenAire[128], OpenDOAR[129] and re3data[130]. Other sources, such as Open Access Button[131] and DOAI[132] will provide data for assessing journal article availability, while sources like Publons[133] and Peerj[134] will provide data regarding open peer review processes. Data collection process shall offer the ability for research administrations to easily attach new sources and make use of their APIs. In practice however, collecting metadata information from repositories is not a trivial procedure in real-world scenarios, as scientific communities often lack well-defined metadata models for their data. This will be taken into account accordingly to allow research administrators to manually harvest or define descriptive metadata that will assist data collection.

Moreover, the Data Harvester includes the appropriate core functionality for parsing the data and assessing/confirming the machine-readability and structure (i.e., validation) of the ingested/collected data from external sources. This is done with the use of the Parser & Validator Module.

At times, a certain degree of cleansing and harmonization of the input is required. This is especially the case when a research administrator initiates a monitoring process for a research resource that is described by data from multiple sources, or even from the same source but with varying formats and structures. For this reason, the Harmonization Module is used to handle data harmonization and homogenization tasks, with the aim of forcing a common conceptual model and structure to the ingested data.

Indicators Processor. This component is responsible for controlling the whole evaluation process, which includes computation of the defined metrics and indicators, normalization of the result, aggregation and combination of the computed indicators, and computation of mixed scoring functions based on the above. The *Indicators Computation Controller* module is responsible for computing the defined indicators. The task of computing one or multiple indicators on the collected data is one of the core functionalities of the OS Monitor. It handles the application of a given function or formula on a set of input data and metadata, to derive a specific result/outcome and map the indicator to this outcome for the measured resource. Furthermore, the *Scoring Engine* module is responsible for aggregations, combinations and other higher-level scoring functionality, which takes place after computing individual indicators. The actual scoring is calculated as a defined function with the individual indicators as the input parameters. For example,

[125] https://i4oc.org/
[126] https://oadoi.org
[127] https://www.datacite.org/
[128] https://www.openaire.eu/
[129] http://www.opendoar.org/
[130] https://www.re3data.org/
[131] https://openaccessbutton.org/
[132] http://doai.io/
[133] https://publons.com/home/
[134] https://peerj.com/

averages, weighted averages, ratios between indicators and sums of indicators are examples of scoring functions.

Monitor Configuration Manager. This component is responsible for handling the configuration and administration functionality for creating new, as well as configuring existing monitoring processes. This part is responsible for directly interacting with the users (i.e., Research Administrators as defined in D5.1) and encodes the users' requirements into process definitions and their metadata. Capturing and storing user preferences is handled by the *User Preferences* module. The *Monitoring Process Controller* module is responsible for single handling the execution of monitoring processes as per the user's configurations, and interacts directly with the *Data Harvester* and *Indicators Processor* components at the left of the figure. Coordination and scheduling of multiple monitoring processes into workflows is handled by the *Workflow Management* module.

At the top of the figure lies the presentation/dissemination layer of the OS Monitor, seen in blue. This is responsible for providing the appropriate interfaces (APIs, frontend components etc.) that handle dissemination of the monitoring output. It consists of the following components:

Visualization Service. This is responsible for converting the data to formats that are required as input from frontend visualizations (e.g., charts, graphs, dashboards, etc.) by use of appropriate open APIs for accessing these data. Depending on the visualization framework used in the implementation, this service could serve raw unprocessed results or map the results to structured formats such as JSON, CSV or XML, to enable machine readability.

Organization Scoring Service. This component is responsible for providing the functionality through which stakeholders such as research organizations, issuing bodies and governments can be scored and assigned badges and accomplishments.

Subscription and Alerting Service. This component is responsible for handling the subscription and alerting mechanisms that enable external users to be notified when their declared resources of interest have been scored, or there have been changes in their scores and analyses.

Personalization Service. This component is responsible for enabling the users to set personalized and customized preferences on how they expect and consume the monitoring output. For example, a user is interested in specific targets and indicators or even in other aspects of the monitoring output, such as the performance of a specific country, a research organization, and so on. The OSM will enable the personalization of the content to different stakeholders' needs and views.

Operationally, the above components of the architecture must be combined in individual or multiple monitoring processes, giving rise to a uniform *workflow of execution*. This workflow consists of the predefined tasks and sub-processes, laid out sequentially. OSM should enable the definition and configuration of multiple data sources and workflows for data ingestion and analysis. It collects data from external sources which are further processed and analysed for the computation of the indicators and the assessment of the monitoring targets. An abstract view of a monitoring process flow performed by the OS Monitor can be seen in Figure 7. The main steps that the flow consists of are as follows:

1. **Input.** As defined in *D5.1: The European Open Science Cloud Architecture: Anatomy and Physiology*, the role of monitoring research outcomes is fulfilled as part of the activities of the *Research Administrator*, who acts to combine and aggregate research results to derive metrics and indicators that help assess and evaluate the past and shape the future. As such, Research Admins are the primary users that initiate and perform monitoring processes by defining the research resources,

configuring the indicators to be computed, selecting the data sources and executing or scheduling monitoring tasks. Thus, they are considered the entry points of the monitoring process flow.

2. **Configuration.** During this step, the process gathers all required metadata and descriptive characteristics that instantiate the monitoring, such as the monitoring targets and associated sets of indicators to be computed, the relevant policies, functional characteristics such as scheduling and execution times and so on. The configuration is performed by the Research Administrator role, which can be seen at the left of the figure.
3. **Harvesting.** After configuring the process, the monitor starts collecting data from the defined external data sources. The collected data will be digested in a harvesting process, which includes data validation (e.g. to exclude or cleanse erroneous and corrupted data), and harmonization, which is the process of homogenizing and providing common representation structure and semantics for the input that originates from remote and diverse data sources.
4. **Analysis and scoring.** After collecting all the required data, the next step is that of the actual analysis. During this step, the defined indicators will be computed. Furthermore, the computed indicators will be combined and aggregated according to the Research Administrator's configuration, for producing scores, schemes (e.g. assignment of a badges to a repository) and insights.
5. **Dissemination.** The output of the analysis step is then ready to be accessed for demonstration and presentation purposes. To facilitate these, the monitor will provide appropriate open APIs and frontend components that serve the data in the required formats. These will be served to the monitor stakeholders, i.e., interested parties that wish to consume the results, which can include but are not limited to Research Administrators (e.g., the general public might be interested in exploring visualizations on the outcomes of OS). To facilitate different needs, the results could be made available in different formats (e.g., database exports, text files, visualization formats and so on).

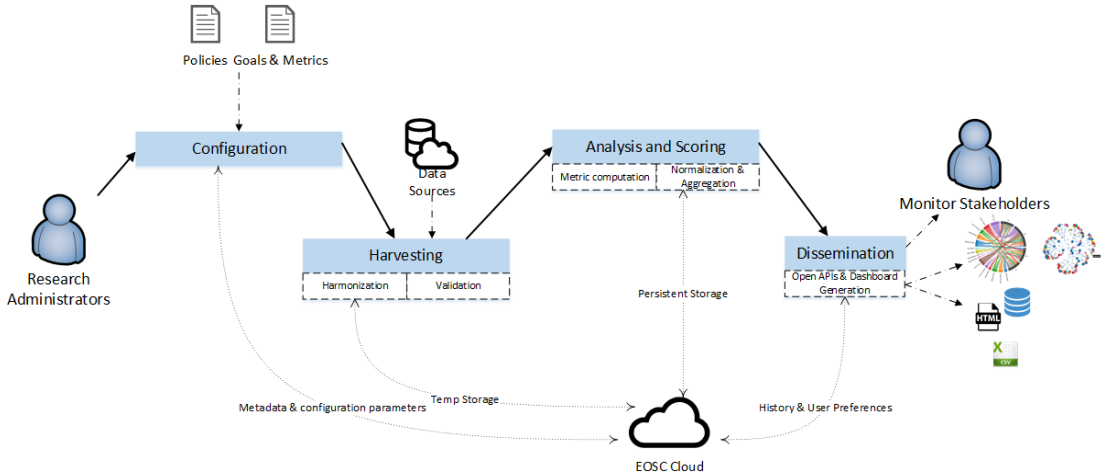

Figure 8: Monitoring Process Flow

Dependencies - Interoperability with the EOSC Cloud. As part of the broader EOSC Cloud architecture, as discussed in detail in D5.1, the presented architecture of the OS Monitor falls in place as part of the EOSC Services for Research Administrators, which also includes the Open Science Policy Registry and the Policy Toolkit. These three are exposed through a common layer of web APIs. The architecture presented herein can easily be integrated with the model of D5.1, by implementing the core functionality layer (shown in red in the figure) together with the Open Science Policy Registry and the Policy Toolkit, and implementing the

dissemination/presentation layer as part of the web APIs that are prescribed in the functionality of the EOSC Services for Research Admins of D5.1.

4.3. The OS Monitor Model

The first step for designing the OS Monitor is to define a high-level representation model for the main elements and semantics of the monitoring landscape. The model itself must provide a high-level description of the core components and their interactions. As was mentioned in Section 1, the purpose of the OS Monitor is to provide both abstractions and specifications for monitoring certain elements (i.e., OS resources), pertaining to specific measurable outputs (monitoring targets), for a given set of interested parties. Furthermore, the model must be flexible enough to allow defining context-specific monitoring flows, instantiations of elements, stakeholders and so on. During the detailed design and implementation phase of the OS Monitor, the OS model will also consider relevant GDPR practices to apply relevant data protection mandates to sensitive and personal information where applicable (e.g., individuals acting as OS Monitor Research Administrators etc.).

To this end, a high-level representation scheme is designed, named the *OS Monitor Model*, which is shown in Figure 8 and will be described in what follows.

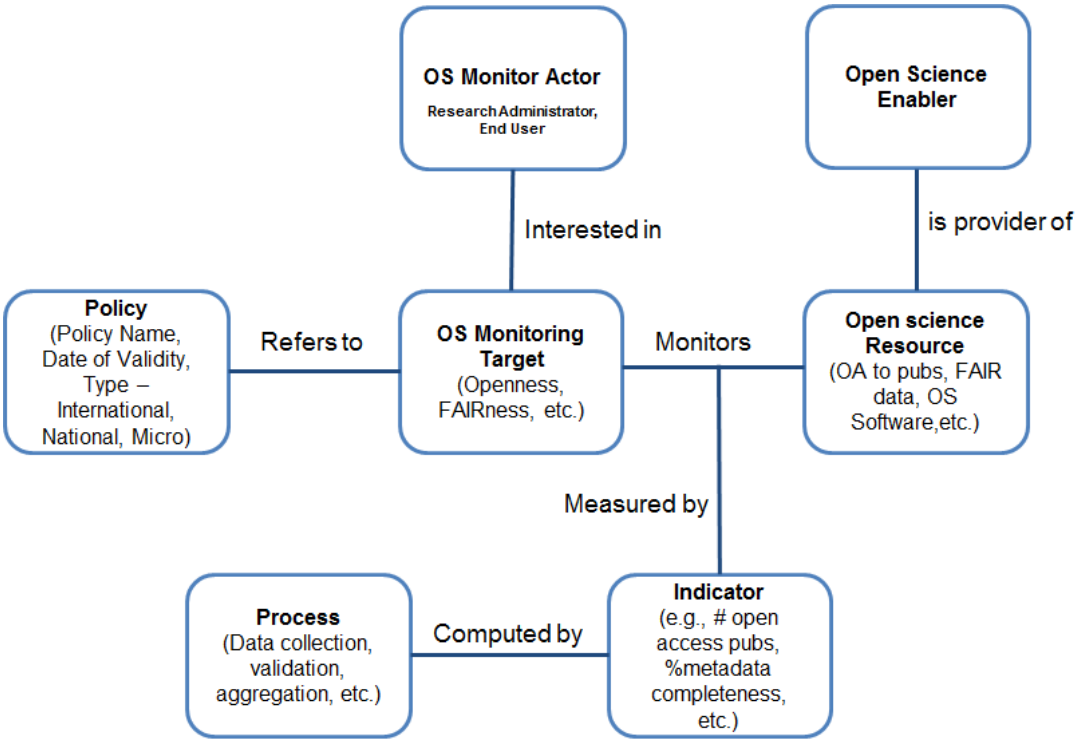

Figure 9: The OS Monitoring conceptual model

In short, the model consists of the following core elements:

OS Monitoring Target. At the core of the OS Monitor Model lies the concept of the *OS Monitoring Target*. An Open Science Monitoring Target is, in essence, the manifestation of an *expected* qualitative or quantitative outcome as defined and declared by relevant Open Science policies and regulations. These targets are defined as "High Level" because they provide a broad categorisation of the motivating targets

behind Open Science. The actual quantification of a monitoring target is defined by relevant *indicators*. However, depending on the stakeholders and the broader context, the goals can differ, thus leading to different subsets of these targets for different instantiations of the OS Monitor.

As the OS Monitor Model is designed to be extensible, more specialized concepts, such as second- and third-level categorisations of the broad targets can be deployed, to create a descriptive and more detailed taxonomy of monitoring targets, with parent-child (i.e., specialization-generalization) relationships between them. In this sense, monitoring targets exhibit a tree structure, where each target potentially consists of a set of sub-targets. This allows analytical operations such as aggregations and specializations depending on the level of the selected target in the tree.

OS Monitor Actor. Open Science Actors are the research administrators and end users, collectively. They are interested in some aspect of the monitoring process either actively by creating and initiating the process, or passively by viewing, using and re-using the results. Specifically, research administrators, as defined in D5.1, are the users that are responsible for defining monitoring targets and creating monitoring processes for specific OS resources. In the context of the OS Monitor, the research administrator is an actor that drives the monitoring process by defining a set of monitoring targets and their respective and relevant indicators to quantify an OS resource's involvement in the OS lifecycle. End-users on the other hand are interested in the output of the monitoring processes. Specifically, they participate passively in the process, by following their interests in specific monitoring targets or resources, and consuming/exploiting the results.

OS Enabler. An OS Enabler is an actor (e.g., government, research organization, publisher etc.) that is the initiator or provider of a specific OS resource. OS Enablers, within the context of the EOSC Monitor are similar to those involved in activities throughout the broader scientific research lifecycle, including:

- Research Infrastructures, e-Infrastructures, VREs or other pertinent H2020 projects, Service Providers (Academic/Commercial), Data repositories:
- Research Funding Bodies, national, regional and local government agencies:
- Learned societies, research communities, scientific and professional associations:
- Enterprises (Industry as providers or consumers, SMEs, Startups, etc)

A comprehensive list of OS enablers, defined as OS stakeholders, is discussed in D8.1.

Policy. Within the context of the OS Monitor Model, a policy as defined in Section 2, represents a formal mandate of the aforementioned form that contextualizes, specifies, defines or in any way supports the existence of a specific monitoring target. Thus, the OSM supports the connection of Monitor Targets with Policies; thus enabling the direct assessment and correlation of policy mandates with OS monitoring results. In accordance with the results of D3.4 and D3.5, the OSM model aims at accommodating machine readable policies.

OS Resource. The term Open Science Resource is a container term, in the sense that it captures all Open Science elements, those being drivers and enablers of the OS movement. An OS Resource is any element of the scientific/research lifecycle that can be measured and evaluated within the context of open science. Consequently, an OS Resource in the OS Monitor Model can be instantiated by multi-level taxonomies, or

flat lists of artefacts, depending on the context. For the purpose of the OS Monitor, primary OS Resources consist in Research Artefacts such as open access publications, open research data, open source software, Open Educational Resources, Research Collaboration and Citizen Science, among others.

Indicator. The success of an OS resource towards a specific monitoring target is quantified with the use of *indicators*. Within the context of the OS Monitor, indicators are metrics that are defined as quantitative functions that can measure the existence, degree or absence of a monitoring target in various ways. Some of these ways are as follows:

- percentage of coverage. Examples include % of open access articles for a specific journal and % of metadata completeness for a publicly accessible dataset.
- binary outcome (e.g., true/false, yes/no, etc.). Examples include existence of any open license for a specific resource, and existence of properly formed URIs for a dataset's records.
- numbers and quantities. Examples include # of available open access articles, # of downloadable datasets and so on.
- selection from fixed set/scale of values.

Furthermore, indicators are accompanied by other characteristics, such as attribution of a temporal dimension (e.g., static or recurring, measured timestamp or timeframe). The majority of the EOSC OSMF metrics are derivatives of existing efforts from stakeholders assessing their research outputs' "openness"[135], FAIRness. Indicators can be combined in the form of scoring functions (e.g. weighted sums of indicators, averages of indicators etc). These scoring functions provide higher level quantification capabilities over a monitoring target (i.e. scoring), This way, a monitoring target is not only measured as the result of one indicator, but as a combination of the results of >1 indicators.

Process. A process is the operational function required to quantify a defined monitoring target over a particular OS resource, given a specific indicator. It consists of all tasks that are relevant to the metric computation procedure, including collection of the needed data, parsing and validation of the data, application of the specific function for computing the indicator. Processes are defined by research administrators. Their execution is controlled by the research administrators and they are schedulable, exhibiting temporal characteristics such as periodicity and frequency.

4.4. Services of the EOSC Monitor Framework

The OS Monitor depends on the described core functionality to execute monitoring tasks. However, a crucial component is the user's ability to interact with the monitor for both initiating monitoring processes and consuming the resulting output, depending on their role (i.e. research admins, stakeholders etc). The required functionality, in the form of services, for the research administrator must cover administrative, as well as configuration tasks for setting up and customizing the monitoring process. In what follows, a list of envisioned services that enable interaction of the research administrator with the OS Monitor is presented:

- **Research Administrator Dashboard.** The OS Monitor must provide an easy way for research admins to see summarizations and overviews of their monitoring contributions. This will be

[135] here to denote all movements regarding "Open" and transparency and their respective aspects.

presented in the form of a dashboard that contains a "bird's eye" overview, containing a description of ongoing monitoring processes, a description of upcoming monitoring processes, as well as the history of previously executed processes. In summary, the user dashboard will consist of the following sub-services:

- Description of current (ongoing) processes
- Description of upcoming processes
- History of executed monitoring processes
- Alerts and Notifications

- **Monitor Administration and Configuration Panel.** The role of research administrators within the OS Monitor is to define, initiate and administer monitoring processes for research resources of interest. To this end, the OSM must provide appropriate interfaces and services for the administration and configuration of monitoring processes and tasks. This will be achieved with the use of an administration and configuration panel, through which the research administrator will be able to:
  - Create new monitoring processes by:
    - Ascribing metadata to monitoring processes
    - Defining input parameters (data sources, policies etc)
    - Defining monitoring targets and assigned indicators to be computed
    - Defining temporal/scheduling characteristics
    - Defining alerts and notifications to be triggered at specified points in the monitoring process (e.g. when data collection is complete, when metric computation is complete, etc)
  - Edit existing monitoring processes by changing any or all the above values
  - Delete/remove existing monitoring processes
  - Share monitoring process definitions with other research administrators

The front-end services for the end users, i.e. the consumers of the monitoring output, are described in the following:

- **User Personalization Service.** A set of end-user services for providing personalization to the monitored resources the end-user is interested in. This can be in the form of a personalized dashboard containing overview and information on trending research resources, tracked scores of resources and other information depending on the preferences of the end user.
- **Monitoring report.** The OSM will provide end-user services for rendering the results of a monitoring process into a human-readable report, in the form of a *monitoring profile*. This will contain more detailed information on a specific monitoring process or monitored resource and will provide different levels of detail on the results (e.g., overview, specific indicator scores etc.). The report will also consist of visualizations (charts, graphs, etc.) which will provide the users the ability to visually interact and explore the monitoring results. Consequently, monitoring reports should:
  - Enable presentation of the monitoring insights in rich visualizations and enable visual exploration (overviews, aggregates, drill down, etc)
  - Allow for comparisons of results and indicators across different monitoring targets, as well as across different resources and stakeholders.
  - Enable specification of the form of output (e.g. machine-readable data, spreadsheet, visual charts, etc)

  - Enable the ability for front-end users to subscribe, unsubscribe and be alerted when measures change / reach a certain target
- **Keyword Search, results browsing and faceted search service.** This is a service that allows end users to search and explore monitored resources and see the results. Resources that are retrieved from this functionality can be added to the end user's interests so they appear in their monitor dashboard in the future. Furthermore, the search results can be shared through social media or other ways on the web.
- **Aggregated scores and Badge systems service.** The OSM will provide incentive for research stakeholders to improve their assessments with the use of a tailored badge system, based on aggregated scores from different indicators. This will provide a categorical way of rewarding monitored research stakeholders, depending on the results of their evaluation and their overall performance. Different or multi-level badge systems can be employed according to the domain or the type of the monitored organization. End-users (i.e., consumers of monitoring results) will also benefit from the badge system, as they will quickly filter, explore and comprehend different types and levels of open science contributors.

In addition to the above specifications, the OS Monitor will support user authentication via the AAI services of the overall EOSC system. The above services are summarized in the following table.

Functionality	Service	OS Monitor Actor
Create new monitoring processes	Monitor Administration and Configuration Panel	Research Admin
Edit existing monitoring processes by changing any or all of the above values	Monitor Administration and Configuration Panel	Research Admin
Delete/remove existing monitoring processes	Monitor Administration and Configuration Panel	Research Admin
Share monitoring process definitions with other research administrators	Monitor Administration and Configuration Panel	Research Admin
Description of current (ongoing) processes	Research Admin Dashboard	Research Admin
Description of upcoming processes	Research Admin Dashboard	Research Admin
History of executed monitoring processes	Research Admin Dashboard	Research Admin
Alerts and Notifications	Research Admin Dashboard	Research Admin
Personalized Results	User Personalization Service	End-user
Personalized Profile	User Personalization Service	End-user
Monitor report viewing	Monitor Report Service	End-user

Monitor report export	Monitor Report Service	End-user
Monitor report comparison	Monitor Report Service	End-user
Subscribing to monitor reports	Monitor Report Service	End-user
Keyword search	Search Service	End-user
Faceted search	Search Service	End-user
Filtered (advanced) search	Search Service	End-user

Table 1: Functional Specifications of OSM end user services

5. CONCLUSIONS AND FURTHER ACTIONS

The EOSCPilot project aims to address the main challenges and problems related to the realization of the EOSC, making an important step towards building a uniform open innovation environment for fostering Open Research in EU, including through the provision of clear incentives and rewards for the sharing of data and resources. The realization of such an environment is a continuous process, whose basic requirements include a *principled approach for monitoring and measuring* the uptake and the impact of Open Science trends and practices, across a clear set of measurable targets, such as the openness, findability and accessibility to open science elements. That need of a monitoring framework to facilitate open science expectations and requirements is clearly mirrored in national pictures which lack relevant services albeit that some of them have provisions and deployment plans in their national strategies.

In this respect, the contribution of this deliverable is multifold. It first provided a thorough review and categorisation of the existing efforts and tools proposed so far for the monitoring of Open science resources in national, regional, European or international level. The review and categorisation of the approaches will enable organizations and initiatives, interested in implementing an OS monitoring mechanism, to identify similar efforts, build upon and extend these tools to adjust them to their specific monitoring goals. It then proposed a methodology with the main concepts and the steps to be followed for the implementation of the OS monitoring framework by the EOSC and its adaptation by other interested organizations. Finally, it described the specifications for the implementation of such a framework, i.e., the key modelling concepts, architectural considerations, standards and processes that an OS monitor framework must support, as well as an overview of the added value services that the framework must offer to end users. The two possible paths foreseen for the deployment and operation of the OS Monitor in EOSC concern that the OS Monitor framework will be one of the services in the overall EOSC system, i.e., a monitor-as-a service tool as part of the EOSC software stack, collecting indicators and monitoring data by the organizations participating in EOSC, and offering to the EOSC stakeholders the functionality for monitoring, visualizing and gaining insights about OS trends and impact; or the implementation, customisation and deployment of the OS monitor framework could be performed by each individual organization participating in the EOSC, which in turn publishes the monitoring results to the EOSC portal.

Again, it should be noted that the whole landscape of OS Monitoring efforts and initiatives is quite dynamic, with new indicators and processes being introduced under different domains and contexts in the EOSC ecosystem. As such, this deliverable provides a first review of the current landscape, covering the most mature efforts in this area, and a first set of specifications based on the current development phase of the EOSC. It is expected that the future development of EOSC will provide more aspects and resources of OS that should be monitored via this framework, such as infrastructures and OS services. OSM is indeed designed as an extensible framework, which can easily accommodate new measurable targets, indicators, badge and certification systems and tools for the future advancements of OS.

In addition, it is foreseen that the activities performed in T3.2 regarding the Policy Registry (D3.4) together with D3.3 regarding Policy Recommendations in Open Science, data protection, procurement and ethics will provide better insights of the policy-compliance indicators that should be measured in the OS Monitor. The goal is to align and enrich this deliverable based on the output of these activities. Moreover, the revised version of this deliverable will be consulting with other WPs, one of them being the WP7 Skills and Training, to address new indicators and find technical solutions for content aggregation and data collection.

Finally, another important aspect to be covered in the future activities concerns the assessment and evaluation of the proposed OS monitoring approach with the relevant stakeholders, namely the EOSC

demonstrators of the EOSCPilot. This task will provide more concrete monitoring targets, indicators and specifications tailored to domain-specific requirements and end-users' needs and will guide the implementation of the framework within the EOSC system.

ANNEX A. MAPPING OF MONITORING TARGETS TO INDICATORS

Openness of publications

Target	Subtarget	OS Actor[136]	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Openness	OA routes costs	funders/governments	HowOpenIsIt: Guide to Research Funders policies ● % of costs coverage by funder (all, specific dollar value) ● % of fees - types (APCs, data storage) ● % of costs in the grant application vs not ● time of costs coverage by funder, e.g. ○ no time limit, ○ one year after grant expiration ○ on a case-by case basis
		research performing organisations/ researchers	Knowledge Exchange: Monitor OA publications and cost data *For Green OA*: ● % of costs for the development of CRIS and equally for IRs ● % of costs for the maintenance and equally for IRs ● % of labour costs of academic and supporting staff ● % of costs for maintenance of subscriptions to journals *For Gold OA*: ● % of APCs expenditures ● % of VAT in APCs, % of discounts in APCs, average % of APC per publisher ● % of costs for administrative purposes paid by authors vs % of costs for administrative purposes paid by APC funds ● % of costs for membership deals and % of fees for membership deals *For Hybrid OA:* ● % of costs paid outside of the offsetting deals (authors paid) ● % of VAT in APCs, % of discounts in APCs, average % of APC per publisher in comparison to the same value for Gold OA ● % of costs per article ● % of open monitoring data Metrics for Openness ● Individual Purchase Index: % of costs paid by a reader to access authors' work ● Openness Cost Index: sum of any access fees/page fees,

[136] here for end-users

			etc including the effort taken ● Practical Openness Index: % of authors' open vs paywalled conference and journal papers; % of authors' peer-reviewed, postprints conference and journal papers ● OI-Broad: same as Practical Openness Index + book chapters
	Licenses	journals/publishers	How Open Is It: A guide for evaluating the openness of journals ● # of articles available to ○ read immediately, ○ after 6 months, ○ more than 6 months ● % of gratis vs libre OA journals ● # of attribution licences ○ generous reuse & remixing rights, e.g., CC-BY / Reuse, ○ remixing with restrictions, e.g., CC BY-NC & CC BY-SA ○ reuse with restrictions, ○ no remixing, e.g., CC-BY-ND ○ % of no exceptions to copyright (no reuse, all rights reserved copyright) ● % of ownership by author vs % of ownership by publisher ● % of restrictions, e.g., ○ no restrictions, ○ restrictions on reuse of published version by publisher ● # of journals and % of automated procedures ● # of journals and % of automated procedures (immediately, 6 months, 12 months) ● # of journals and % of repositories vs datacenters, ● # of journals and % national vs international infrastructure, ● # of journals and % of disciplinary-specific or generic
		funders/governments	HowOpenIsIt: Guide to Research Funders policies ● % of research outputs with generous reuse & remixing rights, e.g., ○ CC BY/CC0 licenses or ○ moderate reuse & remixing rights (e.g., CC BY-SA license) ○ or commercial modification rights (e.g., CC BY-NC license) ○ or modification rights (e.g., CC BY-ND license) or ○ # of DMPs with re-use permissions

		research performing organisations/ researchers	Knowledge Exchange: Monitor OA publications and cost data ● % of monitoring data re-use ● # of offsetting contracts with license statement Metrics for Openness ● % of authors' peer-reviewed, postprints publications without copyright restrictions; ● # of authors' open vs with copyright restrictions publications (meaning all published items); ● # of authors' gratis or libre conference and journal papers OSI 2016-25 "openness score" Creative Commons license, e.g., ● rights to freely read or mine its context ● under embargoes ● pay-walled How Open Is It: A guide for evaluating the openness of journals ● # of journals and % of automated procedures ● # of journals and % of automated procedures (immediately, 6 months, 12 months) ● # of journals and % of repositories vs datacenters, ● # of journals and % national vs international infrastructure, ● # of journals and % of disciplinary-specific or generic

FAIRness of publications

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Findability	Metadata PIDs Directories	journals/publishers	OSI 2016-25 "openness score" Availability measures: ● metadata quality ● % of metadata completeness ● discoverability ● crawling ● machine readability ● links to other resources, e.g., # of papers with links or annotations ● public access to usage data, e.g., # of usage data available to the public

		research performing organisations/ researchers	OSI 2016-25 “openness score” See above in journals/publishers section
Accessibility	Protocols	journals/publishers	Knowledge Exchange: Monitor OA publications and cost data ● # of metadata fields for APC/publication in the Open Archives Initiative (OAI) Protocol for Metadata ● # of OAI-PMH compliant repositories ● # of fields with APC information
		funders/governments	HowOpenIsIt: Guide to Research Funders policies ● # of articles published in fully open access journals or ● # of open repositories or ● # of "hybrid" journals in combination with: ● % of free, immediate readership rights or ● embargo (6-12 months) or ● # of plans detailing how these can be accessed
Interoperability	Machine Readability Standardisation	journals/publishers	OSI 2016-25 “openness score” # of standardised formats (PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml) Open Digital Science ● % of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations) ● % of published works using researcher persistent IDs (e.g. ORCID)
		funders/governments	Knowledge Exchange: Monitor OA publications and cost data ● # of standardized data for APCs ● average % of standardized costs for administrative costs per publisher ● # of CRIS and Universities with DOIs for their accounting systems, % of linkage between these two ● # of publishers providing DOIs of articles in invoices, % of DOIs of articles found in invoices ● # of CrossMark containers on versions and costs

		research performing organisations/ researchers	Knowledge Exchange: Monitor OA publications and cost data ● % of APIs or protocols (community standard or not) ● % of APIs or protocols (permissions to be crawled or not) *Monitoring Green OA*: ● % of metadata completeness as identified through automatic procedures - routing via Crossref and OpenAIRE ● % of metadata standards used for integrations ● % of metadata that include ORCIDs, ● % of metadata that include author's or a correspondance's email address ● # of two-way metadata linkage between CRIS and repositories found in Crossref and OpenAIRE ● # of repositories following Green, Gold and Hybrid OA *Monitoring Gold OA:* ● # of CRIS journal articles ● # of same journal articles in DOAJ *Monitoring Hybrid OA*: ● % of NISO offsetting deals (worldwide) ● #of standardised data formats for author affiliation ● % of standards when depositing articles ● # of DOI, # of PIDs, % of metadata fields found with rights information ● % of commercial sources that are CERIF compliant ● # of OAI-PMH compliant repositories, # of fields with APC information ● % of open data in standardized formats for TDM ● % of publishers using the SWORDprotocol for financial data exchange with repositories ● % of CRIS and repositories following the OpenAIRE guidelines OSI 2016-25 “openness score” Format measures: Per file formats - # of standardised formats examples: PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml Open Digital Science ● % of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)

			• % of published works using researcher IDs
Reusability		journals/publishers	How Open Is It: A guide for evaluating the openness of journals • % of attribution licences (generous reuse & remixing rights eg CC-BY / Reuse, remixing with restrictions e.g., CC BY-NC & CC BY-SA / reuse with restrictions, no remixing eg CC-BY-ND) or • % of no exceptions to copyright (no reuse, all rights reserved copyright) • # of content types of OS Resources regarding Copyrights and exceptions • % of ownership by author and • % of restrictions (with no restrictions, restrictions on reuse of published version) or by publisher and % of restrictions (with some allowances on author and/or user reuse of published version, no author reuse, only fair use by author)
		funders/governments	HowOpenIsIt: Guide to Research Funders policies • % of research outputs with generous reuse & remixing rights (e.g., CC BY/CC0 licenses) or • moderate reuse & remixing rights (e.g., CC BY-SA license) or • commercial modification rights (e.g., CC BY-NC license) or • modification rights (e.g., CC BY-ND license) or • # of DMPs with re-use permissions Knowledge Exchange: Monitor OA publications and cost data Add to offsetting contracts (eg in terms and conditions) that publishers should include in Crossref a licence statement for each publication: • # of offsetting contracts with license statement % of open monitoring data, • % of monitoring data re-use Data should be open in standardized formats allowing for data mining: % of data that can be processed through TDM

		research performing organisations/ researchers	Metrics for Openness ● % of authors'peer-reviewed, postprints publications without copyright restrictions" ● # of authors' open vs with copyright restrictions publications (meaning all published items) OSI 2016-25 "openness score" License measures: -Creative Commons -free to read -free to mine -embargoed and embargo length -pay-walled Open Digital Science ● % of publications with free licensing (public domain, attribution, all kinds of sharing)

Openness of research data and data repositories

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Openness	Licenses	journals/publishers	Metrics to Assess Value of Biomedical Digital Repositories ● L1: ○ Documented: Explicit data use terms (ideally formal licenses) should be defined by the resource providers and easy to find (yes/no) ○ Clear: a. At a minimum, licenses/data use agreements must be clear and easy to understand. A variety of specific examples of data use/reuse conditions should be included. (yes/no) b. Licenses should not require negotiation and licenses themselves should be legally redistributable without engaging legal counsel (yes/no) ○ Minimally restrictive: The licenses and/or data use agreements should explicitly permit downstream data reuse, derivation, and re-dissemination (yes/no) ○ Standard licenses: For data, ideally CC0., Standard software license: For software, ideally Apache version 2 (valid options include GPLv2, GPLv3, AGPLv3, etc) ○ Contactable: person -contact for license information (yes/no) ● L2: Transparent about flowthrough implications # of sources identified with flowthrough implications # of links to the original licenses/data use terms of all redistributed content % of authorization for redistribution
		funders/governments	HowOpenIsIt: Guide to Research Funders policies ● % of costs coverage by funder (all, specific dollar value) and ● # of fees - types (APCs, data storage) and ● % of costs already included in the grant application vs not, and ● time of costs coverage by funder (not time limit, one year after grant expiration, on a case-by case basis) Open Data Portal # of countries that recommends open data license % of data available under open license % of data available free of charge

		research performing organisations/ researchers	Metrics to Assess Value of Biomedical Digital Repositories ● L1: ○ Documented: Explicit data use terms (ideally formal licenses) should be defined by the resource providers and easy to find (yes/no) ○ Clear: a. At a minimum, licenses/data use agreements must be clear and easy to understand. A variety of specific examples of data use/reuse conditions should be included. (yes/no) b. Licenses should not require negotiation and licenses themselves should be legally redistributable without engaging legal counsel (yes/no) ○ Minimally restrictive: The licenses and/or data use agreements should explicitly permit downstream data reuse, derivation, and re-dissemination (yes/no) ○ Standard licenses: For data, ideally CC0., Standard software license: For software, ideally Apache version 2 (valid options include GPLv2, GPLv3, AGPLv3, etc) ○ Contactable: person -contact for license information (yes/no) ● L2: Transparent about flowthrough implications # of sources identified with flowthrough implications # of links to the original licenses/data use terms of all redistributed content % of authorization for redistribution (Re)usable Data Project ● % of public, discoverable, and standard licenses ● % of license requires no further negotiation and its scope is both unambiguous and covers all of the data ● % of data accessibility (data referring to/ covered by the license) ● % of data/datasets with license that has restrictions or no restrictions for re-use ● % of data/datasets with "re-users" license restrictions

FAIRness of research data and data repositories

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Findability	metadata PIDs directories	journals/publishers	FAIR metrics ● Uniqueness: Metric Identifier - FM-F1A : # of identifiers in data/datasets/repositories metadata or # of URLs to a registered identifier scheme ● Persistence: Metric Identifier - FM-F1B : # of identifiers in data/datasets/repositories metadata or # of URLs that resolves to a document containing policy that describes what the provider will do in the event an identifier scheme becomes deprecated ● Machine-Readability: Metric Identifier - FM-F2 : # of identifiers in data/datasets/repositories metadata or # of URLs to a document that contains machine-readable metadata for the digital resource. + file format must be specified ● Resource Identifier in Metadata: Metric Identifier - FM-F3 : # of identifiers in data/datasets/repositories metadata or # of URLs of the metadata and the GUID of the digital resource it describes ● Indexed in a Searchable Resource: Metric Identifier - FM-F4: # of identifiers in data/datasets/repositories metadata or # of persistent identifiers of the resource + one or more URLs that give search results of different search engines Open Data Monitor Metadata completeness: average % of missing metadata (defined set contains: license, author, organisation, date released and date updated) OSI 2016-25 “openness score” Availability measures: -metadata quality - % of metadata completeness -discoverability -crawling -machine readability -links to other resources # of papers with links or annotations -public access to usage data # of usage data available to the public

		research performing organisations/ researchers	Metrics to Assess Value of Biomedical Digital Repositories ● F1: Discoverable through various external mechanisms ○ Registered: % of repositories in re3data or OpenAIRE etc ○ Discoverable via search engines/applications: By name and by features (metadata) - # of successful searches ○ Linked from external resources: % of other OS resources linking to the repositories and vice versa ● F2: Contents/components are well documented and searchable ○ Metadata documented: # of timestamps, versions, counts etc. showing that further effort with the dataset is warranted ○ Indexed: % of indexed contents + % of optimised contents that support common queries ○ Searchable using various mechanisms: % of Search boxes and APIs ○ Contactable: % of contact information found FAIR metrics ● Uniqueness: Metric Identifier - FM-F1A : # of identifiers in data/datasets/repositories metadata or # of URLs to a registered identifier scheme ● Persistence: Metric Identifier - FM-F1B : # of identifiers in data/datasets/repositories metadata or # of URLs that resolves to a document containing policy that describes what the provider will do in the event an identifier scheme becomes deprecated ● Machine-Readability: Metric Identifier - FM-F2 : # of identifiers in data/datasets/repositories metadata or # of URLs to a document that contains machine-readable metadata for the digital resource. + file format must be specified ● Resource Identifier in Metadata: Metric Identifier - FM-F3 : # of identifiers in data/datasets/repositories metadata or # of URLs of the metadata and the GUID of the digital resource it describes ● Indexed in a Searchable Resource: Metric Identifier - FM-F4: # of identifiers in data/datasets/repositories metadata or # of persistent identifiers of the resource + one or more URLs that give search results of different search engines

			Open Data Monitor Metadata completeness: average % of missing metadata (defined set contains: license, author, organisation, date released and date updated) OSI 2016-25 "openness score" Availability measures: -metadata quality - % of metadata completeness -discoverability -crawling -machine readability -links to other resources # of papers with links or annotations -public access to usage data # of usage data available to the public
Accessibility	Protocols	journals/publishers	Metrics to Assess Value of Biomedical Digital Repositories ● A1: Diverse data access mechanisms ○ Dumps: # of (whole) dataset dumps available or % of whole or part ○ Query: % of Query interfaces or exports ○ Downloads: % of records and parts of the

			database that are downloadable (e.g. as JSON/XML/tab delimited, etc.) ○ API: % of the data with APIs ● A2: Well structured and provisioned APIs ○ RESTful: # of RESTful APIs ○ JSON: # of JSON or TSV returns ○ Retrieval: # of single records with PIDs, and for batch retrieval of a list of data entities: # of PIDs listed ○ Paging: Provide a query interface to return matching data entities with paging support (yes/no) ○ Versioned: # of versioned URL patterns, and # of documented policies for change management ○ Uptime: # of API uptime reports ○ Access: % of access requests granted (e.g. new accounts or API keys), and % of write access granted (to make contributions, corrections, suggestions to records) ● A3: Understandable data and scope ○ Audience: Target audience and use cases are well defined and obvious from the homepage (yes/no) ○ Content: The content types included are obvious from the homepage (yes/no) ○ Browsable: % of high-level categories / visualizations to browse data ○ Documented: % of well documented data model, schema, data dictionaries, etc. ○ Tutorials: # of tutorials available to new users, and % of literature cited that previously used the repository FAIR metrics ● Access Protocol: Metric Identifier - FM-A1.1 : # of identifiers in data/datasets/repositories metadata or # of URLs to the description of the protocol + # of those the protocols being open source + # of those protocols being (royalty) free ● Access Authorization: Metric Identifier - FM-A1.2 : # of identifiers in data/datasets/repositories metadata or % of times where authorization was needed + description of the process to obtain access to restricted content ● Metadata Longevity: Metric Identifier - FM-A2 : # of identifiers in data/datasets/repositories metadata or # of URLs to a formal metadata longevity plan (or policy)

			Open Digital Science ● accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects ● access to simulation results

		research performing organisations/ researchers	Metrics to Assess Value of Biomedical Digital Repositories ● A1: Diverse data access mechanisms ○ Dumps: # of (whole) dataset dumps available or % of whole or part ○ Query: % of Query interfaces or exports ○ Downloads: % of records and parts of the database that are downloadable (e.g. as JSON/XML/tab delimited, etc.) ○ API: % of the data with APIs ● A2: Well structured and provisioned APIs ○ RESTful: # of RESTful APIs ○ JSON: # of JSON or TSV returns ○ Retrieval: # of single records with PIDs, and for batch retrieval of a list of data entities: # of PIDs listed ○ Paging: Provide a query interface to return matching data entities with paging support (yes/no) ○ Versioned: # of versioned URL patterns, and # of documented policies for change management ○ Uptime: # of API uptime reports ○ Access: % of access requests granted (e.g. new accounts or API keys), and % of write access granted (to make contributions, corrections, suggestions to records) ● A3: Understandable data and scope ○ Audience: Target audience and use cases are well defined and obvious from the homepage (yes/no) ○ Content: The content types included are obvious from the homepage (yes/no) ○ Browsable: % of high-level categories / visualizations to browse data ○ Documented: % of well documented data model, schema, data dictionaries, etc. ○ Tutorials: # of tutorials available to new users, and % of literature cited that previously used the repository FAIR metrics ● Access Protocol: Metric Identifier - FM-A1.1 : # of identifiers in data/datasets/repositories metadata or # of URLs to the description of the protocol + # of those the protocols being open source + # of those protocols being (royalty) free ● Access Authorization: Metric Identifier - FM-A1.2 : # of identifiers in data/datasets/repositories metadata or % of times

			where authorization was needed + description of the process to obtain access to restricted content ● Metadata Longevity: Metric Identifier - FM-A2 : # of identifiers in data/datasets/repositories metadata or # of URLs to a formal metadata longevity plan (or policy) Open Digital Science ● accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects ● access to simulation results
Interoperability	Machine Readability Standardisation	journals/publishers	Metrics to Assess Value of Biomedical Digital Repositories ● I1: Identifiers ○ Credit any derived content using its original identifier: # of credited identifiers ○ Help local identifiers travel well: % of

			documented prefix and patterns ○ Design new identifiers for diverse uses by others (yes/no) ○ Avoid embedding meaning, or relying on it for uniqueness ○ Opt for simple, durable web resolution: # of resolutionable PIDs ○ Implement an identifier version-management policy: % of relevant policy in place ○ Make URIs clear and findable: # of findable URIs ○ Do not reassign or delete identifiers: # of reassigned or deleted PIDs ○ Document the identifiers you issue and use: % of documented identifiers ○ Reference and display responsibly: % of referenced or displayed identifier ● I2: Vocabularies, Ontologies, and exchange standards Semantics/data structure: ● Data dictionary is provided (yes/no) ● Defined schema or data model is provided (yes/no) ● Services are well aligned to the model and consistent across various access mechanisms (yes/no) ● Structure, format, architecture, and metadata for the repository is consistent with community norms or shared specifications (for example, use of the W3C Dataset Description) (yes/no) Exchange standards: ○ Data are made accessible using common exchange formats, if applicable (for example, use of the HL7 FHIR standard for exchanging healthcare information electronically) (yes/no) ○ Data elements are well-defined using metadata standards (e.g., ISO/IEC 11179, DDI and SDMX/ISO17369) (yes/no) ○ Value set services and value set definition services using the Common Terminology Services 2 (CTS2) standard (yes/no) ● Ontologies: ○ All ontologies in use are documented in one place and are consistently applied to the data (yes/no) ■ Novel ontologies, if any, are registered in public standards repositories (such as the OBO Foundry Library) and released via standard well documented

			mechanisms (for examples ROBOT or the OBO Starter Kit)(yes/no) ■ Appropriate community standards/vocabularies are used to record metadata; preferably standards that are: a) designated or de facto standards within the relevant domain, and b) free to use, see also Licensure section (yes/no) ■ Version of the ontologies used is indicated (yes/no) ■ Ontologies are attributed according to community best practice (yes/no) I3: Versioning ● Data versioning and/or change history is well documented (yes/no) ● Prior versions of each database release (or each record, if appropriate) are accessible (yes/no) Open Data Monitor Machine readable: % of machine readable datasets over the total count of datasets Open Data Watch Machine Readability: % of data that are downloadable in machine readable formats (XLS, XLSX, CSV, Stata, SAS, SPSS, JSON, CDF, RDF, XML, and TXT) + Non-proprietary formats: in combination with % of non-proprietary formats of this data (XLSX, DOCX, CSV, XML,HTML, and JSON, txt) FAIR metrics ● Knowledge Representation Language: Metric Identifier - FM-I1 : # of identifiers in data/datasets/repositories metadata or # of URLs to the specification of the language (eg BNF) ● Use of FAIR Vocabularies: Metric Identifier - FM-I2 : # of identifiers in data/datasets/repositories metadata or # of UUIDs representing the vocabularies used for (meta)data ● Use of Qualified References: Metric Identifier - FM-I3 : # of identifiers in data/datasets/repositories metadata or # of Linksets (in the formal sense) representing part or all of your

			resource Open Digital Science ● % of machine-readable data / metadata ● quality of metadata (versioning, volume, data format, description of fields, etc.) ● % of published works using researcher IDs OSI 2016-25 “openness score” ● # of standardised formats examples: PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml Open Digital Science ● % of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)

		research performing organisations/ researchers	Metrics to Assess Value of Biomedical Digital Repositories ● I1: Identifiers Credit any derived content using its original identifier: # of credited identifiers Help local identifiers travel well: % of documented prefix and patterns Design new identifiers for diverse uses by others Avoid embedding meaning, or relying on it for uniqueness Opt for simple, durable web resolution: # of resolutionable PIDs Implement an identifier version-management policy: % of relevant policy in place Make URIs clear and findable: # of findable URIs Do not reassign or delete identifiers: # of reassigned or deleted PIDs Document the identifiers you issue and use: % of documented identifiers Reference and display responsibly: % of referenced or displayed identifiers ● I2: Vocabularies, Ontologies, and exchange standards Semantics/data structure: ● Data dictionary is provided (yes/no) ● Defined schema or data model is provided (yes/no) ● Services are well aligned to the model and consistent across various access mechanisms (yes/no) ● Structure, format, architecture, and metadata for the repository is consistent with community norms or shared specifications (for example, use of the W3C Dataset Description) (yes/no) Exchange standards: ○ Data are made accessible using common exchange formats, if applicable (for example, use of the HL7 FHIR standard for exchanging healthcare information electronically) (yes/no) ○ Data elements are well-defined using metadata standards (e.g., ISO/IEC 11179, DDI and SDMX/ISO17369) (yes/no) ○ Value set services and value set definition services using the Common Terminology Services 2 (CTS2) standard (yes/no) ● Ontologies: ○ All ontologies in use are documented in one place and are consistently applied to the data (yes/no)

			■ Novel ontologies, if any, are registered in public standards repositories (such as the OBO Foundry Library) and released via standard well documented mechanisms (for examples ROBOT or the OBO Starter Kit)(yes/no) ■ Appropriate community standards/vocabularies are used to record metadata; preferably standards that are: a) designated or de facto standards within the relevant domain, and b) free to use, see also Licensure section (yes/no) ■ Version of the ontologies used is indicated (yes/no) ■ Ontologies are attributed according to community best practice (yes/no) I3: Versioning ● Data versioning and/or change history is well documented (yes/no) ● Prior versions of each database release (or each record, if appropriate) are accessible (yes/no) Open Data Monitor Machine readable: % of machine readable datasets over the total count of datasets Open Data Watch Machine Readability: % of data that are downloadable in machine readable formats (XLS, XLSX, CSV, Stata, SAS, SPSS, JSON, CDF, RDF, XML, and TXT) + Non-proprietary formats: in combination with % of non-proprietary formats of this data (XLSX, DOCX, CSV, XML,HTML, and JSON, txt) FAIR metrics ● Knowledge Representation Language: Metric Identifier - FM-I1 : # of identifiers in data/datasets/repositories metadata or # of URLs to the specification of the language (eg BNF) ● Use of FAIR Vocabularies: Metric Identifier - FM-I2 : # of identifiers in data/datasets/repositories metadata or # of UUIDs

			representing the vocabularies used for (meta)data ● Use of Qualified References: Metric Identifier - FM-I3 : # of identifiers in data/datasets/repositories metadata or # of Linksets (in the formal sense) representing part or all of your resource Open Digital Science ● % of machine-readable data / metadata ● quality of metadata (versioning, volume, data format, description of fields, etc.) ● % of published works using researcher IDs OSI 2016-25 “openness score” Format measures: ● Per file formats - # of standardised formats examples: PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml Open Digital Science ● % of open standards in the research process (standards concerning e.g. the provision of data + metadata, modelling, sharing models, visualisations)

Reusability		journals/publishers	Metrics to Assess Value of Biomedical Digital Repositories ● L1: ○ Documented: Explicit data use terms (ideally formal licenses) should be defined by the resource providers and easy to find (yes/no) ○ Clear: a. At a minimum, licenses/data use agreements must be clear and easy to understand. A variety of specific examples of data use/reuse conditions should be included. (yes/no) b. Licenses should not require negotiation and licenses themselves should be legally redistributable without engaging legal counsel (yes/no) ○ Minimally restrictive: The licenses and/or data use agreements should explicitly permit downstream data reuse, derivation, and re-dissemination (yes/no) ○ Standard licenses: For data, ideally CC0., Standard software license: For software, ideally Apache version 2 (valid options include GPLv2, GPLv3, AGPLv3, etc) ○ Contactable: person -contact for license information (yes/no) ● L2: Transparent about flowthrough implications # of sources identified with flowthrough implications # of links to the original licenses/data use terms of all redistributed content % of authorization for redistribution FAIR metrics - Accessible Usage License: Metric Identifier - FM-R1.1 : # of identifiers in data/datasets/repositories metadata or # of GUIDs of the license (e.g. its URL) for the data license and for the metadata license - Detailed Provenance: Metric Identifier - FM-R1.2 : # of identifiers in data/datasets/repositories metadata or # of URLs points to one of the vocabularies used to describe citational provenance (e.g. dublin core) + # of URLs points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM) - Meets Community Standards: Metric Identifier - FM-R1.3 : # of identifiers in data/datasets/repositories metadata or certification saying that the resource is compliant

		funders/governments	Open Data Portal Licensing norms # of countries that recommends open data license % of data available under open license % of data available free of charge

		research performing organisations/ researchers	Metrics to Assess Value of Biomedical Digital Repositories ● L1: ○ Documented: Explicit data use terms (ideally formal licenses) should be defined by the resource providers and easy to find (yes/no) ○ Clear: a. At a minimum, licenses/data use agreements must be clear and easy to understand. A variety of specific examples of data use/reuse conditions should be included. (yes/no) b. Licenses should not require negotiation and licenses themselves should be legally redistributable without engaging legal counsel (yes/no) ○ Minimally restrictive: The licenses and/or data use agreements should explicitly permit downstream data reuse, derivation, and re-dissemination (yes/no) ○ Standard licenses: For data, ideally CC0., Standard software license: For software, ideally Apache version 2 (valid options include GPLv2, GPLv3, AGPLv3, etc) ○ Contactable: person -contact for license information (yes/no) ● L2: Transparent about flowthrough implications # of sources identified with flowthrough implications # of links to the original licenses/data use terms of all redistributed content % of authorization for redistribution FAIR metrics Measuring Reusability - Accessible Usage License: Metric Identifier - FM-R1.1 : # of identifiers in data/datasets/repositories metadata or # of GUIDs of the license (e.g. its URL) for the data license and for the metadata license - Detailed Provenance: Metric Identifier - FM-R1.2 : # of identifiers in data/datasets/repositories metadata or # of URLs points to one of the vocabularies used to describe citational provenance (e.g. dublin core) + # of URLs points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM) - Meets Community Standards: Metric Identifier - FM-R1.3 : # of identifiers in data/datasets/repositories metadata or certification saying that the resource is compliant

			(Re)usable Data Project ● % of public, discoverable, and standard licenses ● % of license requires no further negotiation and its scope is both unambiguous and covers all of the data ● % of data accessibility (data referring to/ covered by the license) ● % of data/datasets with license that has restrictions or no restrictions for re-use ● % of data/datasets with "re-users" license restrictions

Openness of software/code

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Openness	OA routes & costs		
		funders/governments	HowOpenIsIt: Guide to Research Funders policies ● % of costs coverage by funder (all, specific dollar value) and # of fees - types (APCs, data storage) and % of costs already included in the grant application vs not, and time of costs coverage by funder (not time limit, one year after grant expiration, on a case-by case basis)

		research performing organisations/ researchers	A new way of measuring openness: the open governance index ● Is source code freely available to all developers, at the same time? ● Is source code available under a permissive OSI-approved license? ● Developer support mechanisms – are project mailing lists, forums, bug-tracking databases, source code repositories, developer documentation, and developer tools available to all developers? ● Is the project roadmap available publicly? ● # of project meeting minutes/discussions publicly available such that it is possible to understand why and how decisions are made relating to the project? **Development** ○ Transparency of contributions and acceptance process – is the code contribution and acceptance process clear, with progress updates of the contribution provided (via Bugzilla or similar)? ○ Transparency of contributions to the project – can you identify from whom source code contributions originated? ○ Accessibility to become a committer – are the requirements and process to become a committer documented, and is this an equitable process (i.e., can all developers potentially become committers?). Note that a “committer” is a developer who can commit code to the open source project. The terms “maintainer” and “reviewer” are also used as alternatives by some projects. ○ Transparency of committers – can you identify the committers to the project? ○ Does the contribution license require a copyright assignment, a copyright license, or patent grant? **Derivatives** ○ Are trademarks used to control how and where the platform is used via enforcing a compliance process prior to distribution? ○ Are go-to-market channels for applications derivatives constrained by the project in terms of approval, distribution, or discovery? **Community Structure** ○ Is the community structure flat or hierarchical

			(i.e., are there tiered rights depending on membership status?)
Licenses		research performing organisations/ researchers	Choose a license ● # of MIT licenses (permits use with attribution to the creator), ● # of the Apache license 2.0 (permits use with attribution to the creator securing patent rights) ● # of the GNU GPLv3based (copyleft-share alike securing patent right)

FAIRness of software/code

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Findability	Metadata PIDs Directories	journals/publishers	OSI 2016-25 "openness score" Availability measures: -metadata quality - % of metadata completeness -discoverability -crawling -machine readability -links to other resources # of papers with links or annotations -public access to usage data # of usage data available to the public

		research performing organisations/ researchers	FAIR Software ● indexed in searchable resource ● assigned a unique identifier ● use of code repository for development (GitHub, GitLab, Bitbucker) ● use of digital repository for preservation ● described with rich metadata (eg # of README files), # of software described with CodeMeta (an exchange schema for software metadata) ● # of issued DOIs, # of ORCIDs OSI 2016-25 "openness score" Availability measures: -metadata quality - % of metadata completeness -discoverability -crawling -machine readability -links to other resources # of papers with links or annotations -public access to usage data # of usage data available to the public
Accessibility	Protocols	funders/governments	Open Digital Science ● accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects HowOpenIsIt: Guide to Research Funders policies ● # of code and/or software in open repositories + available immediately upon paper publication or after grant has expired or after embargoes (6-12 months); # of DMPs submitted + their permissions on code and/or software access

Interoperability	Machine Readability Standardisation	journals/publishers	OSI 2016-25 "openness score" ● # of standardised formats examples: PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml
		funders/governments	Open Digital Science ● % of published works using researcher IDs
		research performing organisations/ researchers	OSI 2016-25 "openness score" ● # of standardised formats examples: PDF, PDF-A, HTML, embedded figures, tables, csv, xls, json, xml
Reusability		journals/publishers	FAIR Software ● # of community standards for software (eg ESIP https://esipfed.github.io/Software-Assessment-Guidelines/, IPOL https://tools.ipol.im/wiki/ref/software_guidelines/) ● Software Package Data Exchange (SPDX), # of open source licenses (list here: https://opensource.org/licenses) ● # of Jupyter notebooks (for provenance)
		funders/governments	HowOpenIsIt: Guide to Research Funders policies ● % of research outputs with generous reuse & remixing rights (e.g., CC BY/CC0 licenses) or moderate reuse & remixing rights (e.g., CC BY-SA license) or commercial modification rights (e.g., CC BY-NC license) or modification rights (e.g., CC BY-ND license) or # of DMPs with re-use permissions

		research performing organisations/ researchers	FAIR Software ● # of community standards for software (eg ESIP https://esipfed.github.io/Software-Assessment-Guidelines/, IPOL https://tools.ipol.im/wiki/ref/software_guidelines/) ● Software Package Data Exchange (SPDX), # of open source licenses (list here: https://opensource.org/licenses) ● # of Jupyter notebooks (for provenance) OSI 2016-25 "openness score" ● Creative Commons ● free to read ● free to mine ● embargoed and embargo length ● pay-walled

Trustworthiness of publications repositories

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Trustworthiness	Archiving/long term preservation	Research Performing Organisations/ researchers Journals/publishers	Metrics for Openness ● Preservation-Friendly Openness Index: # of institutional, subject repositories, online journals and libraries that authors' have published their work SPARC HowOpenIsIt? A Guide for Evaluating the Openness of Journals ● % of pre-prints, post-prints in # of repositories vs # of datacenters (specific vs free of choice), ● # of national vs # of international infrastructure (specific vs free of choice), ● # of disciplinary-specific vs # generic infrastructure (specific vs free of choice)
	Certification	Research Performing Organisations/ researchers Journals/publishers	European Framework for Audit and Certification of Digital Repositories ● # of repositories holding basic (coretrustseal) certification ● # of repositories holding extended (nestorseal) certification ● # of repositories holding formal (ISO16363) certification CoreTrustSeal ● % of organisational infrastructure (check the 6 requirements) ● % of digital object management (8 requirements) ● % of technology (2 requirements) OSI 2016-25 “openness score” ● Official certification -Yes - # of certified repositories (coretrustseal, ISO, etc) -No -No but committed to long-term preservation # of repositories with provisions for sensitive data, back ups etc

Trustworthiness of research data repositories

Target	Subtarget	Stakeholder	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Trustworthiness	Archiving/long term preservation	Research Performing Organisations/ researchers Journals/publishers	Metrics for Openness ● Preservation-Friendly Openness Index: # of institutional, subject repositories, online journals and libraries that authors' have published their work SPARC HowOpenIsIt? A Guide for Evaluating the Openness of Journals ● % of pre-prints, post-prints in # of repositories vs # of datacenters (specific vs free of choice), ● # of national vs # of international infrastructure (specific vs free of choice), ● # of disciplinary-specific vs # generic infrastructure (specific vs free of choice)
	Certification	Research Performing Organisations/ researchers Journals/publishers	European Framework for Audit and Certification of Digital Repositories ● # of repositories holding basic (coretrustseal) certification ● # of repositories holding extended (nestorseal) certification ● # of repositories holding formal (ISO16363) certification CoreTrustSeal ● % of organisational infrastructure (check the 6 requirements) ● % of digital object management (8 requirements) ● % of technology (2 requirements) OSI 2016-25 "openness score" ● Official certification -Yes - # of certified repositories (coretrustseal, ISO, etc) -No -No but committed to long-term preservation # of repositories with provisions for sensitive data, back ups etc

Trustworthiness of software/ code repositories

Target	Subtarget	Stakeholder	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Trustworthiness	Archiving/long term preservation	Research Performing Organisations/ researchers Journals/publishers	FAIR Software ● use of digital repository for preservation Metrics for Openness ● Preservation-Friendly Openness Index: # of institutional, subject repositories, online journals and libraries that authors' have published their work SPARC HowOpenIsIt? A Guide for Evaluating the Openness of Journals ● # of repositories vs # of datacenters (specific vs free of choice), ● # of national vs # of international infrastructure (specific vs free of choice), ● # of disciplinary-specific vs # generic infrastructure (specific vs free of choice)
	Certification	Research Performing Organisations/ researchers Journals/publishers	European Framework for Audit and Certification of Digital Repositories ● # of repositories holding basic (coretrustseal) certification ● # of repositories holding extended((nestorseal) certification ● # of repositories holding formal (ISO16363) certification CoreTrustSeal ● % of organisational infrastructure (check the 6 requirements) ● % of digital object management (8 requirements) ● % of technology (2 requirements) OSI 2016-25 “openness score” ● Official certification -Yes - # of certified repositories (coretrustseal, ISO, etc) -No -No but committed to long-term preservation # of repositories with provisions for sensitive data, back ups etc

Open Education resources

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Open Educational Resources - Impact		Research Performing Organisations/ researchers Journals/publishers Governments/Funders	Open Digital Science ● # of PhD theses using OS OR on OS-relevant topics ● # (and quality) of publications on OS (as a subject) ● # of curricula for data science or other new roles ● # of graduates in data science per year OER World Map ● # of organisations ● # of persons ● # of services ● # of projects ● # of stories ● # of primary sector (cross-sector, early childhood education, further education, higher education, school, vocational education) ● # of secondary sector (cross-sector, early childhood education, further education, higher education, school, vocational education) ● audience (first stage of tertiary education, lower secondary education or second stage of basic education, post-secondary non-tertiary education, pre-primary education, primary education or first stage of basic education, second stage of tertiary education, upper secondary education) ● # of licenses (copyright, creative commons, other) ● # of subjects ● # of awards OER Impact Map ● Performance: OER improve student performance/satisfaction (yes/no) ● Openness: People use OER differently from other online materials (yes/no) ● Access: OER widen participation in education (yes/no) ● Retention: OER can help at-risk learners to finish their studies (yes/no) ● Reflection: OER use leads educators to reflect on their practice (yes/no) ● Finance: OER adoption brings financial benefits for students/institutions (yes/no) ● Indicators: Informal learners use a variety of indicators to choose an OER ● Support (informal): Informal learners develop their own forms of study support (yes/no) ● Transition: Open education acts as a bridge to formal education (yes/no) ● Policy: OER use encourages institutions to change their policies (yes/no)

			• Assessment: Informal/open assessments motivate learners using OER (yes/no)

Skills - Expertise and Uptake in Training		Research Performing Organisations/ researchers Journals/publishers Governments/Funders	Open Data Barometer ● # of courses in web technologies, data science, data visualisation, legal aspects, business aspects (patents) OECD ● # of data literacy programs Open Digital Science ● % of research personnel / research disciplines skilled in OS ● % of research personnel active in OS ● % of research personnel aware of standards (is there a standard (relevant to open science), how to we adhere to it, etc.) ● % of research personnel familiar with those standards ● % of curricula that include OS skills (also prior to higher education) ● # of curricula for data science or other new roles ● # of graduates in data science per year ● # of initiatives/training programmes for citizens to engage in science/research ● % of citizens with science literacy Deliverable 7.1 Open Data Science Stewardship skills (p. 40-41) ● Data Management skills - DM (ranging from those required to make data FAIR across domains, to those required to make data actionable for research in at least one domain) ● Data Science Engineering skills - DE (e.g. requirements engineering, scripting or programming, software engineering, database management, security and authentication, storage management) ● Data Science/Analytics skills - DSA (predictive modelling, machine learning, text/ data mining, data integration, or visualisation) ● Domain Research skills - DR (e.g. to enhance research methods or their application to collaborative research) EOSCpilot competence group (p. 60-63): Skills in OS lifecycle activities: ● plan and design (DM/DR, DE) ● capture and process (DE, DM/DR) ● integrate and analyse (DSA, DE, DM/DR) ● appraise and preserve (DM/DR) ● publish and release (DSA, DE, DM/DR) ● expose and discover (DM/DR)

			• govern and access (DE, DM/DR) • scope and resource (DE, DM/DR) • advise and enable (DE, DM/DR)

Research Collaboration - peer review

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Transparency	Peer review	Research Performing Organisations/ researchers Journals/publishers Governments/Funders	What is OPR? A systematic review ● % of Open identities: Authors and reviewers are aware of each other's identities (eg From Wiley website: Types of non open identities in peer review: Single blind review - author does not know who the reviewers are Double blind review - the reviewers don't know the identity of authors, and vice versa) ● % of Open reports: Review reports are published alongside the relevant article. ● % of Open participation: The wider community are able to contribute to the review process. ● % of Open interaction: Direct reciprocal discussion between author(s) and reviewers, and/or between reviewers, is allowed and encouraged. ● % of Open pre-review manuscripts: Manuscripts are made immediately available (e.g., via pre-print servers like arXiv) in advance of any formal peer review procedures. ● % of Open final-version commenting: Review or commenting on final "version of record" publications. ● % of Open platforms ("decoupled review"): Review is facilitated by a different organizational entity than the venue of publication OPR in Royal Society Open Science ● % of open licenses (eg CC-BY) of referee reports What constitutes peer review of data - A survey of published peer review guidelines ● # of data papers including information about reviews ● # of Policies - Review criteria for datasets Open Peer Reviewed Module ● # of repositories having the OPR module Crossref's metadata schema for peer review ● # of citable peer reviews ● # of discoverable peer reviews ● # of creditable peer reviews

Citizen Science - citizen engagement and societal impact

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Citizen engagement		Research Performing Organisations/ researchers Governments/Funders	Citizen Science and Policy: A European Perspective ● policy application areas (differentiate between public policy and policy that facilitates citizen science) ● level of engagement and the type of citizen science activity (passive sensing, volunteer computing, volunteer thinking, full-scale environmental and ecological observations, participatory sensing, and civic/community science) Open Digital Science ● openness in call for proposals ● early involvement of citizens ● % of citizens engaging in open science ● % of researcher who acknowledge Citizen Science as valid form of research ● % of non-academia (citizens, civil society organisations) represented in advisory boards for research projects/programme ● # of research projects using crowd funding ● provision of affordable sets of public interest data / metadata ● # of initiatives/training programmes for citizens to engage in science/research increase in % of citizens with science literacy ● # of efforts to make open data that are most relevant for the public interest Open Data Portal ● Use of Data profile of portal visitors - % of type of visitors (mostly private sector, mostly public sector, a bit of everything, no idea) % of foreign users % of inclusion of marginalised groups ● # of citizens involved in science initiatives ● circulating results outside the academia

Societal impact		Governments/Funders	Open Digital Science ● Decrease of "emotional gap" between science and society ● credibility of science in the opinion of the public ● advancement in closing the gap between the information rich and the information poor Open Data Portal ● % of (review) results from society perspective (social relevance)

Research Impact

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Research impact		Research Performing Organisations/ researchers Journals/publishers funders/governments	Open Digital Science ● open methods ● % of decrease of bad science / fraud % of researchers perceiving a research career as attractive ● instruments for evaluation of the status of OS ● credibility of science in the opinion of the public ● formal recognition of a variety of contributions along the scientific process (e.g. to the selection of research topics, formulation of hypotheses, project participations, review activities); vs. publish and perish # of pilot initiatives for new reward systems ● decrease of # of researchers who have negative attitude towards failure (negative results of research efforts) ● # of shared laboratories (online) ● usability of simulation results (models, data, and code) ● # of researchers NOT publishing in journals Open Data Portal ● observing the merging or natural (re-)formation of science disciplines successful use cases relevant to data The donut and Altmetric Attention Score (also included in Metrics toolkit) ● # of mentions in: policy documents, news, blogs, twitter, post-publication peer reviews, facebook, sina weibo, wikipedia, google+, linkedin, reddit, faculty1000, Q&A stack overflow, youtube, pinterest Metrics toolkit ● # of amazon ratings and reviews ● # of blog mentions ● # of articles citations ● # of books and book chapters citations ● # of data citations ● # of software citations ● # of articles downloads ● # of books and book chapters downloads ● # of software downloads ● # of comments, likes and shares on facebook ● # of FFa, % of ratings and # of reviews in Faculty of 1000 Prime

			• # of forks, collaborators, watchers in Github • % of ratings and # of reviews in Goodreads • % of h-index • % of journal acceptance rates • # of Impact Factor • # of Mendeley readers • # of monograph holdings • % of monograph sales and ranking • # of mentions in the news • # of mentions in policy documents • Publons score • # of pubpeer comments • relative citation ratio • # of mentions on Twitter • # of wikipedia citations

Economic Impact

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets
Economic Impact		funders/governments	Open Data Barometer • % of cited open data (eg peer-reviewed study or government audit showing the contribution of open data to government efficiency) or same for media citations/credible websites or no evidence market value, better service delivery, assessing other features for impact Open Digital Science • # of proposals applying for funding of OS infrastructure creation and use • % of funded projects incorporating costs for data compilation/publication and maintenance (of the repository/data sets) significant external investment or small scale examples of commercial apps, websites or other businesses built with open data. These remain niche or small scale businesses)

Policy Readiness, Adoption and Compliance

Target	Subtarget	OS Actor	Candidate Indicators (the source of the indicators definition is referenced). The same indicators may apply to multiple targets\subtargets

Policy Readiness		funders/governments	Open Data Barometer ● % of regulations completeness: available to the public for free or at reasonable/minimal costs in a variety of venues (e.g., online, government agency offices + can be accessed by citizens within a timeframe as defined by the law + answers the specific request, with explanations for refusal to release information ● % of Governments supporting innovation through activities (Running competitions, Organising hackathon events, Organising incubators, labs and open data boot camps, Offering grant funding or innovation vouchers) ● consistent, open standards related to data formats, interoperability, structure, and common identifiers when collecting and publishing data + Consistent core metadata + Information to understand the source, strengths, weaknesses, and analytical limitations of the data + Accompanying guidance documentation that is written in clear, plain language + Being transparent about data collection, standards, and publishing processes by documenting these processes online Open Data Portal ● % of data collection frequency ● # of countries ensuring up-to-date metadata ● % of automatic (meta)data upload to the ODP ● % of policy presence ● % of changes in policy management approach (mandates etc) ● % of open (research) data in decision making ● # of countries with an open (research) data repository/ portal OECD ● Data availability: content of the open by default policy ● Data accessibility: content of the unrestricted access to data policy Open Digital Science ● is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no) ● availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research) ● is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no) ● % of funding programmes supporting the promotion of data-intensive research

			• % of funded projects incorporating costs for data compilation/publication and maintenance (of the repository/data sets)
Policy Adoption		Research Performing Organisations/ researchers Journals/publishers funders/governments	Open Digital Science • # of researchers having signed an open science pledge • # of research organisations having signed an open science pledge • # of mandates and assigned roles (catalysts/evangelists) • # of sharing policies in research organisations (sharing of data, organisms, etc.) • % of harmonised sharing policies • directives from the European Commission for openness (yes/no) • ratification of those directives by EU member states / adoption by research organisations • % of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards • # of research organisations where OS is strategically anchored (e.g. in guidelines, strategic documents, target agreements) • # of agreed policies, principles, or contracts of openness (national, EU-level)

ANNEX B. NOADS RESULTS FROM SHORT SURVEY ON NATIONAL OA/OS MONITORING MECHANISMS

Country	Has a national monitoring mechanism	Is developing a national monitoring mechanism	Does not have a national monitoring mechanism
Austria		Nationally funded working group to explore ways in developing a national monitor for open access	
Belgium			No national mechanism as of yet
Denmark	Danish National OA Indicator by the Danish Agency for Science and Higher Education		
Finland			
France			No national monitoring mechanism, but Couperin (library consortium) has provided OA statistics.
Germany			The development of an OA monitor is included in the German Strategy; German initiative for APCs
Netherlands	Has a national monitoring Framework and NARCIS		

Slovakia	OA journals and articles are being monitored and results are published on a website (http://openaccess.cvtisr.sk/homepage/zoznam-casopisov/)		
Turkey		Turkish National Science Foundation (TUBITAK) established an Open Science Committee to explore OS monitoring activities	
United Kingdom	Monitor Local, Monitor UK provided by JISC		

ANNEX C. GLOSSARY

Term	Explanation
Article Processing Charges	fees paid by organisations or individuals in order to publish their research in OA journals. Usually such fees include peer reviewing process but refer to other processes taking place prior to publication, such as editing and articles database hosting charges.
Open Science	(from the OpenAIRE portal) "follows e-Science, i.e. the computationally-intensive/digitalised research process and practice, while positioning collaboration and re-usability as its driving forces. Open Science opens up the research lifecycle, from the concept of an idea and the collection of relevant material (papers, data, etc) to the publication, archiving and re-use of the research outcomes, including metadata and research data. It creates a new modus operandi for science, where all stakeholders (researchers, funders, research performing organisations, ITs, librarians, citizens, even governments) are involved and research is organised, linked, verified, facilitated by new technologies and, enhanced with collaborative and coordinative activities. Legal barriers in accessing and sharing information and data, as well as, utilisation of data-intensive, cost-demanding infrastructures are among the issues that are eliminated with Open Science."
Open Science Resources (OSR)	i.e. research artefacts such as open access publications, FAIR (research) data, open source software etc; open educational resources as part of tertiary education and open scholarship; processes and practices strengthening research collaboration such as open peer review, open data citation, open workflows; citizen science activities engaging the public in research by making use of open science drivers and/or enablers of open science.

Open Science Monitor	(definition) according to the D5.1 The European Open Science Cloud Architecture: Anatomy and Physiology, is a set of services for supporting Research Performing Organisations (RPOs), Research Funding Organisations (RFOs) and Government Bodies to measure: ● levels of compliance with European Union's laws, regulations and policies regarding research and research results dissemination; ● Open Science Resources' (i.e. research artefacts, educational resources, research collaboration, citizen science) levels of openness, trustworthiness and FAIRness that cover each stage of the research lifecycle; ● impact of science on society and economy
EOSC Open Science Monitor (OSM)	a monitoring framework of specifications measuring open science elements in the context of the European Open Science Cloud (EOSC). It can be used by stakeholders interested in developing such mechanism to get a better view/understanding of open science activities performance within their organisation, field. EOSC OSM is envisioned to become the European (Open) Science Monitor service as described in D5.1: The European Open Science Cloud Architecture: Anatomy and Physiology (please see also definition above). Such achievement is feasible with iterations of information/data received from the two-way communication established/ensured between the EOSC OSM and the stakeholders.
EOSC OSM Stakeholder	also considered as actors. Please also see below for EOSC OSM Actor.
EOSC OSM Actor	since EOSC OSM consists of a two-way communication process between stakeholders and is rather dynamic, actors may play both direct and indirect roles in its ecosystem. Actor of the EOSC OSM is thus defined as the research admin who uses the proposed specifications to build a monitoring mechanism as well as an (end) user who provides the EOSC OSM with information about new Open Science trends or metrics.
Indicators	are proposed measurable sets highlighting targets and expected outcome of research results, processes and trends in Open Science.

Monitoring Target and Sub-targets	are specific monitoring goals derived by the mapping activity of the landscape review. Monitoring targets and sub-targets are the main drivers of the monitor providing a more coherent framework between different approaches of stakeholders/actors OS activities.
Research Admin	according to the D5.1 The European Open Science Cloud Architecture: Anatomy and Physiology, research admins are primary actors in generating and maintaining a monitoring culture in an organisation since they "are involved in informing their organization leadership in advising on policies or IPR. In order to address their mission, they perform activities that aggregate research results for their organization in order to produce metrics and indicators allowing to assess the past (impact) and shape the future (trends)."