Plan-Track-Assess Pathways

Deliverable 1.1

Deliverable Number	D1.1
Deliverable Title	Plan-Track-Assess Pathways
Version nº	1.0
Status	Final
Type	Report
Dissemination Level	Public
Work Package	1
Lead Beneficiary	Graz University of Technology
Delivery Date	31 July 2024
Author(s)	Stefan Reichmann, Miguel Rey Mazón, Ilire Hasani-Mavriqi, Laura Thaci, David Eckhard
Editor(s)	Elli Papadopoulou (ARC)
Reviewers	Tomasz Miksa (TU Wien), Natalia Manola (OpenAIRE), Daniel Spichtinger, Victoria Eisenheld (Uni Wien)
Approved by	Natalia Manola (OpenAIRE)

This project has received funding from the European Union's Horizon Europe framework programme under grant agreement No. 101130187. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the European Research Executive Agency can be held responsible for them.

Funded by the European Union

Revision History

VERSION	STATUS	DATE	DESCRIPTION	AUTHOR(S)
0.0	Draft	15 May 2024	Agreement on structure & References	Miguel Rey Mazón, Stefan Reichmann, Ilire Hasani-Mavriqi
0.1	Draft	17 June 2024	First Draft	Miguel Rey Mazón, Stefan Reichmann, David Eckhard, Ilire Hasani-Mavriqi, Laura Thaci
0.2	Draft	1 July 2024	Second Draft (incorporated feedback WP1-WP3)	Miguel Rey Mazón
1.1	Review	26 July 2024	Addressed feedback by all reviewers	Stefan Reichmann, Miguel Rey Mazón
1.0	Final	29 July 2024	Tables, Arrows, ToC	Miguel Rey Mazón, Stefan Reichmann
1.1	For Approval	31 July 2024	Edited the content for quality assurance	Elli Papadopoulou

Authors List

ORGANISATION	NAME	CONTACT INFORMATION
Graz University of Technology	Stefan Reichmann	stefan.reichmann@tugraz.at, +43 316 873 30687
	Miguel Rey Mazón	m.reymazon@tugraz.at, +43 316 873 30684
	Ilire Hasani-Mavriqi	ilire.hasani-mavriqi@tugraz.at, +43 316 873-30679
	Laura Thaci	lthaci@tugraz.at +43 316 873 - 30686
	David Eckhard	david.eckhard@tugraz.at

Contributors List

ORGANISATION	NAME	CONTACT INFORMATION
Technical University of Vienna	Tomasz Miksa, Lukas Arnhold	tomasz.miksa@tuwien.ac.at, lukas.arnhold@tuwien.ac.at
OpenAIRE	Paolo Manghi, Natalia Manola, Tassos Stavropoulos	paolo.manghi@openaire.eu, natalia.manola@openaire.eu, tassos.stavropoulos@openaire.eu
University of Oxford (UOXF)	Susanna-Assunta Sansone; Allyson Lister; Milo Thurston	susanna-assunta.sansone@oerc.ox.ac.uk allyson.lister@oerc.ox.ac.uk milo.thurston@oerc.ox.ac.uk
Tampere University. Finnish Social Science Data Archive (CESSDA)	Katja Moilanen; Matti Heinonen	katja.moilanen@tuni.fi matti.heinonen@tuni.fi
ESRF (PaNOSC)	Renaud Duyme	renaud.duyme@esrf.fr
Centre Inria de l'Université de Grenoble Alpes (ESCAPE)	Maud Medves	maud.medves@inria.fr
Austrian Academy of Sciences (OEAW) (SSHOMP)	Matej Ďurčo	matej.durco@oeaw.ac.at
CLARIN/KNAW Humanities Cluster	Menzo Windhouwer	menzo.windhouwer@di.huc.knaw.nl
Czech Technical University of Prague	Marek Suchánek, Jana Martinková	marek.suchanek@cvut.cz, jana.martinkova@cvut.cz
ObsParis	Baptiste Cecconi	baptiste.cecconi@obspm.fr
Codevence	Jakub Jirka	jakub.jirka@codevence.com
Universidad Politénica de Madrid	Mark Wilkinson; Daniel Garijo	mark.wilkinson@upm.es daniel.garijo@upm.es
KNAW-DANS	Wilko Steinhoff	wilko.steinhoff@dans.knaw.nl
	Wim Hugo	wim.hugo@dans.knaw.nl
LifeWatch ERIC	Christos Arvanitidis; Joaquín López Lérida	ceo@lifewatch.eu joaquin.lopez@lifewatch.eu
CITE	Georgios Kakaletris	gkakas@cite.gr
ATHENA Research Center (ARC)	Elli Papadopoulou	elli.p@athenarc.gr

Table of Contents

List of Figures

List of Tables

Disclaimer

This document contains description of the OSTrails project findings, work, and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure to ensure that its content is accurate, consistent, and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The publication herein was produced in the framework of the EU-funded OSTrails project (Grant Agreement No 101130187). Views and opinions expressed are however those of the OSTrails consortium and do not necessarily reflect those of the European Union or the European Research Executive Agency.

Funded by
the European Union

Abbreviations list

COARA	Coalition for Advancing Research Assessment
(ma)DMP	(machine-actionable) Data Management Plan
DO	Digital Object
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
IRA	Interoperability Reference Architecture
PIDs	Persistent Identifiers
PTA	Plan-Track-Assess d
RDA	Research Data Alliance
RDM	Research Data Management
SKG	Scientific Knowledge Graph

Executive Summary

This report forms the first technical deliverable of the project aimed at identifying and co-designing the Plan-Track-Assess (PTA) framework and the interactions of its components, i.e. the PTA pathways.

The PTA framework constitutes a high-level, tool-independent framework that outlines how researchers, research managers, and funders can plan, track, and assess the management of Digital Objects (DO) in alignment with the FAIR Guiding Principles. Its main components are Data Management Plans (DMPs), Scientific Knowledge Graphs (SKGs), and FAIR assessment tools. To develop the pathways conceptual diagrams of the PTA components were prepared consolidating input received from respective service providers of the consortium, which were then validated by the pilots of the project.

Based on these preparatory activities and input, the OSTrails consortium defined three pathways, each representing user needs from a different perspective.

- The "Plan" Pathway focuses on refining the creation and assessment of Data Management Plans (DMPs), with recommendations for improved integration of DMP platforms with Scientific Knowledge Graphs (SKGs) and repositories, enhanced metadata evaluation capabilities, and user-friendly interfaces.
- The "Track" Pathway addresses the utilization of SKGs to monitor and evaluate the FAIRness of research outputs, recommending enhancements in tracking functionalities, seamless integration with DMP platforms, and advanced FAIR assessment tools.
- The "Assess" Pathway emphasizes the need for comprehensive evaluation tools for DMPs and research outputs, ensuring interoperability with SKGs and repositories, and offering both automated and manual assessment options.

This deliverable is a foundational step in our workplan for it informs the OSTrails architecture highlighting key requirements for Interoperability Frameworks. The findings and recommendations support further technical developments in the project towards enhanced harmonisation and automations.

1. Introduction

As the first of several technical deliverables, D1.1 presents the groundwork for the development of the OSTrails architecture, to be refined within the Interoperability Reference Architecture.[1] The deliverable describes three pathways, each representing user needs from a different perspective, starting from DMPs platforms, SKGs, and FAIR Assessment tools, respectively. The text is structured as follows: Starting from a description of the Plan-Track-Assess Framework, **Section 1** develops a definition of the pathways and a high-level summary of the approach adopted in creating the pathways. The pathways diagrams, including the modifications suggested by project partners, are presented and discussed in **Section 2** and **3** along with the consultation process, detailing how we included the user perspective into the pathways' development and how the pathways diagrams were subsequently refined based on the feedback from the consortium. This is followed by a summary of the process, some conclusions, and recommendations for the architecture (**Section 4**). The annexes present the feedback received from partners during the consultation process and the document and instructions shared to gather their input.

1.1. Background: FAIR implementation and FAIRness literacy

The FAIR Guiding Principles ("FAIR principles" for short) for scientific data management and stewardship (Wilkinson et al. 2016) provide a set of implementation-agnostic principles (Jacobsen et al. 2020) to enable science practitioners to make research outputs (including data, software, etc.) easy to share between machines and support their reuse by others. The FAIR Principles express the idea, not new in itself, that scientific data should be both human- and machine-readable (Boeckhout et al. 2018), condensing disparate efforts to streamline research data management practices (Wilkinson et al. 2016). The FAIR principles mandate that research data be Findable, Accessible, Interoperable, and Reusable, constituting a concise and measurable (Wilkinson et al. 2018) attempt to ensure that digital research objects can be discovered and reused.[2] However, as they do not dictate a specific technical implementation (Purnama Jati et al. 2022), the FAIR principles have been largely open to interpretation, spawning inconsistent, non-interoperable implementations in some cases (Jacobsen et al. 2019). In light of this, Mons et al. (2017)

[1] OSTrails project Milestone 4, submitted in month 6 of the project, describes a FAIRness model baseline as well as a draft of the OSTrails Interoperability Specifications and Architecture. It refines the pilots as well as the project's "Dissemination & Engagement Roadmap".

[2] In this, they remain aspirational insofar as they merely "describe a continuum of features, attributes, and behaviors that will move a digital resource closer to that goal" (Wilkinson et al. 2018).

suggested the development of a high-level interpretation, discussing early emerging misinterpretations of the FAIR foundational principles, and clarifying their original intent. However, as the FAIR principles are specifically intended to enable the findability, accessibility, interoperability, and reusability of data by machines (i.e. computers), they are insufficient to describe the mechanisms to measure the adoption of FAIR *practices* (i.e. the daily activities carried out by science practitioners to ensure that research outputs comply with the FAIR principles).[3] As a consequence, there is a gap between the progress achieved in making research outputs machine-readable (i.e. FAIR-compliant) and the adoption of FAIR data practices (David et al. 2020). It then becomes necessary to develop practical measures to evaluate "FAIRness" that can guide researchers in their daily tasks. This has led in recent years, to the development of a range of frameworks, metrics, and maturity indicators to measure the FAIR compliance of research outputs (Wilkinson et al. 2018; Wilkinson et al. 2019), including software, semantic artefacts, and training materials[4] (Jacobsen et al. 2019) such as the RDA FAIR Maturity Model.[5] Considering “the full range of research outputs, such as scientific publications, data, software, models, methods, theories, algorithms, protocols, workflows, exhibitions, strategies, policy contributions, etc.”[6] requires moving away from approaches based on bibliometric indicators towards using large-scale data collections that provide information on the broader ecosystem of “Digital Objects” (DOs), their use and openness metrics, to track and assess the FAIRness level reached.

To understand the extent to which researchers need support in handling the increasing amounts of research data (Awre et al. 2015), a growing body of research has studied research data management (RDM) practices (Tenopir et al. 2011) together with the uptake of RDM policies (Kalichman et al. 2014), regulation (Higman and Pinfield 2015), and the role of infrastructures (Amorim et al. 2017). These studies find that the adoption of the FAIR principles, and of RDM practices more generally, is complicated by the situatedness/contextuality of research data (Bezuidenhout et al. 2017; Leonelli 2017), along with varying definitions of openness (Levin et al. 2016; Levin and Leonelli 2017). Research data is highly context-dependent with respect to data formats meanings and uses. The context of data (re)use varies, for instance, with the readiness level of data formats (Weinshall and Epstein 2020) and with the involvement of human subjects (Cychosz et al. 2020), with the consequence that “one-size-fits-all” approaches risk privileging research fields that are already more prone to FAIR data practices (Ross-

[3] As the EOSC Task Force (TF) on FAIR Metrics and Data Quality (https://eosc.eu/advisory-groups/fair-metrics-and-data-quality/) has highlighted: "FAIR compliance is currently “stuck” between being an increasingly common research and publishing requirement while remaining an unmeasurable set of ideals.”
[4] In early 2022, the Coalition for Advancing Research Assessment (COARA) called for a fresh approach to research evaluation. Drafted by a group representing the European University Association, Science Europe, the EC, the agreement has almost 500 signatories (as of March 2023).
[5] https://zenodo.org/records/3909563#.YGRNnq8za70
[6] https://coara.eu/agreement/the-agreement-full-text/

Hellauer et al. 2022). Assessment of the benefits of FAIR data practices, such as increased citation rates (Piwowar et al. 2007), better reproducibility of findings (Gilmore et al. 2017), and improved research quality (Fecher et al. 2015), have often failed to appreciate these differences (Rappert and Bezuidenhout 2016). Another well-documented gap concerns the recognition of the importance of (FAIR) data skills, and their prevalence in day-to-day research practices (Fuhr 2019), with considerable differences between scientific disciplines (Kalichman et al. 2015). Where open research practices are not standard, researchers are often described as lacking FAIRness literacy (David et al. 2020), despite possessing the relevant data literacy (Atenas et al. 2020).

To remedy this situation, Data Management Plans[7] (DMPs) have emerged as tools for effective data stewardship (Bishop et al. 2023), as a precursor to implementing the FAIR principles throughout a project's lifecycle[8] (Jones et al. 2020). However, to date there is no evidence that DMPs have resulted in better RDM practices[9] or FAIRer digital objects. Since DMPs continue to be text-based (Pham et al. 2023) at present, with their creation often happening as an afterthought, it has been suggested that DMPs need to be transformed from plans defined at the beginning of a project into living, (machine)actionable guidance to allow for automatic, standardized exchange, integration, and validation (Miksa et al. 2019), to be used and revised at any point in the research life cycle, and enable FAIR-by-design (Jones et al. 2020). Making DMPs machine-actionable in the required sense is supported by a de-facto convergence of DMP requirements across funders (Jones et al. 2020: 209). There have been attempts to increase the FAIRness of DMPs by treating them as DOs and by including maDMPs in SKGs (Papadopoulou et al. 2020). These large networks describe the actors, research products, contextual information, and research knowledge, as well as their relationships, and are becoming increasingly popular as infrastructures for representing scholarly knowledge.[10] There is a variety of stand-alone SKGs, collecting, cleaning, integrating, organising, and publishing research outputs within or across disciplines. As a result, the landscape of SKGs is fragmented, making the need for some level of interoperability critical. Realising the above vision requires transitioning towards more networked scholarship, without losing the specificity that each discipline needs.

---

[7] DMPs "are documents accompanying research proposals [to] describe the data that are used and produced during [...] research, where the data will be archived, [...] licenses and constraints [...], and [...] credit." DMPs are "awareness tools to help researchers manage their data and ensure that it will be of high quality, accessible, and reusable after the project has ended". They are "typically created manually, mostly by researchers using checklists and online questionnaires" (Miksa et al. 2019).

[8] Their dissemination has led to the establishment of supporting roles within Research Performing Organisations (RPOs) to take on the development and promotion of DMPs and FAIR-enabling practices and tools to assist researchers (Losoff 2009; Das 2018, Boeckhout 2018).

[9] https://www.rd-alliance.org/active-data-management-plans-are-we-there-yet/

[10] Popular SKGs include OpenAlex, AMiner, ScholarlyData.org, Semantic Scholar, PID Graph, Open Research KG, Computer Science KG, OpenCitations.

1.2. The Plan-Track-Assess Framework

To support the adoption of FAIR data practices across disciplines and countries, this complex landscape of tools, services and practices needs to be superseded by a framework which codifies solutions to similar problems appearing in different scientific or geographical contexts, based on the actual needs of research communities. The Plan-Track-Assess framework (**Figure 1**)[11] put forward by the OSTrails project provides such a high-level, tool-independent framework that outlines how researchers, research managers, and funders plan, track, and assess the management of Digital Objects in alignment with the FAIR Principles.

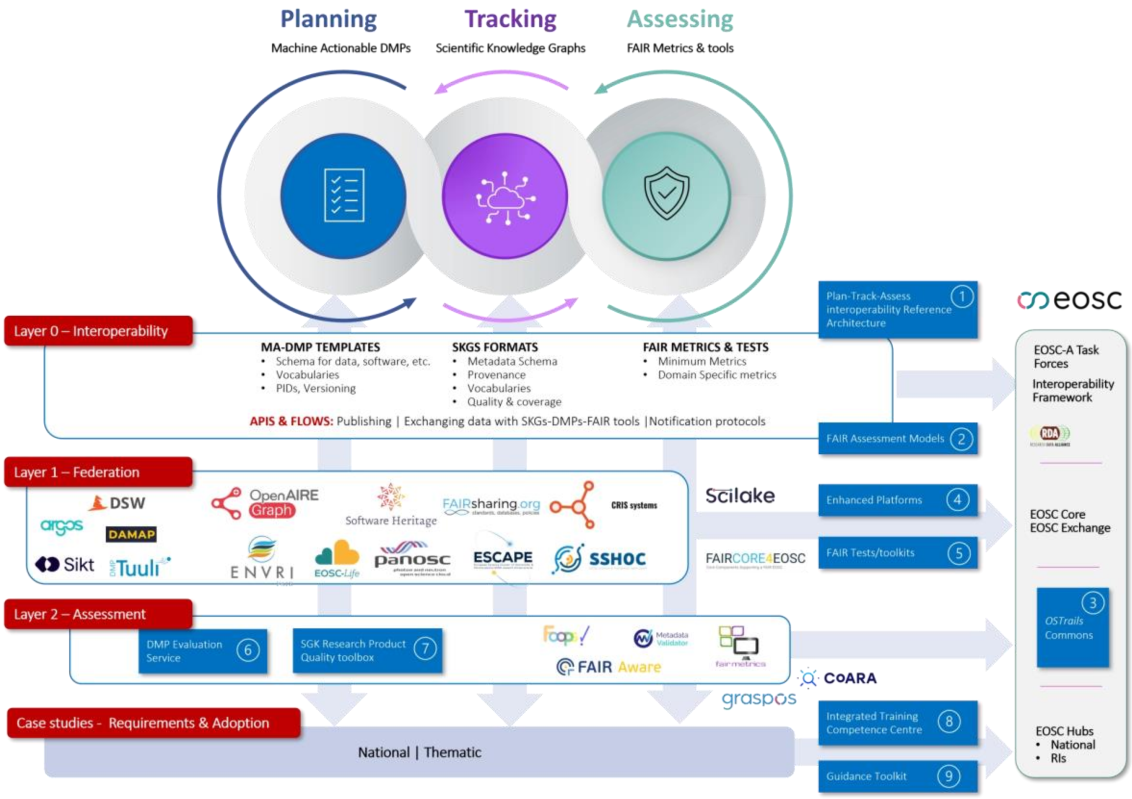


Figure 1: Plan-Track-Assess process

In what follows, we use the term "Digital Object" (DO) for pieces of information in digital format on which the tools of the PTA framework act.[12] Digital Objects can be broadly defined as "a sequence of bits identified by a persistent identifier and described by metadata" (Berg-Cross et al. 2015; see also Schwardmann 2020; De Smedt et al. 2020).

[11] See https://ostrails.eu/methodology for full-size image.
[12] "Digital Object" replaced the term "record" employed in the first discussions in the project, and in the diagrams and subsequent consultation.

For our purposes, a DO spans from research outputs in digital form (i.e. datasets or data) to any other digital information related to research activities (e.g. publications, software, protocols, peer-reviews, patents, DMPs, FAIR tests, etc.), and also includes digital information about physical or abstract objects.

The Plan-Track-Assess framework comprises three main components: Data Management Planning platforms, Scientific Knowledge Graphs, and FAIR assessment tools. By identifying the assumptions underlying the (context-dependent) practices of researchers and co-creating a framework built by interoperable tools and services, it will be possible to extend the “FAIR ecosystem” to serve researchers across purposes and borders.

1.3. Towards Open Science Pathways

A quick glance at the landscape of available tools to implement the FAIR principles is sufficient to justify the need for a high-level view like that provided by the PTA framework, as different research communities have created a wide variety of solutions for each of the three pillars of the PTA framework. For example, the OSTrails consortium, which represents a small sample of the European landscape despite its size (it comprises 38 partners across the EU), already encompasses over 80 different tools as particular implementations of the three components—DMPs, SKGs, and FAIR tools—of the PTA Framework. Although they are all related, only FAIR Assessment tools originate from the FAIR principles. DMPs and SKGs do not originate from the FAIR principles but are part of infrastructures and ecosystems supporting research workflows, and need to enable FAIRness as a consequence, by adopting best practices to e.g. assign PIDs for data published in catalogues.

The solutions differ in their design and implementation due to the different needs of the research communities they serve, sometimes rendering them incompatible with each other and making it difficult to see the relations among them. This results in a suboptimal use of resources due to duplicate solutions to the same problem that can drive away less advanced users, or those who have fewer incentives to adopt FAIR practices in their daily tasks. OSTrails aims at reducing the complexity of the ecosystem by increasing the interoperability of existing tools.

To that end, the consortium identified common practices in different scientific disciplines and geographical locations (i.e. national/institutional research settings) to describe how existing tools may be adapted or extended in a way that fits different needs and requirements. It is these practices we refer to as *Open Science pathways*, or *pathways* for short. The pathways represent the interplay between the components of the PTA framework as they are employed by different user groups; as is typical in software system design, they are represented in relatively simple diagrams that provide a sufficiently abstract representation of the underlying practices. They abstract from the details of individual use cases by focusing on how the tools and services of the PTA framework are used in practice by different actors, with a focus on the *role* in the data lifecycle, rather than on *how* they achieve it. The pathways thus contribute to creating an OSTrails Plan-Track-Assess Interoperability Reference Architecture (OSTrails-IRA), representing generic conceptual components to a logical view of the OSTrails architecture. The PTA framework components are shown as rounded boxes and the interactions between them as labelled arrows (see **Figure 2**). The beginning of an arrow shows from which tool an interaction originates, with the tip pointing to the tool it targets.

Another way to understand the pathways is as representations of the actions a user can perform starting from a certain tool. Defining joint pathways for DMP platforms, SKGs, and FAIR Assessment tools thus meant determining how these tools are interconnected, to understand how different users employ them by way of devising persona scenarios/user stories,[13] and documenting each of these scenarios to understand what exists already, what is needed, and what is currently missing that could therefore suggest areas for improvement. The work builds on the expertise of the OSTrails consortium in the various stages of research data management, the outcomes of the RDA Working Groups on SKGs[14] and maDMPs,[15] the outcomes of the EOSC Association FAIR Metrics and Data Quality Task Force,[16] collaborations with other EOSC-related projects such as SciLake[17] and FAIRCORE4EOSC,[18] and the analysis of user stories discussed at the International Data Curation Conference (IDCC) 2024 workshop (Sisu et al. 2024).

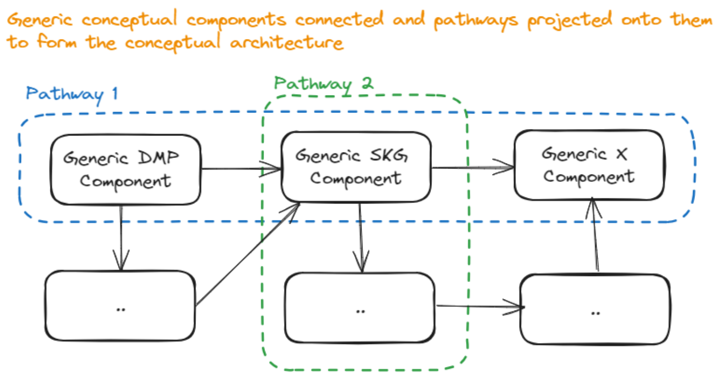


Figure 2: Generic conceptual components and projected pathways

[13] https://github.com/RDA-DMP-Common/user-stories
[14] https://www.rd-alliance.org/groups/scientific-knowledge-graphs-interoperability-framework-skg-if-wg/
[15] https://archive.rd-alliance.org/groups/dmp-common-standards-wg
[16] https://eosc.eu/advisory-groups/fair-metrics-and-data-quality/
[17] https://scilake.eu/
[18] https://faircore4eosc.eu/

2. Methodology

To provide a high-level, tool- and service-independent conceptual overview of the three pillars of the PTA framework (DMP platforms, SKGs, and FAIR assessment tools), we created *user stories.* These stories show how the PTA framework's components are utilised across different stages of research:

- Planning (Plan): How Digital Objects (DOs) are planned for collection, processing/analysis, and storage before research begins.
- Tracking (Track): How the (re)use of DOs is tracked during the research process.
- Assessing (Assess): How compliance with the FAIR principles for DOs is assessed.

While pathways are analytical instruments showing the relationships between the generic conceptual components of the OSTrails architecture, generic user stories codify common ideas, underlying tools and services, thus preparing the ground for modifications/adaptations to become (more) interoperable or indicating new components that need to be created to fill the gaps identified with "built-in" interoperability.[19] To achieve this, we evaluated the abstract (i.e. implementation-independent) pathways diagrams against actual research practices to understand how the various components are currently used by different stakeholders. This task requires connecting the (relatively) abstract diagrams with actual ("real world") practices adopted by researchers. To accomplish this, we retrieved the expertise from project partners, including the use cases represented by pilot projects.[20]

In order to abstract from concrete instantiations of the architecture, we started from the data lifecycle concept as representing data practices from the perspective of data (re)use (see **Figure 3**)[21] to develop initial sketches based on the PTA framework.

[19] These two complementary tasks will be carried out elsewhere in OSTrails.

[20] Recognising that there is no one-size-fits-all solution, and to ensure buy-in at scale, a total of 15 national (HR, GR, NL, NO, SRB, AUT, PL, IL, PT, FR, FL, DE, ES, CZ, and SWE) and 9 thematic pilots (physics, marine/coastal science, cross-domain, social sciences, SSH, linguistics, astronomy, and biodiversity) will engage RPOs, RFOs, and all other organizations in the ESFRI Clusters to co-create, implement, validate, and adopt OSTrails results.

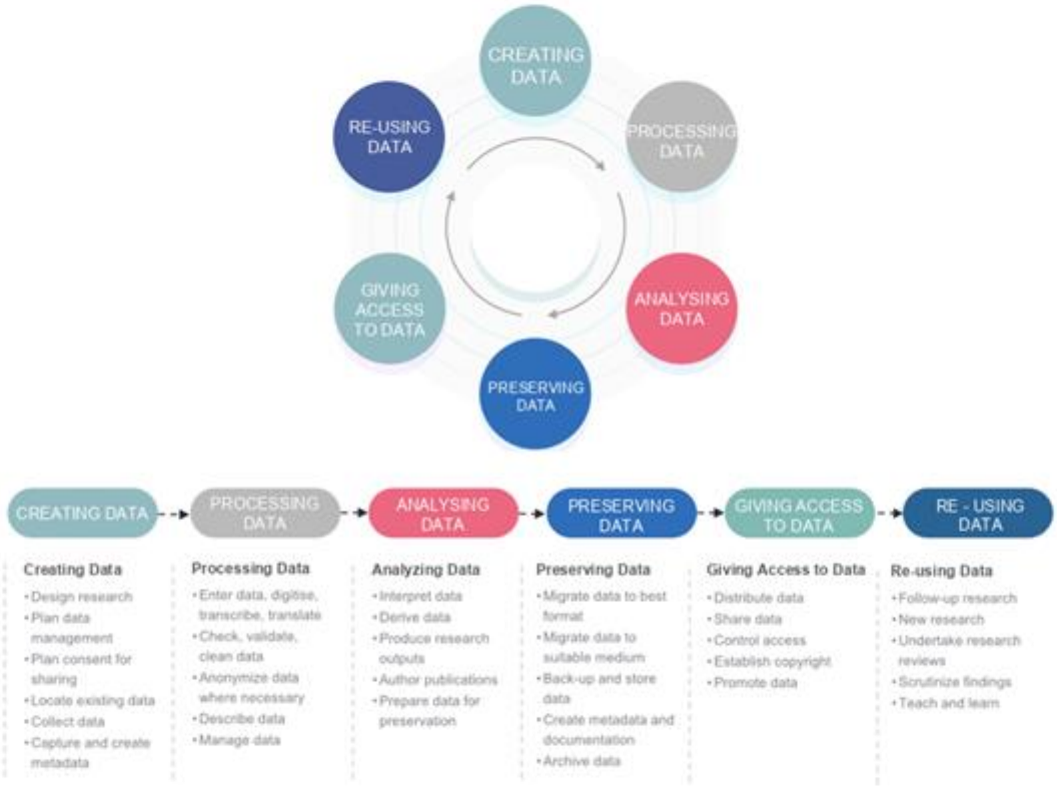


Figure 3: The research data lifecycle

The sketches were then subjected to the scrutiny of the consortium. The discussion about pathways made partners aware of the differences among them in the theoretical framework and practical solutions they have employed to implement tools and services.

Feedback from partners was collected, analysed, and iteratively incorporated to refine the pathways diagrams, resulting in final versions that guide the consortium's future work. The pathways diagrams were developed as abstract representations of system components, their interactions, and user actions using a graphical methodology typical for system design and analysis. Components and interactions were depicted using "black boxes" to represent tools in the PTA framework and arrows to represent actions performed by ideal user types. This methodology provided the desired level of abstraction, focusing on generic components (represented by the boxes) and their interactions (represented by arrows). Different user group perspectives (e.g., researchers, research managers, funders) were incorporated via the direction and attachment of arrows. The specific action or type of information exchanged between components is shown by a label added to the arrows. Most of the interactions depicted target specific research outputs (e.g. datasets), with some exceptions, since an SKG may contain information on organisations, repositories, or other entities.
The target user perspective is represented by displaying only the steps that may be actioned by users and leaving out those carried out automatically by machines. However,

the focus remains on "abstract" user actions, meaning that it is the types or classes of integrations or interfaces that matter, and not the individual case or specific implementation. It is also important to note that the diagrams show the desired ("ought") future state rather than the present.

The user stories represented in the pathway diagrams show how components are used to deliver value to stakeholders, and how they interact with each other. This "use case view"[22] only depicts in an abstract fashion the integrations and interfaces related to OSTrails' PTA framework that users *should* be able to do, without reference to any particular implementation. Below, we give an overview of the initial sketches:

A) "**Plan" stage (starting from DMPs):** When preparing a research project, a researcher creates a Data Management Plan using a DMP Tool following the request from funders, or on their own initiative. For this, they may search for "records" (i.e. Digital Objects, see previous paragraph) in a repository, or retrieve their metadata via SKGs, or (later in the process) evaluate completeness and FAIRness of a DMP via a DMP evaluator (**Figure 4**).

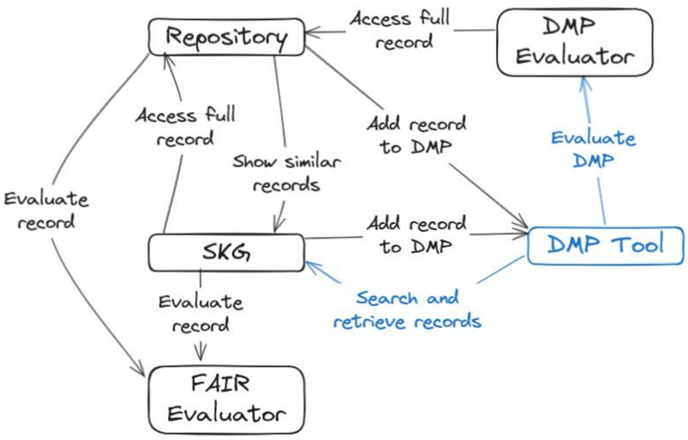


Figure 4: "Plan" Pathway

B) "**Discover/Track" stage (starting from SKGs):** During the active research phase, SKGs are typically employed by researchers to access a full record of a Digital Object (DO) in a repository, evaluate the FAIRness of metadata via a FAIR evaluator, and/or add a reference to a DO to the DMP via a DMP tool (**Figure 5**).

[22] See e.g. https://guides.visual-paradigm.com/4-1-views-in-modeling-system-architecture-with-uml/ or https://www.jsware.io/uml2/architectural-views.html for brief descriptions of the different architectural views employed.

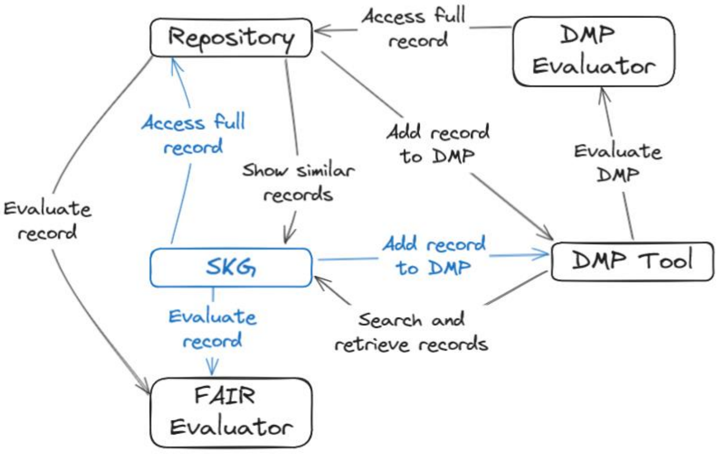


Figure 5: "Track" pathway

C) "**Assess" stage (starting from FAIR Assessment Tools):** In the evaluation of a DMP, a researcher may access full records of the research output mentioned in the DMP via the repository where they are stored (**Figure 6**).

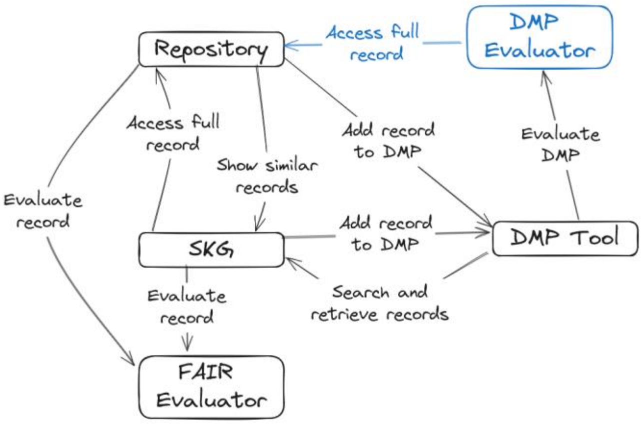


Figure 6: "Assess" pathway

3. Pathways: User Actions and Machine Interactions

In this section, we describe the three pathways, and the changes done to the diagrams after incorporating the feedback received from project partners. Using these results, a roadmap to implement the changes in the remainder of the project will be established by the consortium. The three diagrams represent two viewpoints (see **Table 1**): Each pathway can be interpreted as representing 1) the actions potential users might (or actually do) take (section 3.1), and 2) the potential or actual interactions between the various components (section 3.2). This section is therefore organised in two subsections, each discussing one of the two interpretations for each of the three diagrams. We describe the mature pathways as representing user actions as well as interactions between the PTA framework tools. Taken together, these interpretations constitute the OS pathways which are the main result of the deliverable. In total, we developed three pathways, each allowing for two distinct interpretations, either as representing user actions or as representing tool interactions.

Table 1: Pathways as user actions versus pathways as tool interactions

IDENTIFIER	TITLE	DESCRIPTION	ACTIONS	FIGURE
Pathway 1-PU	PLAN-User	PLAN-User describes how a potential user might start from a DMP platform to interact with the other components (SKGs, repositories, DMP evaluators).	Search for research output via Repository Search for digital objects via an SKG Retrieve metadata for research output via SKGs/Repositories Add a DMP to a repository Evaluate completeness and FAIRness of a DMP and/or research output listed in the DMP via FAIR Assessment	Figure 8
Pathway 2-TU	TRACK-User	TRACK-User describes user actions that may occur during the tracking phase of a research project and typically start from an SKG to interact with the other components (DMP platforms, repositories, DMP evaluators).	Access full research output via the respective repository Evaluate FAIRness of research output metadata via FAIR Assessment Add research output to a DMP via DMP platform	Figure 9

Pathway 3-AU	ASSESS-User	ASSESS-User describes user actions typically starting from a FAIR assessment tool, to interact with the other components (DMPs, SKGs, repositories).	Access full research output mentioned in a DMP via the respective repository When research output or the Repository itself is not available, users may access metadata and related research output via SKGs	Figure 10
Pathway 1-PT	PLAN-Tools	PLAN-Tools describes how a potential interaction with the other components (FAIR assessment tools, SKGs, repositories) might be triggered from a DMP, either by some other user action, or via a background task.	Request metadata for research output from Repositories Request metadata for digital objects from SKGs Provide user search to SKGs/Repositories Provide DMP metadata to SKGs/Repositories/FAIR assessment/DMP platforms Request DMP metadata from DMP platforms	Figure 12
Pathway 2-TT	TRACK-Tools	TRACK-Tools describes how a potential interaction with the other components (FAIR assessment tools, DMPs, repositories) might be triggered from an SKG, either by some other user action, or via a background task	Provide Repositories research output similar to those that were requested Request metadata for research output from Repositories Provide metadata for research output to Repositories and/or FAIR assessment tools Provide metadata about Digital Objects to SKGs/DMP platforms Request metadata about Digital Objects from SKGs Request DMP metadata from DMP platforms	Figure 13
Pathway 3-AT	ASSESS-Tools	ASSESS-Tools describes how a potential interaction with the other components (SKGs, DMPs, repositories) might be triggered from a FAIR assessment tool, either by some other user action, or via a background task.	Request metadata for research output (via a PID) from an SKGs and/or a repository Provide an evaluation result to SKGs and/or a repository and/or a DMP platform	Figure 14

3.1. Actions by users

Figure 7 (below) generically represents the interactions between the PTA framework tools as they are triggered by actions carried out by various user groups when working with the tools. User groups include researchers, research managers, and research funders (and possibly others). The arrows in the diagrams depict the actions users perform themselves. Importantly, this diagram does not depict the processes in the background (i.e. what information is exchanged between the tools), even as they are triggered by the user. For instance, the diagram shows that a researcher might wish to “search for digital objects and retrieve metadata” (DMP platform → SKG, mid-right) and/or use the tools to “Show similar research output” (repository → SKG, mid-top).

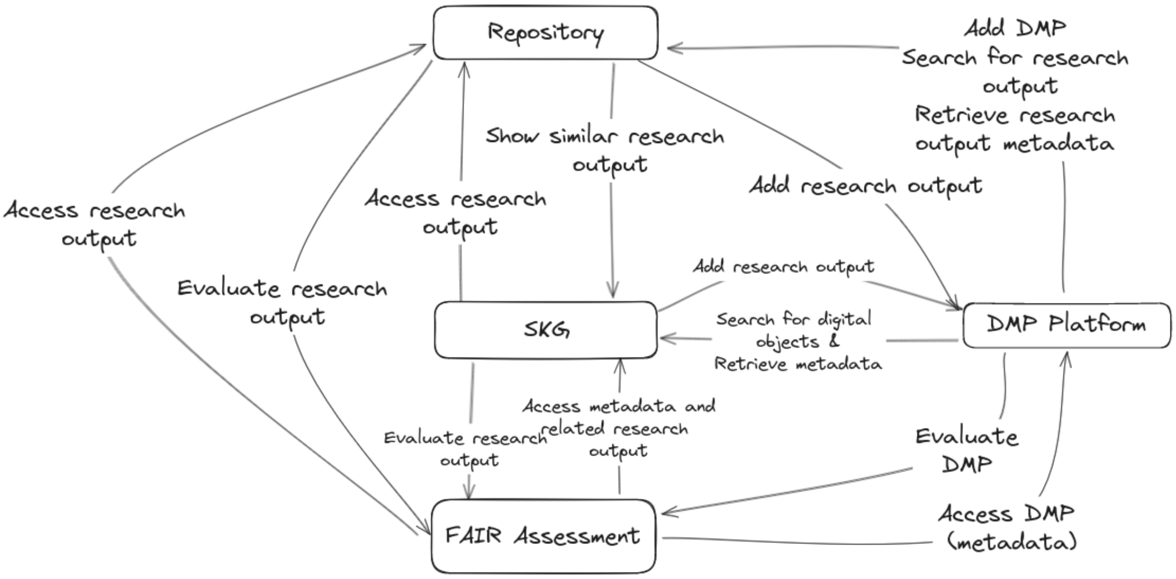


Figure 7: Generic pathways diagram, including all the feedback received

In what follows, this generic diagram (**Figure 7**) is discussed for each of the three components (DMPs, SKGs, FAIR Assessment tools): first, the “Plan Pathway” discusses potential user actions during the planning phase (starting from a DMP); this is followed by the “Track Pathway”, where the potential user actions during the tracking phase (starting from an SKG) are discussed; finally, the potential user actions during the assessment phase (“Assess Pathway”, starting from a DMP Evaluator) are discussed. We start by discussing user actions that may occur during the planning phase of a research project. This type of user action typically starts from a DMP. **Figure 8** presents a subsection of **Figure 7**, describing how a potential user might start from a DMP platform to interact with the other components (SKGs, repositories, DMP evaluators).

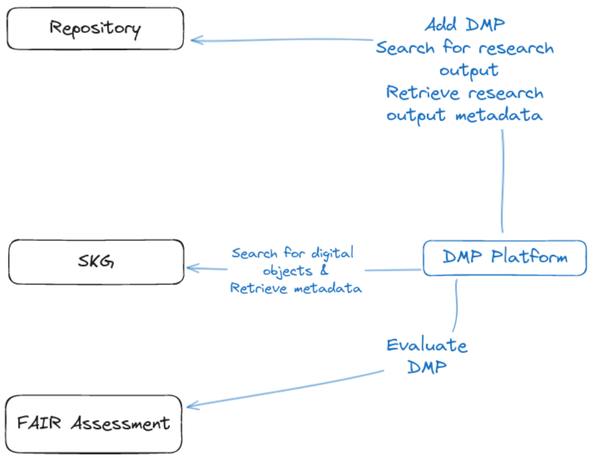


Figure 8: Pathway 1-PU

In **Pathway 1-PU** (**Figure 8**), user actions are highlighted in blue. The figure abstracts from all potential user actions starting from any of the other components (and likewise for all subsequent Figures). During the "planning" phase (planning data collection and further treatment of data), users create a Data Management Plan via a DMP platform. Creating a DMP can trigger the following interactions with the other components (note that not all steps need to take place for all DMPs):

1) Search for research output via Repository
2) Search for digital objects via an SKG
3) Retrieve metadata for research output via SKGs/Repositories
4) Add a DMP to a repository
5) Evaluate completeness and FAIRness of a DMP and/or research output listed in the DMP via FAIR Assessment

Next, we discuss user actions that may occur during the tracking phase of a research project. This type of user action typically starts from an SKG. **Pathway 2-TU** (**Figure 9**) describes how a potential user might start from an SKG to interact with the other components (DMP platforms, repositories, DMP evaluators) during the "tracking" phase.

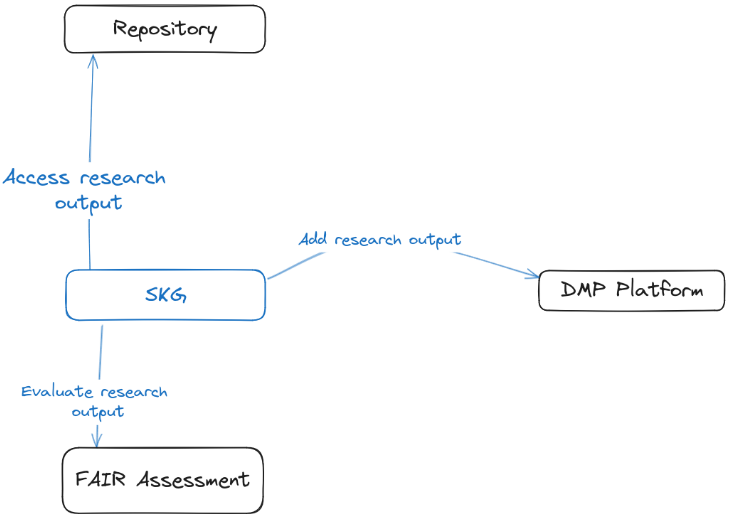


Figure 9: Pathway 2-TU

During active research, researchers retrieve research outputs and other Digital Objects via SKG. For instance, a researcher might wish to "add a research output" (SKG → DMP platform) and/or use the SKG to "Access research output" (SKG → repository). More specifically, the actions that users might take to achieve this are the following:

1) Access full research output via the respective repository
2) Evaluate FAIRness of research output metadata via FAIR Assessment
3) Add research output to a DMP via DMP platform

We continue with user actions that may occur during the "assess" phase of a research project. This type of user action typically starts from a FAIR assessment tool. **Pathway 3-AU** (**Figure 10**) represents the relevant subsection of **Figure 7**, describing how a potential user might start from a FAIR assessment tool to interact with the other components (DMPs, SKGs, repositories) during the "assess" phase.

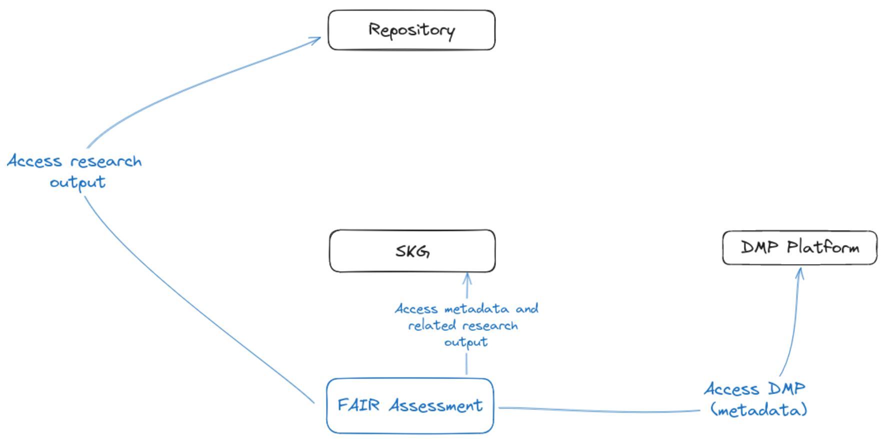


Figure 10: Pathway 3-AU

During active research, researchers may wish to access research outputs and/or DMPs via a FAIR assessment tool. For instance, a researcher might wish to "access research output" (FAIR assessment tool → repository) and/or use the FAIR assessment tool to

"Access metadata and related research output" (FAIR assessment tool → SKG). To evaluate the FAIRness of DMPs or research outputs, a user may:

1) Access full research output mentioned in a DMP via the respective repository
2) When research output or the Repository itself is not available, users may access metadata and related research output via SKGs

3.2. Interactions between tools

From the point of view of the PTA framework tools, the interactions between them may be triggered by a request from a user, or by a request placed by another tool; in this case, they may happen through actions in the background that are not triggered (or even noticed) by users. **Figure 11** shows these interactions between tools, with the direction of the arrows indicating which tool initiates the respective action and from which of the others information is requested. Due to the level of abstraction chosen, no technical details are shown here; they will be tackled elsewhere in the project.

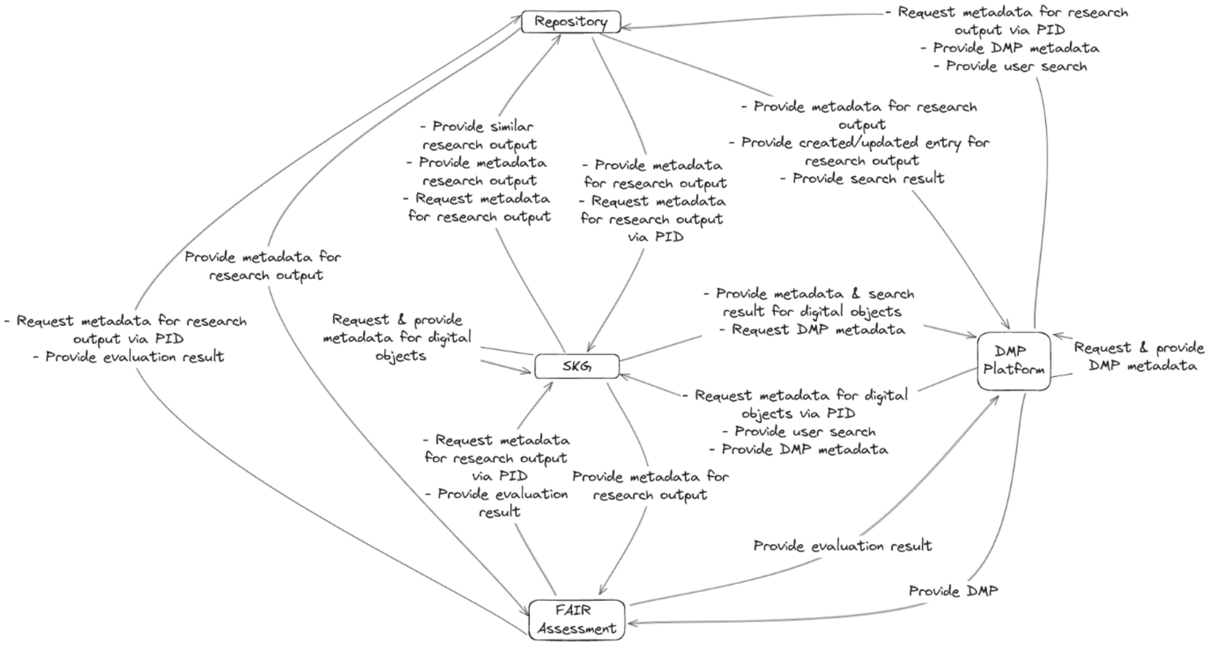


Figure 11: Generic Tool interaction diagram

We begin by discussing **Pathway 1-PT**, which represents potential tool interactions founded in DMPs during the "planning" phase, highlighting potential tool interactions in blue (**Figure 12**). The Figure abstracts from all potential interactions starting from any of the other components (and likewise for all subsequent Figures). This type of tool interaction typically starts from a DMP. **Figure 12** presents the relevant subsection of **Figure 11**, describing how a potential interaction with the other components (FAIR assessment tools, SKGs, repositories) might be triggered from a DMP, either by some other user action, or via a background task:

1) Request metadata for research output from Repositories
2) Request metadata for digital objects from SKGs
3) Provide user search to SKGs/Repositories
4) Provide DMP metadata to SKGs/Repositories/FAIR assessment/DMP platforms
5) Request DMP metadata from DMP platforms

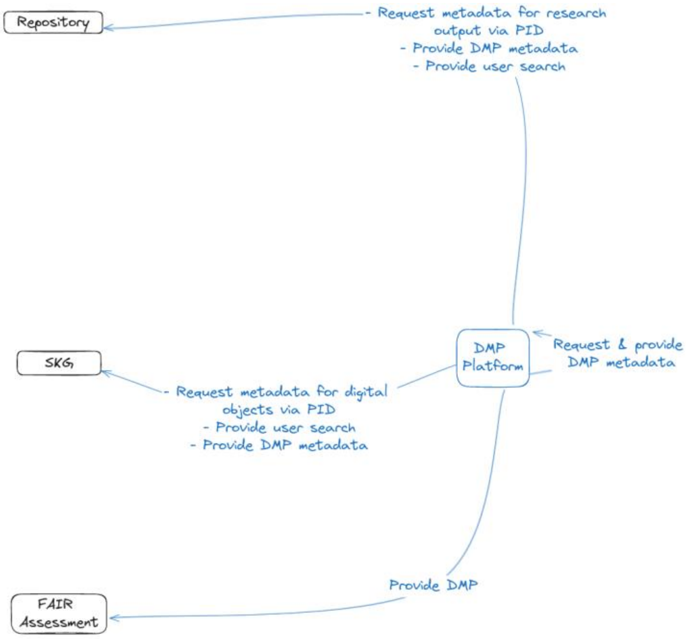


Figure 12: Pathway 1-PT

We continue with **Pathway 2-TT**, which represents tool interactions starting from SKGs (**Figure 13**) during the "tracking" phase; as above, component interactions are highlighted in blue to abstract from potential interactions starting from any of the other components. This type of tool interaction typically starts from an SKG. **Figure 13** represents how a potential interaction with the other components (FAIR assessment tools, DMPs, repositories) might be triggered from an SKG, either by some other user action, or via a background task:

1) Provide Repositories research output similar to those that were requested
2) Request metadata for research output from Repositories
3) Provide metadata for research output to Repositories/FAIR assessment tools
4) Provide metadata about Digital Objects to SKGs/DMP platforms
5) Request metadata about Digital Objects from SKGs
6) Request DMP metadata from DMP platforms

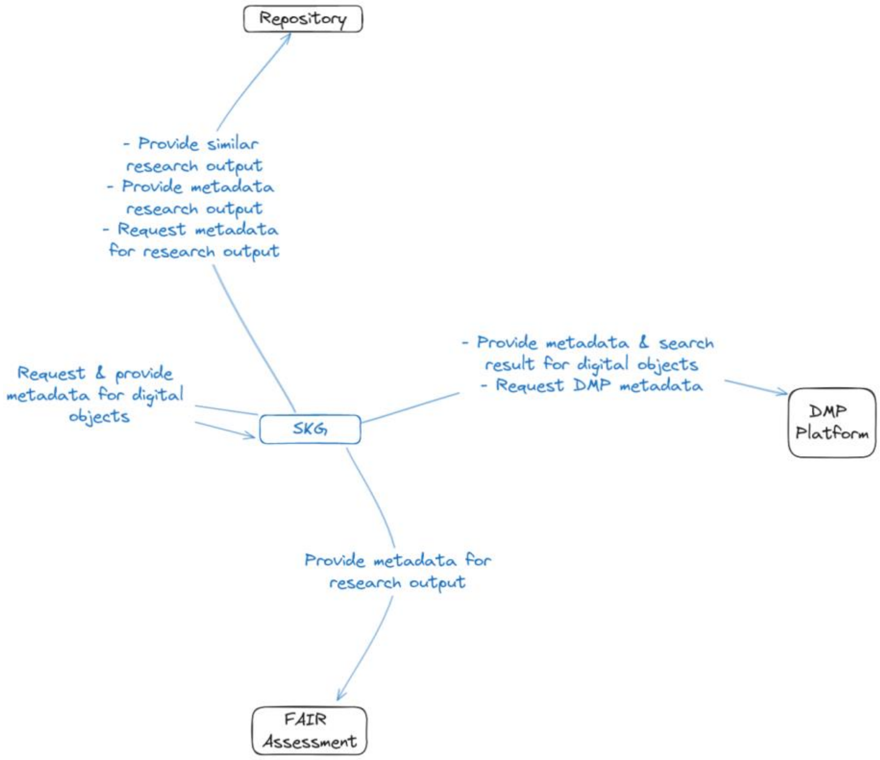


Figure 13: Pathway 2-TT

Lastly, **Figure 14**, represents **Pathway 3-AT** where tool interactions are described from the viewpoint of a FAIR assessment tool, with component interactions highlighted in blue to abstract from potential interactions starting from any of the other components. **Figure 14** represents a potential interaction with the other components (SKGs, DMPs, repositories), as triggered from a FAIR assessment tool, either by some other user action, or via a background task.

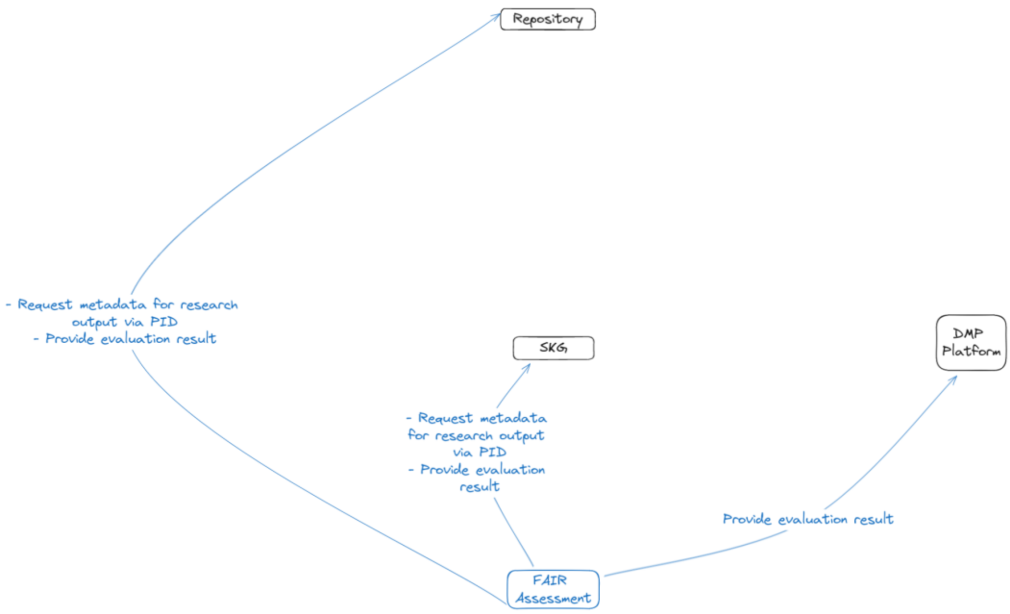


Figure 14: Pathway 3-AT

When triggered by users, or by a background task, a FAIR assessment tool may:

1) Request metadata for research output (via a PID) from an SKGs and/or a repository
2) Provide an evaluation result to SKGs and/or a repository and/or a DMP platform.

4. Conclusion and recommendations

The key objectives of this deliverable are to capture the needed pathways that represent visually how stakeholders use the Plan-Track-Assess (PTA) framework components, validate these pathways with partners and pilots to ensure they reflect current and ideal practices, and promote the interoperability of existing tools by identifying common practices across different scientific disciplines and geographic locations.

Starting from a high-level, tool-independent framework, we outlined how various stakeholders—researchers, research managers, and funders—can plan, track, and assess the management of research data in alignment with the FAIR principles. The framework comprises three main components: Data Management Planning platforms (DMPs), Scientific Knowledge Graphs (SKGs), and FAIR assessment tools. Building upon visual diagrams representing these processes, we outlined the development and validation of the pathways through consultations with project partners and pilot projects.[23] The process involved initial sketches by the authors of this deliverable, as well as user stories adopted from various data management workshops. The sketches were then validated through a structured consultation process with consortium partners and thematic pilots, and subsequently revised based on the feedback received. Going forward, feedback from the national pilots will need to be collected in a future iteration.

The landscape showed that tools and services supporting FAIR principles is fragmented, leading to difficulties in interoperability and suboptimal resource use. This fragmentation often results in duplicate solutions and confusion among users who are less advanced in FAIR practices. In addition, the feedback highlighted gaps between current practices and the idealized future state depicted in the pathways. Issues such as unclear terminology, ambiguous interactions, and the lack of temporal context in data transfers were identified as areas needing improvement.

Based on these, we identified the following three pathways and a list of recommendations going forward:

- The "Plan" Pathway outlines the process by which researchers develop Data Management Plans (DMPs), retrieve metadata, and assess these DMPs to enhance their adherence to FAIR principles.
  - Enhance DMP Platforms Integration: Improve the integration of Data Management Plan (DMP) platforms with Scientific Knowledge Graphs (SKGs) and repositories. This includes enabling seamless searches for

[23] https://ostrails.eu/case-studies

research outputs and digital objects via SKGs and retrieving metadata efficiently.

  - Metadata Retrieval and Evaluation: Develop capabilities for DMP platforms to not only create and store DMPs but also to evaluate the completeness and FAIRness of DMPs and associated research outputs. This involves implementing tools that can assess the FAIR (Findable, Accessible, Interoperable, and Reusable) principles of research data.
  - User-Friendly Interfaces: Create intuitive user interfaces that guide researchers through the process of creating, managing, and evaluating DMPs. This should facilitate easy addition of DMPs to repositories and ensure that metadata retrieval is straightforward and user-friendly.

- The "Track" Pathway demonstrates how Semantic Knowledge Graphs (SKGs) are utilized to access Digital Objects (DOs), evaluate the FAIRness of metadata, and incorporate records into DMPs.
  - Enhanced Tracking Capabilities: Improve the functionalities of SKGs to allow comprehensive tracking of research outputs. This includes providing users with the ability to access full research outputs via respective repositories and ensuring the continuous evaluation of metadata FAIRness.
  - Integration with DMP Platforms: Enable seamless integration between SKGs and DMP platforms to allow users to add research outputs directly into DMPs from SKGs. This integration should also support the continuous updating and tracking of research outputs within the DMP.
  - Develop and integrate FAIR assessment tools that can be accessed through SKGs to evaluate the FAIRness of research outputs. This should include providing detailed feedback to users on how to improve their research outputs' compliance with FAIR principles.
- The "Assess" Pathway depicts the evaluation process of DMPs and DOs.
  - Comprehensive Assessment Tools: Enhance FAIR assessment tools to evaluate DMPs and research outputs comprehensively. These tools should be able to access full records of research outputs mentioned in DMPs and provide detailed assessments of their FAIRness.
  - Interoperability with SKGs and Repositories: Ensure that FAIR assessment tools are interoperable with SKGs and repositories, enabling them to retrieve metadata and full research outputs efficiently. This interoperability will support more accurate and comprehensive assessments.
  - Automated and Manual Assessment Options: Develop both automated and manual assessment options within the FAIR assessment tools to cater

to different user needs. Automated assessments can provide quick feedback, while manual options allow for more detailed and customized evaluations.

These findings and recommendations set the stage for further harmonization of PTA tools, aiming to enhance the interoperability and adoption of FAIR practices in research data management. The updated diagrams offer a robust foundation for the PTA framework informing the OSTrails Interoperability Reference Architecture to better support the management of Digital Objects in alignment with FAIR principlesserving as a valuable resource for Open Science and EOSC ecosystems.

5. References

5.1. Peer-reviewed publications

Amorim, Ricardo Carvalho, João Aguiar Castro, João Rocha da Silva, and Cristina Ribeiro. 2015. 'A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential'. In New Contributions in Information Systems and Technologies, edited by Alvaro Rocha, Ana Maria Correia, Sandra Costanzo, and Luis Paulo Reis, 101–11. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-16486-1_10.

Atenas, Javiera, Leo Havemann, and Cristian Timmermann. 2020. 'Critical Literacies for a Datafied Society: Academic Development and Curriculum Design in Higher Education'. Research in Learning Technology 28:2468. https://doi.org/10.25304/rlt.v28.2468.

Awre, Chris, Jim Baxter, Brian Clifford, and Janette Colclough. 2015. 'Research Data Management as a "Wicked Problem"'. Library Review 64 (4/5): 356–71. https://doi.org/10.1108/LR-04-2015-0043.

Berg-Cross, Gary, Raphael Ritz, and Peter Wittenburg. 2015. 'Data Foundation and Terminology Work Group Products', August. https://doi.org/10.15497/06825049-8CA4-40BD-BCAF-DE9F0EA2FADF.

Bezuidenhout, Louise M., Sabina Leonelli, Ann H. Kelly, and Brian Rappert. 2017. 'Beyond the Digital Divide: Towards a Situated Approach to Open Data'. Science and Public Policy 44 (4): 464–75. https://doi.org/10.1093/scipol/scw036.

Bishop, Bradley Wade, Peter Neish, Ji Hyun Kim, Raphaëlle Bats, A. J. Million, Jake Carlson, Heather Moulaison-Sandy, and Minh T. Pham. 2023. 'Data Management Plan Implementation, Assessments, and Evaluations: Implications and Recommendations'. Data Science Journal 22 (1). https://doi.org/10.5334/dsj-2023-027.

Boeckhout, Martin, Gerhard A. Zielhuis, and Annelien L. Bredenoord. 2018. 'The FAIR Guiding Principles for Data Stewardship: Fair Enough?' European Journal of Human Genetics 26 (7): 931–36. https://doi.org/10.1038/s41431-018-0160-0.

Borgman, Christine L. 2012. 'The Conundrum of Sharing Research Data'. Journal of the American Society for Information Science and Technology 63 (6): 1059–78. https://doi.org/10.1002/asi.22634.

Bugaje, Maryam, and Gobinda Chowdhury. 2018. 'Identifying Design Requirements of a User-Centered Research Data Management System'. In Lecture Notes in Computer Science, Information Systems and Applications, Incl. Internet/Web, and HCI, 11279:335–47. https://doi.org/10.1007/978-3-030-04257-8_35.

Cychosz, Margaret, Rachel Romeo, Melanie Soderstrom, Camila Scaff, Hillary Ganek, Alejandrina Cristia, Marisa Casillas, Kaya de Barbaro, Janet Y. Bang, and Adriana Weisleder. 2020. 'Longform Recordings of Everyday Life: Ethics for Best Practices'.

Behavior Research Methods 52 (5): 1951–69. https://doi.org/10.3758/s13428-020-01365-9.

David, Romain, Laurence Mabile, Alison Specht, Sarah Stryeck, Mogens Thomsen, Mohamed Yahia, Clement Jonquet, et al. 2020. 'FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles'. Data Science Journal 19 (1). https://doi.org/10.5334/dsj-2020-032.

De Smedt, Koenraad, Dimitris Koureas, and Peter Wittenburg. 2020. 'FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units'. Publications (Basel) 8 (2): 21-. https://doi.org/10.3390/publications8020021.

Fecher, Benedikt, Sascha Friesike, and Marcel Hebing. 2015. 'What Drives Academic Data Sharing?' PLOS ONE 10 (2): e0118053. https://doi.org/10.1371/journal.pone.0118053.

Fuhr, Justin. 2019. '"How Do I Do That?" A Literature Review of Research Data Management Skill Gaps of Canadian Health Sciences Information Professionals'. Journal of the Canadian Health Libraries Association 40 (2). https://doi.org/10.29173/jchla29371.

Gilmore, Rick O., Michele T. Diaz, Brad A. Wyble, and Tal Yarkoni. 2017. 'Progress toward Openness, Transparency, and Reproducibility in Cognitive Neuroscience'. Annals of the New York Academy of Sciences 1396 (1): 5–18. https://doi.org/10.1111/nyas.13325.

Higman, Rosie, and Stephen Pinfield. 2015. 'Research Data Management and Openness'. Program: Electronic Library and Information Systems 49 (4): 364–81. https://doi.org/10.1108/PROG-01-2015-0005.

Jacobsen, Annika, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, et al. 2020. 'FAIR Principles: Interpretations and Implementation Considerations'. Data Intelligence, January. https://doi.org/10.1162/dint_r_00024.

Jones, Sarah, Robert Pergl, Rob Hooft, Tomasz Miksa, Robert Samors, Judit Ungvari, Rowena I. Davis, and Tina Lee. 2020. 'Data Management Planning: How Requirements and Solutions Are Beginning to Converge'. Data Intelligence 2 (1–2): 208–19. https://doi.org/10.1162/dint_a_00043.

Kalichman, Michael, Monica Sweet, and Dena Plemmons. 2014. 'Standards of Scientific Conduct: Are There Any?' Science and Engineering Ethics 20 (4): 885–96. https://doi.org/10.1007/s11948-013-9500-1.

———. 2015. 'Standards of Scientific Conduct: Disciplinary Differences'. Science and Engineering Ethics 21 (5): 1085–93. https://doi.org/10.1007/s11948-014-9594-0.

Kurata, Keiko, Mamiko Matsubayashi, and Shinji Mine. 2017. 'Identifying the Complex Position of Research Data and Data Sharing Among Researchers in Natural Science'. Sage Open 7 (3): 2158244017717301. https://doi.org/10.1177/2158244017717301.

Leonelli, Sabina. 2017. 'Global Data Quality Assessment and the Situated Nature of "Best" Research Practices in Biology'. Data Science Journal 16 (0): 32. https://doi.org/10.5334/dsj-2017-032.

Levin, Nadine, and Sabina Leonelli. 2017. 'How Does One "Open" Science? Questions of Value in Biological Research'. Science, Technology, & Human Values 42 (2): 280–305. https://doi.org/10.1177/0162243916672071.

Levin, Nadine, Sabina Leonelli, Dagmara Weckowska, David Castle, and John Dupré. 2016. 'How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice'. Bulletin of Science, Technology & Society 36 (2): 128–41. https://doi.org/10.1177/0270467616668760.

Miksa, Tomasz, Stephanie Simms, Daniel Mietchen, and Sarah Jones. 2019. 'Ten Principles for Machine-Actionable Data Management Plans'. PLOS Computational Biology 15 (3): e1006750. https://doi.org/10.1371/journal.pcbi.1006750.

Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, and Mark D. Wilkinson. 2017. 'Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud'. Information Services & Use 37 (1): 49–56. https://doi.org/10.3233/ISU-170824.

Papadopoulou, Elli, Alessia Bardi, Georgios Kakaletris, Diamantis Tziotzios, Paolo Manghi, and Natalia Manola. 2020. 'Data Management Plans and Linked Open Data: Exploiting Machine Actionable Data Management Plans through Open Science Graphs'. Application/pdf. https://doi.org/10.4126/FRL01-006423283.

Purnama Jati, Putu Hadi, Yi Lin, Sara Nodehi, Dwy Bagus Cahyono, and Mirjam van Reisen. 2022. 'FAIR Versus Open Data: A Comparison of Objectives and Principles'. Data Intelligence 4 (4): 867–81. https://doi.org/10.1162/dint_a_00176.

Rappert, Brian, and Louise Bezuidenhout. 2016. 'Data Sharing in Low-Resourced Research Environments'. Prometheus 34 (3–4): 207–24. https://doi.org/10.1080/08109028.2017.1325142.

Ross-Hellauer, Tony, Stefan Reichmann, Nicki Lisa Cole, Angela Fessl, Thomas Klebel, and Nancy Pontika. n.d. 'Dynamics of Cumulative Advantage and Threats to Equity in Open Science: A Scoping Review'. Royal Society Open Science 9 (1): 211032. https://doi.org/10.1098/rsos.211032.

Schwardmann, Ulrich. 2020. 'Digital Objects – FAIR Digital Objects: Which Services Are Required?' Data Science Journal 19 (1). https://doi.org/10.5334/dsj-2020-015.

Sisu, Diana, Tomasz Miksa, Marie-Christine Jacquemot, Elli Papadopoulou, Maria Praetzellis, and Marek Suchanek. 2024. 'Elevating Data Management Planning: Interoperability of RDM Services through Machine-Actionability', March. https://zenodo.org/records/10794257.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. 'Data Sharing by Scientists: Practices and Perceptions'. PLOS ONE 6 (6): e21101. https://doi.org/10.1371/journal.pone.0021101.

Unal, Y., G. Chowdhury, S. Kurbanoglu, and J. Boustany. 2019. 'Research Data Management and Data Sharing Behaviour of University Researchers'. Information Research-An International Electronic Journal 24 (1).

Weinshall, Keren, and Lee Epstein. 2020. 'Developing High-Quality Data Infrastructure for Legal Analytics: Introducing the Israeli Supreme Court Database'. Journal of Empirical Legal Studies 17 (2): 416–34. https://doi.org/10.1111/jels.12250.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. Scientific Data 3 (March):160018.

Wilkinson, Mark D., Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, et al. 2019. 'Evaluating FAIR Maturity through a Scalable, Automated, Community-Governed Framework'. Scientific Data 6 (1): 174. https://doi.org/10.1038/s41597-019-0184-5.

Wilkinson, Mark D., Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos, and Michel Dumontier. 2018. 'A Design Framework and Exemplar Metrics for FAIRness'. Scientific Data 5 (June):180118. https://doi.org/10.1038/sdata.2018.118.

5.2. Other sources

Machine-actionable Data Management Plans, presentation by Tomasz Miksa, (2020): https://snd.se/sites/default/files/2020-06/Miksa%20-%20Machine-actionable%20Data%20Management%20Plans.pdf and ref and references therein.

V. K. Chaudhri *et al., An introduction to knowledge graphs*, https://ai.stanford.edu/blog/introduction-to-knowledge-graphs/.

5.3. Output from RDA Working Groups

RDA SKG Working Group: https://ostrails.eu/skgs

RDA Common Standard for maDMPs: https://www.sba-research.org/wp-content/uploads/2020/09/MLDM_RDA-DMP-Common-Standard-for-Machine-actionable-Data-Management-Plans_WEB.pdf

RDA Common Standards WG: https://www.rd-alliance.org/groups/dmp-common-standards-wg/members/all-members/

https://github.com/RDA-DMP-Common

RDA Github: https://github.com/RDA-DMP-Common

5.4. Deliverables from relevant INFRAEOSC projects

SciLake: Blog Post on SKGs: https://scilake.eu/survey-results-insights-into-the-potential-of-scientific-knowledge-graphs

Annex I: Feedback from consultation

Table 2: Summary of changes suggested to the pathway diagrams

ARROW	SUGGESTION	REASON
"Add Record to DMP"	Rename to "Query repository for record metadata"	This would better represent when DMP tools requests information about DOs from a repository
N.A.	Introduce new arrow DMP▸Repository with label "Add DMP into the repository"	A new arrow would allow to state where a DMP is stored
"Access full record" arrow between repository and SKG	The arrow should be bi-directional	
"Access full record/DO"	The arrow should (also) point in the opposite direction	The process opposite to that shown takes place when the DMP evaluator which requests metadata about repositories, standards, and policies described in the DMP
"Evaluate record/DO"	Rename to "Access record/DO for evaluation"	The current meaning is considered unclear since it does not indicate who placed the request
General comment on arrows	Labels should also state which DOs are accessed, i.e. whether the datasets mentioned in the DMP, or the DMP itself. If the latter, then there is no direct step where the DMP is added to the repository.	To express the steps via links between components,

5.4.1. Feedback on the DMP Pathway

Table 3: Overview of the feedback received on the DMP pathway diagram

QUESTION	ANSWER	ASSESSMENT
Does the diagram provide a good representation of the steps taken when using DMPs?	The sketch is incomplete, as more than one SKG might be involved in preparing a DMP	Sketches are generic representations, the SKG box is intended to represent all SKGs
	Difficulty to assess sketch since it may not reflect all interactions or steps	

	Sketch misrepresents current situation, as it implies some steps are machine actionable	Diagrams depict the desired state of things
How well does the sketch represent current situation, workflows, or practices?	The connection between DMP evaluator and DMP tool should be bi-directional, as the evaluator reports back to the tool	
	For SSHOMP, researchers may have to complement information in the DMP through SSHOMP as a second SKG, which would mean users would have to do additional work	
How well does the sketch represent an ideal situation? Where do you see room for improvement?	– What happens if there are different versions of a DO? Should metadata be merged, or should metadata in one repository be preferred? – Is it necessary to add the whole metadata record to the DMP? The bibliographic citation and a persistent URL to the data source may be enough. – Researchers would like to be able to search and retrieve records from repositories, and to retrieve metadata providing a URL.	
Is there anything missing from the sketch? Are there other ways for DMPs to interact with other components?	– DMP tools should have more ways to combine information from different repositories – References to controlled vocabularies that can be used for DMP missing – SKGs usually store only metadata; also, DMP tools are not suitable for long term preservation, since their purpose is (just) to enable the creation of DMPs.	
	DMPs need to include the whole metadata record, instead of only bibliographic citation and (persistent) URL to the data source.	May be solved by indicating that the full record would be retrieved if available; if it is not, DMPs have to make do with the available information
Does this diagram diverge from the sketches you/your team has provided?	– Clearer representation of information flows (i.e. directions and initiators). – Researchers expect DMP tools to enquire FAIR evaluators about an entry in a repository by e.g. using its DOI, so that the user could	

	employ the “live” feedback while editing the DMP – DMP evaluation (which includes the evaluation of FAIRness) should be part of the DMP tool – use of other persistent identifiers to search for other types of entities (e.g. ORCID for persons, ROR for organisations, URN for research literature/data, etc.), and to clarify what a ‘record’ is	

5.4.2. Feedback on the SKG pathway

Table 4: Overview of the feedback received on the SKG pathway diagram

QUESTION	ANSWER	ASSESSMENT
Does the diagram provide a good representation of the steps taken when using SKGs?	the “thin line” separating repositories from SKGs, as recognised by e.g. CESSDA and PaNOSC, makes it sometimes difficult to interpret and assess the diagrams	
	information about a record in an SKG might come from more than one repository (due possibly to duplicated entries), or include data added by users (which may be less structured than those from a repository)	Implicit in the diagram
	The diagram, in which “Repository” could be understood as included several of them, and even other cases where input into an SKG comes from other sources that do not qualify as repositories	
How well does the sketch represent the current situation, workflows, or practices?	SKGs already have a full record of a DO by harvesting metadata from various sources, but they can only provide information on their current cache, since the searches are not federated.	The request that the diagram shows that any PID or URI should be valid to retrieve information from the SKG is difficult to show. This is already implicit since no identifier was specified.
How well does the sketch represent an ideal situation? Where do you see	Diagram should how the relation between DMP (FAIR) evaluators that query SKGs to get metadata from as part of the evaluation	This needs further assessment to clarify whether evaluators retrieve information about records from

room for improvement?		repositories only, or if they also use SKGs
	The diagram should show whether data have been processed (curated, edited, etc.) before being passed on from one component to the next.	This could be addressed by creating different diagrams depending on the status of the DO
Does this diagram diverge from the sketches you/your team has provided?	use of SKGs by ESRF seems not to be included in the diagram	

5.4.3. Feedback for FAIR & DMP Evaluators

Table 5: Overview of the feedback received on the FAIR assessment tools pathway diagram

QUESTION	ANSWER	ASSESSMENT
Does the diagram provide a good representation of the steps taken when using FAIR tools?	lack of clarity about what is it that the DMP evaluators actually assess (is it FAIR assessment of the digital objects, or rather compliance with policy requirements?) and their interaction with FAIR evaluators (do DMP evaluators request information on FAIRness of a DMP from a FAIR evaluator, or do they assess it directly?).	
	representation of the search functionality of DMP evaluators was deemed insufficient, as it leaves unclear how a user would find a record/DO they need to retrieve.	
How well does the sketch represent the current situation, workflows, or practices?	metadata evaluation, which may happen directly (i.e. not via the SKG), so the DMP evaluator could access the same information. To express the steps via links between components, labels should also state what DOs are accessed, whether the datasets mentioned in the DMP, or the DMP itself. If the latter, then there is no direct step where the DMP is added to the repository.	
	how the evaluator finds a DO in a repository should be explicitly depicted	

How well does the sketch represent an ideal situation? Where do you see room for improvement?	the diagram should do a better job at representing ambiguity in DOs (e.g. outdated or duplicate entries	
	The consortium should also consider interactions between DMP evaluator and SKGs, to accommodate cases where the DO/repository is not available.	
Does this diagram diverge from the sketches you/your team has provided?	the diagram misses a connection between the SKG and the DMP evaluator	

Annex 2: Consultation preparations

This document was prepared to collect feedback from OSTrails partners regarding the definition of Plan-Track-Assess pathways documenting current requirements and pinpointing future needs.

Background and definitions to date

Conceptual Components: Understand how tools and services interact, what data they use, and what they provide

Generic Conceptual Elements: Abstract components to create universal representations not tied to specific tools

Pathways: Offer a process / use case view, showing how components work together to deliver value. They prioritize integrations, revealing common use cases. For example, pathways illustrate how DMP tools are paired with other components to automate feedback on research quality.

Conceptual Architecture: Combines pathways and generic conceptual elements to form a high-level framework. It doesn't specify tools but guides implementation in pilot projects.

Pathways - working definition for the consultation process

- Actions a user can perform, starting from a certain tool
- It may not depict all integrations and interfaces which are required to make it work, and only shows what a user should be able to do in an abstract way

Requirements

Defining joint pathways for Data Management Plans (DMPs), Scientific Knowledge Graphs (SKGs), and FAIR Assessment tools (FATs) requires us to

- show how these types of tools are interconnected
- understand how different user types use these tools (e.g. by way of devising persona scenarios/user stories)
- document each of these scenarios to understand what exists already, what is needed, and what is currently missing (and could therefore define areas for improvement/development)

Approach and desired outcomes

TU Graz has prepared preliminary sketches of OSTrails Pathways that were presented at the dedicated meeting on 25 April 2024. With this consultation process, TU Graz expects to get additional input from project partners on the Pathways based on their context, conceptual components and sketches. The input will be used to identify gaps in the current landscape for DMPs, SKGs, and FAIR Assessment tools, assess their validity, and update the sketches.

The document is structured as follows: After a brief set of instructions on the consultation process, readers are presented with three Pathways, each with their separate diagram, followed by a set of open questions to consider when assessing the sketches. We have prepared this for all three sketches and kindly ask you to participate and provide feedback and comments directly in this document.

Instructions

Below you will find **sketches of possible pathways** to connect DMPs, SKGs, and FAIR Assessment tools. These sketches are the objects of this consultation process. Please assess them against your own conceptual components and sketches, with a view to the following **dimensions** and **questions**:

Dimensions of the assessment: representing the current situation/research process (given a role, a discipline, a maturity level, etc.); representing desirable improvements to the process; bearing in mind what different user groups want to achieve using a specific tool.

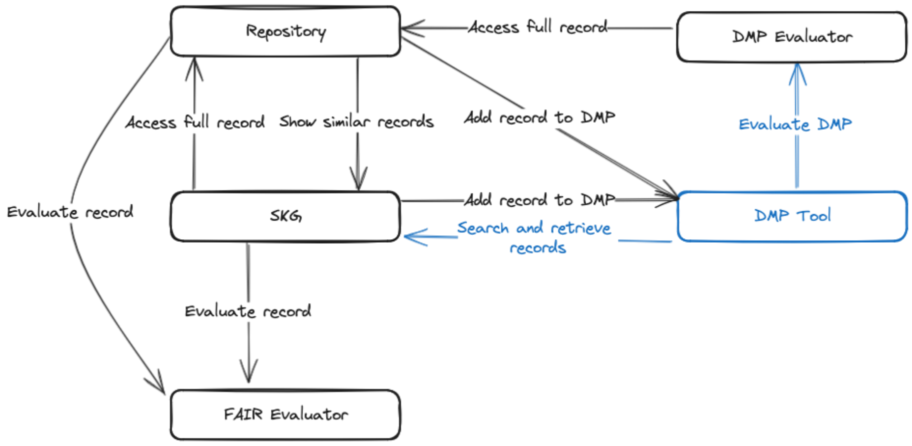
Questions:

- From what position are you assessing the sketch (i.e. researcher, research manager, research evaluator, etc.)? Please indicate this in your responses.
- How well do the sketches represent the current situation/workflows/practices? Please assess the sketches based on the position(s) you are assuming. *Example: How well does the pathways sketch represent the current use of DMPs? As a [researcher, research evaluator, etc.], I use DMP tools to do…*
- How well do the sketches represent an ideal situation? Where do you see room for improvement based on your own experience? Example: Where would the sketch need to be improved to optimize DMP usage? As a [researcher, research evaluator, etc.], I would like DMP tools to be able to do…
- Is there anything missing from these sketches based on the above considerations?

To submit your assessment, please work on this document, and store a copy with the following file name: **OSTrails T1.1_OS Pathways Consultation_FirstNameLastName** in the Consultation Process folder in the OSTrails MS Teams space (this link). Alternatively, you can send it to Miguel Rey Mazón (m.reymazon@tugraz.at) and Stefan Reichmann (stefan.reichmann@tugraz.at). We kindly ask you to complete this by **Tuesday 7 May 2024** at the latest. We will collect and collate the feedback received. Please also indicate the name of your tool/service/institution and/or (national/thematic) pilot. You can access the sketches / diagrams here:
https://excalidraw.com/#room=a87117f81e154692b999,02piqBdHvm9D8zsVWJvukw

Consultation Process

Pathway 1: Plan (Research Planning, starting from DMPs)



During creation of a DMP, a user may:

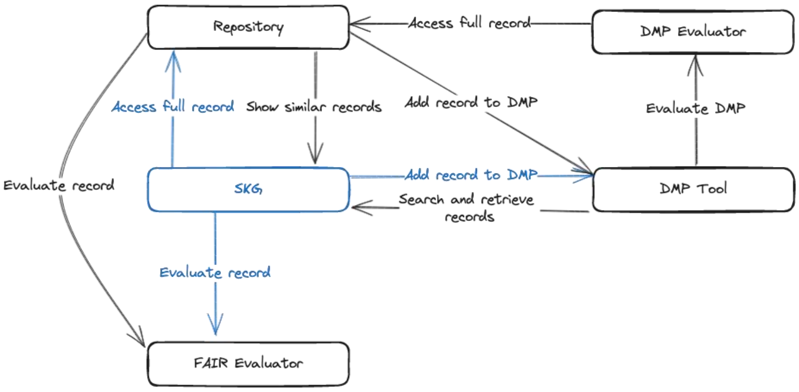
- Search for records via SKGs
- Retrieve record metadata via SKGs by providing DOI
- Evaluate completeness and FAIRness of DMP via DMP evaluator

Table 6: Table with the same content as for DMPs but changing DMP to FAIR tools.

DIMENSION/TOPIC	ROLE	QUESTION	YOUR RESPONSE
Current situation/workflows /practices	Researcher Research Manager Research Evaluator …	How well does this diagram represent the current steps in using DMPs? Is this how DMPs are being used?	
Current situation/workflows /practices	Researcher Research Manager Research Evaluator …	How well does the diagram represent the interactions between the DMP tool and the other components?	
Current situation/workflows /practices	Researcher Research Manager Research Evaluator …	How well does the diagram represent the actions a prospective user might take (researcher, research evaluator, research manager, etc.)?	
Room for improvement	Researcher Research Manager Research Evaluator …	How well does the diagram represent an ideal situation? Where do you see room for improvement based on your own experience? Example: Where would the diagram need to be improved to optimize DMP usage? As a [researcher, research evaluator, etc.], I would like DMP tools to be able to do…	

Room for improvement	Researcher Research Manager Research Evaluator …	Is there any information missing from the above diagram? Are there other relevant ways in which DMPs could/should interact with the other components?	
Room for improvement	Researcher Research Manager Research Evaluator …	Does this diagram diverge from the sketches you/your team has provided? If yes, where does it diverge?	
Additional comments		Is there anything else you would like to add?	

Pathway 2: Discover/Track (starting from SKGs)

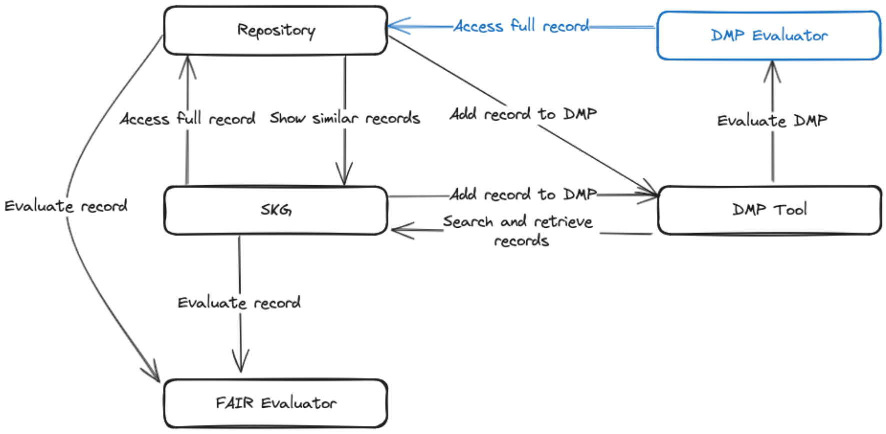


During active research and usage of an SKG a user may:

- Access full record via the respective repository
- Evaluate FAIRness of record metadata via FAIR evaluator
- Add record to DMP via DMP tool

  Table with the same content as for DMPs but changing DMP to SKG.

Pathway 3: Assess (starting from FAIR Assessment Tools)



E.g., during evaluation of a DMP, a user may:

- Access full records mentioned in DMP via respective repository