

graspOS

open research assessment dataspace

DOI: 10.5281/zenodo.13618531

Deliverable D4.2: Infrastructure Architecture

Due Date of Deliverable	31 August 2024 (M20)
Actual Submission Date	30 August 2024
Work Package	WP4
Type	R - Report
Approval Status	Submitted
Version	v1.0
Number of Pages	25

The information in this document reflects only the author's views and the European Commission is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/her sole risk and liability.

Abstract

This report describes the high-level architecture of the GraspOS federated infrastructure, which aims to support responsible research assessment processes that take into consideration Open Science practices.



This project has received funding from the European Union's Horizon Europe framework programme under grant agreement No. 101095129. 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.

Revision history

VERSION	DATE	REASON	REVISED BY
0.0	28/06/2023	First draft created - structure agreed	Giulia Malaguarnera, Andrea Mannocci
0.1	31/7/2024	Version ready for internal review	Thanasis Vergoulis, Giulia Malaguarnera, Andrea Mannocci
0.2	28/8/2024	Internal review comments incorporated	Thanasis Vergoulis, Giulia Malaguarnera, Andrea Mannocci
1.0	30/8/2024	Final Version after proofreading	Thanasis Vergoulis, Giulia Malaguarnera, Andrea Mannocci

Author List

ORGANISATION	NAME	CONTACT INFORMATION
OpenAIRE	Giulia Malaguarnera	giulia.malaguarnera@openaire.eu
CNR-ISTI	Andrea Mannocci	andrea.mannocci@isti.cnr.it
ARC	Thanasis Vergoulis	vergoulis@athenarc.gr
ARC	Serafeim Chatzopoulos	schatz@athenarc.gr
OPERAS	Suzanne Dumouchel	suzanne.dumouchel@operas-eu.org
CWTS	Clifford Tatum	c.c.tatum@cwts.leidenuniv.nl
UniBo	Silvio Peroni	silvio.peroni@unibo.it

Reviewer List

ORGANISATION	NAME	CONTACT INFORMATION
TSV	Dragan Ivanovic	dragan.ivanovic@uns.ac.rs
OPERAS	Fotis Mystakopoulos	fotis.mystakopoulos@operas-eu.org
TUBITAK	Selcuk Bilmis	selcuk.bilmis@tubitak.gov.tr

Table of Contents

1. Executive Summary	5
2. Introduction	6
3. Design	8
3.1. Methodology	8
3.2. Functional goals	9
3.3. Alignment with relevant initiatives	11
4. Architectural Overview	14
5. Federated Resources	17
5.1. Types of Federated Resources	17
5.2. Rules of Resource Inclusion	18
6. Core Components	19
6.1. Catalogues & Registries	19
6.2. Interoperability & Access Layer (DIAL)	20
6.3. Infrastructure Front-end UI	24
7. Conclusions	25
8. References	25

List of Figures

- **Figure 1:** Overview of the GraspOS infrastructure core components (p. 15)

Abbreviation List

- API - Application Programming Interface
- CoARA - Coalition for Advancing Research Assessment
- DIAL - Data Interoperability & Access Layer
- EDI - Equity, Diversity, and Inclusion
- EOSC - European Open Science Cloud
- EU - European Union
- HE - Horizon Europe
- OI4RRA - Open Infrastructures for Responsible Research Assessment
- OS - Open Science
- OSAF - Open Science Assessment Framework
- QoS - Quality of Service
- RAiD - Research Activity Identifier
- RDA - Research Data Alliance
- RFO - Research Funding Organisations
- RPO - Research Performing Organisations
- RRA - Responsible Research Assessment
- SKG - Scientific Knowledge Graph
- SKG-IF - Scientific Knowledge Graph Interoperability Framework
- UI - User Interface
- WG - Working Group

1. Executive Summary

Research assessment is crucial in a variety of research applications, from informing hiring and promotion decisions of research staff to influencing strategic investments and policymaking. However, widely used assessment practices face significant challenges, such as the use and abuse of a small pool of often misunderstood and non-transparent quantitative indicators, the explicit focus on publications overlooking other types of research activities and contributions, the fact that the effort given on practising Open Science is not properly acknowledged, and the reliance on evidence from proprietary data sources with restricted access, to name a few.

To address such challenges and pave the way for an *Open-Science-aware Responsible Research Assessment*, this deliverable outlines the architecture of the GraspOS (federated) infrastructure, which aims at enabling the creation of a Research Assessment Dataspace that is transparent, inclusive, and interoperable. The GraspOS infrastructure enables the integration of diverse (meta)data sources and facilitates their interoperability and the discoverability of their contents. This will allow (among others) for improved coverage of research products and more comprehensive assessments that are based on a wider range of evidence.

The GraspOS infrastructure is designed to be modular and scalable, allowing for the integration of new resources and data sources as needed. It is also designed to be open and accessible, enabling the research community to contribute to its development and use. The infrastructure is built on a federated architecture, which means that it is distributed across multiple locations. This enables the infrastructure to be more resilient and adaptable to changing needs and contexts. Of course, a series of centralised, core components exist that help the discovery of the included resources.

The current report introduces the readers to the main objectives of the GraspOS infrastructure, outlines the design process followed for its architecture, and provides details on its contents and core components.

2. Introduction

Research assessment refers to the process of evaluating research activities, outputs, practices, and roles in terms of various characteristics such as their quality, novelty, scientific/financial/societal impact, usage, and accessibility. This includes the process of supporting researchers in their career progression, evaluating research performed by Research Performing Organisations (RPOs) and measuring the scientific, societal, and economic impact of research projects/grants for Research Funding Organisations (RFOs). Consequently, research assessment plays a crucial role, ranging from informing the hiring and promotion processes in RPOs to shaping strategic investments and policymaking.

Commonly used research assessment practices face major challenges. In particular, there is a growing recognition that research and researchers need to be assessed in a more responsible way (de Rijcke et al., 2023). This requires a shift away from over-relying on a small pool of quantitative indicators, which are usually not calculated transparently and often misunderstood, towards a more comprehensive approach that considers different aspects and merits of research work and incorporates qualitative evidence. In addition, since research assessment practices shape research practices, they have to be well-thought-out to avoid amplifying problems such as the “Publish-or-Perish” culture and the Matthew effect (Merton, 1968), or fostering bad research practices (Strinzel et al., 2021). In response to these challenges, various initiatives for Responsible Research Assessment (RRA) have been launched (Moher et al. 2020, Curry et al., 2020).^{1,2,3} Furthermore, research assessment processes should be tailored to the scope and the context of the evaluation following an appropriate approach, such as the SCOPE Framework (SCOPE, 2023) introduced by the INORMS Research Evaluation Group.⁴

Another important problem is that research assessment practices currently often rely on indicators and evidence from proprietary data sources behind paywalls, hindering the transparency of the assessment processes and restricting the ability of the research community to independently scrutinise the findings or to offer alternative interpretations of the data. To make matters worse, these data sources are not designed to be interoperable; hence, it is difficult to combine their contents to achieve improved coverage and capitalise on their plurality of views of the global research track record. Furthermore, the evaluation outcomes are deeply bound to the data source selected for the assessment process, as moving to another data source could, in general, yield profoundly different results.

¹ The Declaration of Research Assessment (DORA), <https://sfedora.org>

² Leiden Manifesto for Research Metrics, <http://www.leidenmanifesto.org>

³ Coalition for Advancing Research Assessment (CoARA), <https://coara.eu>

⁴ INORMS, <https://inorms.net>

Finally, research assessment practices often focus on traditional research outputs, mainly on English-language journal articles and their impact in terms of citations, and fail to properly acknowledge efforts to make research processes more transparent and research outputs more open. Although Open Science (OS) initiatives have gained popularity in recent years, reforming research assessment to acknowledge the contributions researchers and institutions have in OS is essential. A movement towards enabling an *Open-Science-aware Responsible Research Assessment* (OS-aware RRA) is important to encourage and reward the adoption of OS practices. At the same time, organising research assessment on the basis of OS principles, through the use of open data sources and transparent indicators, is increasingly seen as a precondition for realising the ambition of making research assessment more responsible. This shows that OS and RRA need to reinforce each other mutually.

GraspOS aims to alleviate the aforementioned problems by designing and delivering an *Open and Federated Research Assessment Infrastructure* (for brevity, we will refer to it as the *GraspOS infrastructure* from now on). This infrastructure paves the way for the creation of an *Open Research Assessment Dataspace* by aggregating open resources (e.g., data, tools, services, templates and guidance materials offered by different sources) that can catalyse the implementation of policy reforms towards an OS-aware RRA framework.⁵ The value of this infrastructure will be evaluated and showcased in practice by the GraspOS piloting activities (WP5) at different levels of granularity: from a researcher (individual/group) and institutional to an organisational and country level.

An initial version of the architecture to be used for the GraspOS infrastructure was introduced in the D4.1 report⁶, titled "Infrastructure Architecture", in August 2023 and has been refined in Section 3 of the D4.3 report⁷, titled "Federated Open Metrics Infrastructure", in December 2023. In this report, we detail the current version of the architecture, elaborating on various concepts that have been further refined over the past year. This version is considered the primary architecture (GraspOS architecture v1.0) to be presented throughout the project's duration. Any minor refinements that may occur before the project's conclusion will be documented in D4.5, the final infrastructure implementation report.

In the following sections, we discuss the technical details of the architecture to support the GraspOS infrastructure. Section 3 details the process followed for the design of the architecture. Section 4, provides a high-level overview of the architecture, while Section 5 presents the main types of resources that can be hosted on the GraspOS infrastructure and

⁵ This framework (also known as Open Science Assessment Framework - OSAF) will be developed based on the recommendations made by WP2.

⁶ GraspOS Deliverable 4.1: <https://zenodo.org/doi/10.5281/zenodo.8302197>

⁷ GraspOS Deliverable 4.3: <https://zenodo.org/doi/10.5281/zenodo.10475567>

Section 6 delves into important design choices of its core components. Finally, Section 7 concludes the report.

3. Design

This section presents matters related to the design of the GraspOS architecture infrastructure. Section 3.1 presents the adopted methodology. Section 3.2 outlines the main functional goals identified from the project plan and interactions with pilot case representatives. Finally, Section 3.3 discusses the alignment of the design with related initiatives such as CoARA, RDA, and EOSC.

3.1. Methodology

The GraspOS infrastructure is being developed using an *agile, continuous co-design approach*, involving not only the technology-providing project partners (actively contributing in WP3 and WP4) but also research assessment experts (contributing in WP2) and pilot representatives (WP5), who bring perspectives from various stakeholders involved in research assessment processes. Consequently, throughout the project's lifetime, the respective architecture is subject to changes and refinements according to the input collected from the aforementioned expert categories. This input is collected through questionnaires, workshops, regular technical meetings, and internal working groups.

Feedback from research assessment experts primarily pertains to general guidance on responsible research assessment. This feedback influences the metadata specifications used by the catalogues and registries, which collect and expose information for the resources (i.e., data, tools, services, templates and guidance materials) of the GraspOS infrastructure. The main objective is to ensure that the provided metadata will be adequate to serve OS-aware RRA use cases. In addition, the design of the GraspOS API specification (part of the GraspOS Interoperability & Access Layer) is also similarly affected by the same input.

The feedback from the research assessment experts was mainly gathered through discussions during monthly technical meetings, which included participants from WP2, WP3, and WP4. Moreover, the respective GraspOS partners participated in two related internal working groups that have been coordinated by WP4 participants: the "Internal WG on Resource Metadata Schemas" and the "Internal WG on Data Interoperability and Access". The former developed the initial metadata specifications for the entries in the registries and catalogues of the GraspOS infrastructure, while the latter created a first version of the API specification and the corresponding data model to be used by all GraspOS data providers.

Both working groups held several meetings between March and May 2024 and will continue to meet in the future for further refinements. Finally, a limited number of ad-hoc meetings, both virtual and during face-to-face workshops, were arranged to discuss these matters in more detail.

Feedback from pilot representatives is also crucial for the co-creation process, since pilots represent diverse stakeholder categories that conduct research assessment events. As a result, their feedback reflects the needs of the end-users of the GraspOS infrastructure. A first input from the pilots was collected using template documents designed to gather information on the state of affairs, evaluation context, and ambitions for each use case. The respective documents have been analysed, and the results of the analysis have been presented in Deliverable D5.1⁸, titled “Report on pilot setup, current practices and initial requirements”. Moreover, a series of virtual and face-to-face workshops were then held to help pilot representatives delve into the details of their use cases and better understand the OSAF framework, GraspOS resources, and the support provided by the GraspOS infrastructure. More specifically, an online meeting was held in October 2023 to assist pilots in identifying the stakeholders involved in their use cases. In November 2023, a face-to-face workshop was organised in Espoo, Finland, to help pilots better understand the values and context relevant to their use cases. Another face-to-face workshop took place in February 2024, focusing on informing pilot representatives about the tools and services to be delivered by the project’s technical partners. Finally, in May 2024, the last face-to-face workshop was conducted to finalise the roadmaps for the GraspOS piloting activities and the development of the tools and services. All these workshops revealed (among others) valuable information for the architectural design of the GraspOS infrastructure. In addition, various ad-hoc meetings were also arranged with the aim to gather additional requirements from the pilot representatives.

3.2. Functional goals

Based on the initial project plan, and after refinements made to it according to feedback collected using the methodology described in Section 3.1, the GraspOS architecture should support and facilitate the following functional goals:

- The GraspOS infrastructure should federate multiple open (meta)data sources that can offer evidence valuable for the implementation of OS-aware RRA processes, paving the way for the creation of an open research assessment dataspace. Such a dataspace will increase transparency in research assessment processes, helping to move away from

⁸ GraspOS Deliverable 5.1: <https://zenodo.org/doi/10.5281/zenodo.8304450>

paywalled data sources. The GraspOS architecture will support the implementation of this dataspace, supporting important functionalities such as communication and coordination between data providers and consumers, data discovery, and data integration.

- Additionally, the infrastructure should support the promotion of OS and collaboration, as well as the development of new research assessment indicators and methods.
- All GraspOS data sources and services should be interoperable by design, being compliant with relevant interoperability specifications and guidelines. The focus should be on the design of interoperability specifications and guidelines that will facilitate the federation of scholarly (meta)data sources and enable access to the respective contents by end-users and value-added tools and services. These specifications and guidelines will essentially create a set of “onboarding” requirements that can be used for the federation of additional sources in the future. The specifications should be compliant to relevant EOSC guidelines and well-established relevant standards.
- Regarding the contents of the federated data sources, the aim will be for them to collectively offer various types of evidence that are useful for the implementation of OS-aware RRA processes. Indicatively, the objective is to offer bibliographic records, relationships between research-related entities (e.g., authorship linkages, citations) and information for the relevant semantics (e.g., author contribution roles, citation intent), pre-calculated research performance indicators, usage data for research products (e.g., views and downloads of publications), scientific texts (e.g., contents of Open Access publications), research service usage indicators and accounting metadata (e.g., from EOSC accounting system), narrative CVs, research assessment protocols, and, of course, data on practising OS.
- Regarding the value-added tools and services that we are going to build upon the above-mentioned data sources, the focus will be on two categories:
 - *Enrichment tools and services* to enhance research-related entity metadata records with additional metadata (e.g., missing attribute values, link semantics, impact metrics) and missing links (e.g., missing citations, affiliation/authorship links).
 - *Monitoring services* to report and visualise information and indicators to support OS-aware RRA and/or to monitor the uptake and impact of OS from multiple perspectives (e.g., institutional-level, national-level, scientific, and societal).
- It should be noted that GraspOS, of course, aims to also cover *Data services*, which are essentially based on the onboarded data sources that are providing open APIs (compliant with the specifications released by the GraspOS consortium).

- Most of the services will be based on the extension and adaptation of relevant, well-established services already provided by the technology-providing partners in the GraspOS project (e.g., OpenAIRE, UniBo, OPERAS, ATHENA). The developments will partly depend on the requirements of the Open Science Assessment Framework (OSAF; WP2) to ensure that the GraspOS services will facilitate a systematic approach to the implementation of OS-aware RRA processes. In addition, special requirements coming from the use cases of GraspOS pilots (WP5) will be taken into consideration as well.
- Finally, GraspOS services will be ready to be integrated into EOSC (e.g., being available in the EOSC marketplace), contributing to the expansion of EOSC with important services in the field of research assessment and OS monitoring. In addition, all services will leverage existing EOSC components where applicable to reduce development efforts and increase the level of integration with the EOSC ecosystem.

3.3. Alignment with relevant initiatives

The design work for the GraspOS architecture was not conducted in isolation. From the outset, the goal was to integrate important related activities into the process. This includes the work of well-established organisations such as the Research Data Alliance (RDA)⁹ and UNESCO¹⁰, as well as related initiatives like CoARA¹¹ and relevant R&D projects. Additionally, any work related to EOSC was given serious consideration. In the following paragraphs, we delve into the most important examples of this type of alignment work.

Highly relevant to the design of the GraspOS infrastructure architecture are the activities of RDA's interest group on "Open Science Graphs for FAIR Data"¹², particularly the Scientific Knowledge Graph Interoperability Framework¹³ (SKG-IF) that the group is developing. As it is elaborated in Section 6.2, the activities to develop the GraspOS API specification and the respective data model for sharing metadata valuable to research assessment processes, are building upon SKG-IF. The GraspOS model and API specification can be considered to be extensions of SKG-IF specifications, focusing on how to better serve use cases related to research assessment.

Naturally, the architectural design of the GraspOS infrastructure is also taking into serious consideration the activities and the outputs of CoARA, especially those of its OI4RRA

⁹ RDA's website: <https://www.rd-alliance.org/>

¹⁰ UNESCO: <https://www.unesco.org/en>

¹¹ CoARA: <https://coara.eu/>

¹² RDA OSG for FAIR Data IG: <https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig/>

¹³ SKG-IF: <https://skg-if.readthedocs.io/en/latest/>

("Towards Open Infrastructures for Responsible Research Assessment") working group¹⁴. This working group studies the sustainability, interoperability, openness, and accountability of infrastructures of open research information offering data and tools that can enable the reform towards responsible research assessment. Its mission is to mobilise existing Open Infrastructures for RRA and establish a path toward a global OI4RRA ecosystem. It is evident that the objectives of the OI4RRA group are closely aligned with those of the GraspOS project, offering mutual benefits from collaboration. This essential connection has been established since many members of the GraspOS consortium participate in the CoARA working group and some of them have coordination roles in various working group activities. In practice, elements of the GraspOS architectural design have been discussed within the OI4RRA group, providing valuable insights and starting points for discussions, since GraspOS was in a mature stage when OI4RRA was established. Additionally, GraspOS has gained further feedback and consultation from the extended pool of experts participating in the group, influencing the conceptualisation of the architecture.

The subject of open infrastructures for research information is, in general, of particular importance for GraspOS. Until recently, research assessment heavily relied on proprietary data sources and analysis tools. This hindered the transparency of the respective assessment processes and restricted the options to independently scrutinise the findings or offer alternative interpretations of the data. However, in recent years, the Open Science community has created and maintains a multitude of open datasets (e.g., OpenAIRE Graph, OpenAlex, OpenCitations, COKI) that offer competitive alternatives to the proprietary sources in terms of data coverage and quality. In this context, GraspOS is also interested in the activities related to the Barcelona Declaration on Open Research Information¹⁵. This initiative advocates for the use of open infrastructures for research information, aiming to help these infrastructures grow, improve, and gain widespread recognition and adoption. Many GraspOS partners are among the early signees and supporters of the Barcelona Declaration, and they are participating in related meetings and activities. As a result, the main commitments from the Barcelona Declaration have also been considered during the architectural design of the GraspOS infrastructure.

Moreover, the GraspOS architectural design is influenced by various reports discussing the reform towards responsible research assessment and related matters. Indicatively, both the "EC report on Research Assessment"¹⁶ and the "Action Plan by the Commission to implement

¹⁴ OI4RRA: <https://coara.eu/coalition/working-groups/wg-towards-open-infrastructures-for-responsible-research-assessment-oi4rra/>

¹⁵ The Barcelona Declaration: <https://barcelona-declaration.org/>

¹⁶ EC report on Research Assessment: <https://op.europa.eu/en/publication-detail/-/publication/219aa5ea-fae2-11ee-a251-01aa75ed71a1/language-en>

the ten commitments of the Agreement on Reforming Research Assessment (ARRA)¹⁷ contain useful information that has been considered for the design of the GraspOS infrastructure. In addition, various reports addressing relevant topics (e.g., the draft report on "Principles of Open Science Monitoring"¹⁸ submitted to UNESCO by a group of Open Science experts) have been considered to identify any important information that might impact the infrastructure design.

GraspOS architecture is also influenced by the work being done in other Horizon Europe (HE) projects, especially those related to EOSC. OPUS¹⁹, is another HE-funded project focusing on reforming research assessment to incentivise Open Science. Naturally, GraspOS representatives took into consideration the outputs of this project during the design of the GraspOS architecture. In addition, the federated SKG ecosystem introduced by SciLake²⁰ has heavily influenced the architecture adopted by GraspOS. Finally, another notable example is the FAIRCORE4EOSC²¹ project, since some prototype implementations in the GraspOS infrastructure (e.g., the Assessment Registry) will be based on RAiD²² technologies and this project is implementing RAiD service points that can be potentially useful. Finally, GraspOS representatives are closely following the outputs of the EOSC Beyond²³ project to identify options for the better integration and testing of GraspOS components into the EOSC landscape.

Given that GraspOS is a project funded under an EOSC-related Horizon Europe call, the design of the architecture was informed by the latest developments in the EOSC landscape. GraspOS experts aimed to facilitate EOSC integration. Initially, the plan was to ensure that all GraspOS resources would be registered in the EOSC Marketplace, effectively integrating GraspOS infrastructure with EOSC. However, recent extensive redesigns in the EOSC ecosystem have impacted these integration plans. GraspOS experts are closely monitoring all relevant developments participating in all important EOSC-related events and communication channels and will endeavour to find the most effective methods for integrating all valuable GraspOS resources into EOSC, as all details of the new EOSC landscape starts to become more crystallised.

¹⁷ https://research-and-innovation.ec.europa.eu/document/download/e69aff11-4494-4e5f-866c-694539a3ea26_en

¹⁸ <https://www.enssib.fr/bibliotheque-numerique/documents/72242-principles-of-open-science-monitoring.pdf>

¹⁹ OPUS project: <https://opusproject.eu/>

²⁰ SciLake: <https://scilake.eu/>

²¹ FAIRCORE4EOSC: <https://faircore4eosc.eu/>

²² RAiD: <https://raid.org/>

²³ EOSC Beyond: <https://www.eosc-beyond.eu/>

4. Architectural Overview

Figure 1 offers a high-level representation of the current version of the GraspOS infrastructure architecture. As explained in Section 2, the presented architecture builds upon the concepts introduced in D4.1 and further refined in D4.3. The main objective of this architecture is to enable the discovery and access of *GraspOS resources* (data, tools, services, templates, and guidance materials valuable to implement RRA processes) from the *end-users*, and facilitate the work of research assessment tool and service developers in consuming (meta)data assets from a set of supported, federated scholarly data sources. The *core components* of the GraspOS architecture are the following:

- The **Data Registry**, which facilitates the discovery of the data included in the GraspOS federated infrastructure and serves as an inventory of the data sources and assets.
- The **Data Interoperability & Access Layer** (also known as “DIAL”), which comprises the union of all the API endpoints that offer access to the contents of the federated data sources included in the GraspOS infrastructure.
- The **Tools Catalogue**, which contains information for all stand-alone GraspOS enrichment & monitoring tools and facilitates their discovery.
- The **Services Catalogue**, which contains information for all enrichment, monitoring, and data services provided by the GraspOS infrastructure and facilitates their discovery.
- The **Templates & Guidance Materials Catalogue**, which facilitates the discovery of research assessment templates, guidelines, checklists, and indicator toolboxes (see also Section 4 of deliverable D2.2 “OSAF”²⁴).
- The **Infrastructure front-end UI**, which is the entry point of the GraspOS infrastructure for the end users, which also offers high-level exploration functionalities for its resources. It provides easy access to the GraspOS catalogues and registries.

²⁴ GraspOS Deliverable 2.2: <https://zenodo.org/doi/10.5281/zenodo.10475459>

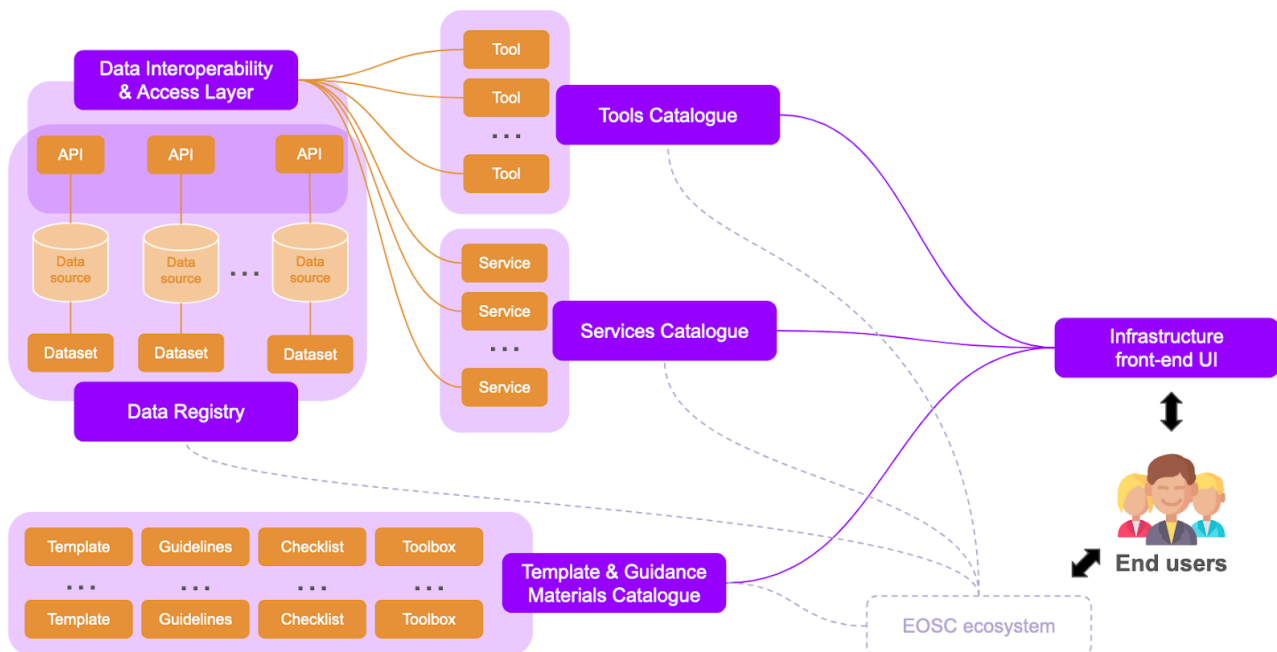


Figure 1 Overview of the GraspOS infrastructure core components

The aforementioned GraspOS core components implement an environment that aims to enable the federation of resources (tools, services, datasets, templates, guidance materials), which can be combined for the implementation of research assessment and monitoring processes focused on OS-aware RRA. More details on the current array of such resources that are already hosted (or planned to be hosted) in the current deployment of the GraspOS infrastructure can be found in Deliverable D3.2, titled “Tools and services”,²⁵ and in Deliverable D2.2²⁶, titled “OSAF”. Implementation details for each core component are provided in Deliverable D4.3²⁷ and in its iterations (Deliverables D4.4²⁸ and D4.5 - to be published).

In general, any individual accessing one of the core components or using a GraspOS resource is considered to be an *end user* of the GraspOS infrastructure. The GraspOS architecture is designed with the aim to create an ecosystem that not only supports evaluators in their job, but also facilitates the work of scholarly data managers and developers of tools and services which can be useful in research assessment. For instance, the architecture aims to facilitate the programmatic access of the supported data sources by the developers of monitoring tools and services and the curation of scholarly data sources by their managers leveraging the available enrichment tools and services. This is implemented by the GraspOS Data

²⁵ GraspOS Deliverable 3.2: <https://zenodo.org/doi/10.5281/zenodo.13619176>

²⁶ GraspOS Deliverable 2.2: <https://zenodo.org/doi/10.5281/zenodo.10475459>

²⁷ GraspOS Deliverable 4.3: <https://zenodo.org/doi/10.5281/zenodo.10475567>

²⁸ GraspOS Deliverable 4.4: <https://zenodo.org/doi/10.5281/zenodo.13617815>

Interoperability and Access Layer, which offers unified APIs connecting the data sources to the tools and services (as indicated by the orange links in Figure 1).

Moreover, as it is indicated in Figure 1, all GraspOS catalogues and registries are designed to be ready for integration into the EOSC ecosystem. The initial plan was to make sure that all resources provided will be registered to the EOSC Marketplace, essentially offering integration of the GraspOS infrastructure with EOSC. However, the EOSC ecosystem has been heavily redesigned during the last months, affecting any integration plans for services and other resources. In the context of this redesign, EOSC Marketplace has been discontinued. In theory, GraspOS approach remains relevant as the implementation of catalogues for all resources, excluding services, ensures that these resources will be findable through the Resource Hub of the EOSC EU Node²⁹ (for more details, see Deliverable D4.4 - to be published). However, the policy for onboarding services to EOSC has not yet been fully crystallised. GraspOS representatives are closely monitoring developments in this area and will strive to identify the best methods for integrating all valuable GraspOS services into EOSC.

Finally, to keep Figure 1 legible, we did not include all the details about the ways the end users can access the various components and resources of the infrastructure. For instance, the GraspOS registries and catalogues, the API endpoints of the various GraspOS services, and the Data Interoperability & Access Layer can be directly accessed by the end users.

It's important to emphasise that the GraspOS infrastructure functions as an ecosystem designed to easily federate assessment resources, rather than as a centralised infrastructure. While the core components are centralised, all resources, such as the services provided, are hosted in various locations. The architecture, through its central components, facilitates the flow of information between the onboarded resources, enhances their discoverability, and simplifies their integration into research assessment workflows.

²⁹ EOSC EU Node Resource Hub: <https://open-science-cloud.ec.europa.eu/resources/all>

5. Federated Resources

This section outlines the various types of resources that can be hosted by the GraspOS federated infrastructure, as well as the approach for the inclusion of resources in the infrastructure.

5.1. Types of Federated Resources

The GraspOS infrastructure is designed to federate an array of resources that can be valuable for research assessment, making them easy for end users to discover and access. More specifically, the following main types of resources are planned to be included in the GraspOS infrastructure:

- **Datasets**, including raw or structured data that can be used to either directly support the conduction of OS-aware RRA events or produce indicators, track records, narratives or other evidence that could be valuable in such processes.
- **Tools**, including stand-alone pieces of software, such as scripts, executables, libraries, or even workflows, which can be used by installing and executing them locally on a computer or computational cluster owned by the end-user to conduct a particular type of analysis and/or produce data.
- **Services**, including pieces of software that provide a set of functions or features for the end user and that are typically hosted on a remote server or cluster and accessed via a Web interface or API. Essentially, the service providers aim to offer them in a 24/7 manner giving guarantees for the respective Quality of Service (QoS), such as the service availability.
- **Templates & guidance materials**, including various templates for documents that can be used during the research assessment process (e.g., related to assessment readiness, stakeholder mapping, value and purpose statement), guideline documents (e.g., OS assessment guide, EDI guide), checklists, and indicator toolboxes. More details for these resources can be found in Section 4 of D2.2 "OSAF".³⁰

For each type of resources, a dedicated registry or catalogue has been established in the GraspOS infrastructure architecture to facilitate discoverability and accessing. More specifically, a Data Registry³¹ has been created for the datasets and a Tools³² and a Services³³

³⁰ GraspOS Deliverable 2.2: <https://zenodo.org/doi/10.5281/zenodo.10475459>

³¹ GraspOS Data Registry: <https://graspos-data.athenarc.gr/>

³² GraspOS Tools Catalogue: <https://zenodo.org/communities/graspos-tools/records>

³³ GraspOS Services Catalogue: <https://graspos-services.athenarc.gr/home>

Catalogue has been created for the respective resource types. The need for the Templates & Guidance Materials Catalogue has been identified based on the work performed in the context of D2.2, and GraspOS partners are in the process of implementing it (first version is expected to be released soon). Brief introductions for all aforementioned components can be found in Section 6.1, while implementation details (and the respective code bases) can be found in deliverables D4.3 “Federated Open Metrics Infrastructure (FOMI)”³⁴ and D4.4 “Federated Open Infrastructure (FOMI)”³⁵.

5.2. Rules of Resource Inclusion

The first group of resources (i.e., data, tools, services, templates, and guidance materials) that have been included in the GraspOS infrastructure was selected based on the competencies of the partners of the project. However, the GraspOS infrastructure is planned to act as an open and evolving ecosystem. This means that the list of assessment resources currently included in the infrastructure is always subject to changes (mostly additions). For instance, it is desirable for data and software creators to include additional resources in the future. To make this possible a resources inclusion policy should be determined.

At this stage, the focus of the project is to establish the basic infrastructure and to make available and test in practice an initial set of resources. Consequently, the details related to the inclusion policy will be refined and finalised at a later point. However, all resources to be included should:

- be relevant to the research assessment domain
- provide the required metadata so that they can be included in the respective GraspOS catalogues and/or registries
- have in place API endpoints which are compliant to the specifications followed by GraspOS, where applicable
- offer the minimum QoS guarantees that will be decided by the GraspOS infrastructure administrators for all services
- follow the recommendations made by the OSAF framework (where applicable)
- be compliant to other relevant best practices which have been identified as relevant to the GraspOS infrastructure (e.g., the EOSC criteria for Service Providers).

³⁴ GraspOS Deliverable 4.3: <https://zenodo.org/doi/10.5281/zenodo.10475567>

³⁵ GraspOS Deliverable 4.4: <https://zenodo.org/doi/10.5281/zenodo.13617815>

6. Core Components

In this section, we discuss important design choices for the GraspOS infrastructure core components, and we outline their basic technical information. Implementation details for all these components are provided in Deliverables D4.3³⁶ "Federated Open Metrics Infrastructure (FOMI)" and D4.4³⁷ "Federated Open Infrastructure (FOMI)" - to be published.

6.1. Catalogues & Registries

An array of catalogues and registries, that enable the discovery of and access to the resources included in the GraspOS federated ecosystem have been designed to comprise central components of the architecture.

The **GraspOS Data Registry** will allow users to be informed about and understand the metadata contents within the infrastructure, facilitating searching for or browsing all available datasets. The Data Registry is built upon the CKAN³⁸ open-source software, and it will collect metadata and access information for the onboarded data sources. The CKAN basic metadata model is being extended to keep research-assessment-specific information (details in D4.4) based on the recommendations made by the Internal WG on "Resource Metadata Schemas" (see also Section 3).

The **Tools Catalogue** contains information for all stand-alone GraspOS enrichment & monitoring tools and facilitates their discovery. Its implementation is based on Zenodo communities (details in D4.4). The collected metadata are based on those that Zenodo is asking to be collected for Software entries. A set of predefined keywords have also been determined to be used for those entries according to the recommendations made by the Internal WG on "Resource Metadata Schemas".

The **Services Catalogue** contains information for all enrichment, monitoring, and data services provided by the GraspOS infrastructure and facilitates their discovery. Its implementation is based on the OpenAIRE Catalogue.³⁹ The main reason for this selection was related to the fact that this catalogue software was supporting a metadata model for services which was extending the one used by the EOSC Marketplace, hence would facilitate the

³⁶ GraspOS Deliverable 4.3: <https://zenodo.org/doi/10.5281/zenodo.10475567>

³⁷ GraspOS Deliverable 4.4: <https://zenodo.org/doi/10.5281/zenodo.13617815>

³⁸ CKAN: <https://ckan.org/>

³⁹ OpenAIRE Catalogue: <https://catalogue.openaire.eu/>

alignment with EOSC practices. The Internal WG on “Resource Metadata Schemas” decided which subset of the fields would be used for the first stage (details in D4.4).

Finally, the **Templates & Guidance Materials Catalogue** facilitates the discovery of research assessment templates, guidelines, checklists, and indicator toolboxes. It is a recent addition to the list of GraspOS registries and catalogues, based on the work reported in D2.2. Its implementation has not started yet, however, the intention is to follow a similar approach like the one used for the implementation of the Tools Catalogue (i.e., it will be based on Zenodo communities, as well).

6.2. Interoperability & Access Layer (DIAL)

One of the primary objectives of the GraspOS infrastructure is to enhance the interoperability of various scholarly (meta)data sources, facilitating the consumption and integration of (meta)data assets from diverse origins. The Interoperability & Access Layer (DIAL) of the GraspOS infrastructure architecture is the component responsible for achieving this objective.

DIAL comprises the union of the API endpoints that offer access to the contents of the federated data sources included in the GraspOS infrastructure. This means that each GraspOS data service is expected:

- *Requirement 1:* To have an API deployment exposing its contents to end users (researchers, developers)
- *Requirement 2:* To share metadata related to this deployment (including accessing information) through the GraspOS services catalogue and data registry (see Section 6.1).
- *Requirement 3:* To implement a set of API endpoints and datasets compatible with a common API and metadata model specification that has been determined to increase the interoperability of the GraspOS data services.

Essentially, the implementation of DIAL consists of preparing and putting together GraspOS-compliant API endpoints for all GraspOS data services. For each data service, a phased implementation approach is followed: during the initial phase, the effort is given to addressing Requirements 1 & 2; during the production phase, the data service managers are required to meet Requirement 3.

Most GraspOS data services already had an API deployment in place before the project started; hence, for the majority of services, Requirement 1 has already been addressed. Addressing Requirement 2 mainly involves registering API-related metadata in the GraspOS

services catalogue and data registry. Regarding Requirement 3, the data services managers are requested to transform or/and extend their API endpoints and any open datasets provided so that they will be compliant with the *GraspOS API and metadata model specification* (see below).

Following a common specification for (meta)data sharing (through APIs and/or datasets) is essential to ensure the interoperability of data services included in the GraspOS ecosystem and facilitate accessing and querying their contents to build value-added services. Developing this common GraspOS API and metadata model specification was the main objective of the "Internal WG on Data Interoperability and Access" (see Section 3.1). This specification was selected to be an extension of RDA's SKG-IF (see Section 3.3), focusing on missing pieces that are, however, important for research assessment. In the rest of the section, we further elaborate on the GraspOS metadata model and API specification.

The mission of RDA's working group that is developing the SKG-IF model is to target the high fragmentation, heterogeneity and replication of scholarly information across different Scientific Knowledge Graphs (SKGs), reduce duplication of effort, and capitalise on synergies and complementarity. By comparing and putting to a common factor the modelled information contained in different SKGs publicly available such as Crossref, OpenAIRE, OpenCitations, and OpenAlex, the SKG-IF

- captures a set of core entities participating in scholarly communication processes such as research products, persons, grants, and organisations and their most relevant properties and relationships, and
- provides guidelines towards a uniform and pragmatic exchange of core information.

In its current version (the last release at the time of writing being V1.1⁴⁰), the entity types modelled by SKG-IF are as follows:

- **Research product**, which represents tangible outputs of research processes and may be of four types:
 - *Literature*, intended for reading by humans (article, thesis, peer-review, blog posts, books, reports, patents, etc.)
 - *Research data*, self-contained, persistently identified digital assets intended for processing (e.g., files containing tables, metadata collections, dumps; persistent dynamic queries to scientific databases)

⁴⁰ <https://skg-if.readthedocs.io/en/v1.1/>

- *Research software*, including source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose⁴¹.
- *Other products*, i.e., any digital asset, uniquely identified, whose nature does not fall into the first three types.
- **Person**, which represents individuals who are involved in the creation, publication, and dissemination of Research products. A Person can be an author, a reviewer, an editor, a publisher, a researcher, or any other stakeholder involved in the scholarly communication process.
- **Organisation**, representing academic institutions, research centres, funders, or any other institutions taking part in the research process.
- **Venue**, that models publishing “gateways” used by a Person to make their Research products available to others.
- **Data source**, represents services where published material (metadata and files) are stored, preserved, and made discoverable and accessible. A data source is described by the EOSC Profile for [data sources](#).
- **Grant**, describes funding awarded to a Person or an Organisation by a funding body. These bodies, both public and private, can be funding organisations, foundations, governments, agencies or institutions.
- **Topic**, describing the scientific disciplines, the subjects and the keywords potentially relevant to a Research product.

It is evident that the previous entity types are very useful in research assessment use cases; hence, many of the datasets and data sources included in the GraspOS ecosystem are expected to cover them, offering valuable metadata. However, at the moment, the SKG-IF does not offer details about the modelling or exchange of information that is conducive to the GraspOS context (e.g., indicators and narratives associated with research products, individuals, and institutions). Because of this, it was essential for GraspOS experts to implement an extension that focuses more on aspects which are important for research assessment.

Based on the discussions that took place in the GraspOS working group on Data Interoperability and Access, a first version of the planned SKG-IF extension has been created,

⁴¹ It is worth mentioning that software components (e.g., operating systems, libraries, dependencies, packages, scripts, etc.) that are used for research but were not created during or with a clear research intent should be considered software in research and not Research Software.

with the focus given on the inclusion of indicators and narratives, since these were the most important missing notions that have been identified. The inclusion of indicators was implemented by the addition of a relevant field in the Research product, Person and Organisation entity types. The inclusion of narratives was implemented by the addition of a brand-new entity and its specific relationships with the SKG-IF core entities. Full details on these extensions can be found in Deliverable D4.4 " FOMI (2/3)".⁴²

It is worth mentioning that the SKG-IF working group approved an extension mechanism⁴³ of the SKG-IF model, which is precisely devised to capture peculiarities and requests coming from user communities or application contexts and projects. GraspOS experts are planning to leverage this mechanism and submit an application for an official SKG-IF extension, focusing on research assessment (based on the extensions discussed in the previous paragraphs).

The principles guiding the development and implementation of extensions within the SKG-IF framework emphasise the following key factors:

- **Shared Interest/Need:** Extensions should address a collective interest or requirement identified within projects, communities, or domains rather than catering to individual or isolated needs. This principle ensures that extensions serve a broader purpose and contribute to the overall advancement of scientific knowledge representation.
- **Non-Interference:** Extensions should not interfere with or disrupt the entities already defined within the SKG-IF. They should not serve as shortcuts for sharing information that should be placed elsewhere within the framework. By adhering to this principle, the integrity and coherence of the SKG-IF are preserved.

By adhering to these guiding principles, the development and adoption of SKG-IF extensions are driven by a commitment to collaboration, standardisation, interoperability, and continuous improvement in representing and sharing scientific knowledge effectively across diverse research domains and communities.

A community SKG-IF extension is structured in three parts:

- **Extension-specific entities:** entities that reflect semantics and structure different from the core entities, i.e. no core entity can be seen as a super-entity of community entities;
- **Core entity extensions:** properties that extend the set of properties of core entities to address the needs of the specific community;

⁴² GraspOS Deliverable 4.4: <https://zenodo.org/doi/10.5281/zenodo.13617815>

⁴³ SKG-IF extensions: <https://skg-if.readthedocs.io/en/latest/extensions.html>

- **Relationships introduced by the extension:** relationships that link entities (either core's or community's) reflecting community-specific semantics.

SKG-IF extensions are regulated and tracked by a process starting with the submission of a case statement. Once the extension is endorsed by the working group, the extension maintainers will have to provide documentation pages that will be included in the official SKG-IF documentation within a section dedicated to supported extensions. Once the extension is endorsed, its lifecycle will be monitored. If the purposes an extension serves are proved to be conducive for many applications, the extension (or the relevant subparts) can be marked for inclusion in the "core" SKG-IF model.

6.3. Infrastructure Front-end UI

This component is a front-end user interface (UI) that can be used by the GraspOS end-users to help them navigate between the various catalogues and registries of the GraspOS infrastructure⁴⁴, but also to explore the resources (data, tools, services, templates, and guidance materials) hosted and exposed by the GraspOS infrastructure. This UI is essentially the entry point to the infrastructure.

Accessing the GraspOS resources can be done without the need for using a centralised "entry point", since the whole infrastructure is a distributed system of federated resources. However, an entry point is useful to help end users discover and browse the contents of the GraspOS infrastructure. As a result, a front-end UI has been developed and can be used for this purpose.

As a Web-UI component, a Web development framework was selected for the respective implementation activities. More specifically, the development was based on Docusaurus,⁴⁵ a framework that allows building and deploying customised and highly optimised project websites.

⁴⁴ GraspOS Infrastructure front-end UI: <https://graspos-infra.athenarc.gr/>

⁴⁵ Docusaurus: <https://github.com/facebook/docusaurus>

7. Conclusions

This deliverable report outlines the architecture of the GraspOS (federated) infrastructure. This infrastructure is designed to support the implementation of OS-aware RRA processes, the promotion of OS practices, and the development of new research assessment and monitoring services, tools, and datasets. The infrastructure is also designed to be flexible and adaptable, allowing for the easy integration of new resources.

The GraspOS infrastructure is composed of a set of core components that facilitate the interoperability of an array of scholarly (meta)data sources and improve the discoverability of research assessment resources of various kinds (data, tools, services, templates, and guidance materials). Essentially, it aims to become the de-facto entry point for everyone interested in implementing research assessment events in a responsible way.

8. References

de Rijcke , S., Cosentino , C., Crewe , R., D'Ippoliti, C., Motala-Timol , S., Binti A Rahman , N., Rovelli , L., Vaux, D., & Yupeng , Y. (2023). The Future of Research Evaluation: A Synthesis of Current Debates and Developments. <https://doi.org/10.24948/2023.06>

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810), 56-63. <https://doi.org/10.1126/science.159.3810.56>

Curry, S., de Rijcke, S., Hatch, A., Pillay, D. (Gansen), van der Weijden, I., & Wilsdon, J. (2020, November 18). The changing role of funders in responsible research assessment: progress, obstacles and the way ahead. [Rori.figshare.com](https://doi.org/10.6084/m9.figshare.13227914.v2). <https://doi.org/10.6084/m9.figshare.13227914.v2>

Strinzel, M., Brown, J., Kaltenbrunner, W., de Rijcke, S., & Hill, M. (2021). Ten ways to improve academic CVs for fairer research assessment. *Humanities and Social Sciences Communications*, 8, 251. <https://doi.org/10.1057/s41599-021-00929-0>

Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biol* 18(7): e3000737. <https://doi.org/10.1371/journal.pbio.3000737>

International Network of Research Management Societies - Research Evaluation Group (2023). The SCOPE Framework. <https://doi.org/10.26188/21919527.v1>