Federated Open Metrics Infrastructure (FOMI)

Deliverable 4.3

Due Date of Deliverable	31.12.2023
Actual Submission Date	31.12.2023
Work Package	WP4
Tasks	T4.1, T4.2
Type	Other
Approval Status	Accepted
Version	1.0
Number of Pages	27
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 deliverable report offers details on the development of FOMI, the GraspOS federated infrastructure, and contains references to the code repositories and the documentation websites of the respective software packages.

Revision history

VERSION	DATE	REASON	REVISED BY
0.1	18.12.2023	First Draft	Thanasis Vergoulis, Serafeim Chatzopoulos, Giulia Malaguarnera, Andrea Mannocci
0.2	22.12.2023	Peer review	Ivan Heibi, Kumar Guha, Alban Thomas, Odile Hologne
0.3	28.12.2023	Peer review comments addressed	Thanasis Vergoulis, Serafeim Chatzopoulos, Giulia Malaguarnera, Andrea Mannocci
1.0	30.12.2023	Final Version after proofreading	Thanasis Vergoulis, Serafeim Chatzopoulos

Author List

ORGANISATION	NAME	CONTACT INFORMATION
ATHENA RC	Thanasis Vergoulis	vergoulis@athenarc.gr
OpenAIRE	Giulia Malaguarnera	giulia.malaguarnera@openaire.eu
CNR	Andrea Mannocci	andrea.mannocci@isti.cnr.it
ATHENA RC	Serafeim Chatzopoulos	schatz@athenarc.gr

Contributor List

ORGANISATION	NAME	CONTACT INFORMATION
ATHENA RC	Mary Melekopoglou	marmel@athenarc.gr
INRAE	Alban Thomas	alban.thomas.2@inrae.fr
INRIA	Kumar Guha	kumar.guha@inria.fr
UNIBO	Ivan Heibi	ivan.heibi2@unibo.it
INRAE	Odile Hologne	odile.hologne@inrae.fr

Table of Contents

List of Tables

List of Figures

Abbreviation List

- *API:* Application Programming Interface
- *DB:* Database
- *DBLP:* database systems and logic programming (a widely know database for CS publications)
- *EOSC:* European Open Science Cloud
- *FOMI:* Open & Federated Research Assessment Infrastructure (previously called "Federated Open Metrics Infrastructure")
- *ID:* Identifier
- *ORAD:* Open Research Assessment Dataspace
- *OS:* Open Science
- *OS-aware RRA:* Open-Science-aware Responsible Research Assessment
- *OSAF:* Open Science Assessment Framework
- *PID: Persistent Identifier*
- *QoS:* Quality of Service
- *RDA:* Research Data Alliance
- *RFO:* Research Funding Organization
- *RPO:* Research Performing Organization
- *RRA:* Responsible Research Assessment
- *UI:* User Interface
- *URL:* Uniform Resource Locator

1. Executive Summary

Research assessment covers a pivotal role in a variety of applications in research, ranging from informing the hiring and promotion of research staff to shaping future strategic investments and policymaking. GraspOS aims to deliver a federated infrastructure, called FOMI, which will support research assessment processes by incorporating and making available a set of valuable data sources, tools, and services. The focus of this infrastructure will be on facilitating Open-Science-aware Responsible Research Assessment (OS-aware RRA) essentially catalysing the creation of an open, federated research assessment dataspace.

The infrastructure is designed to be modular and scalable, allowing for the integration of new components 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 based on a federated architecture and is distributed across multiple components to be more resilient and adaptable to changing needs and contexts.

Overall, the GraspOS infrastructure is composed of a set of federated Data Asset Sources providing access to and delivering metadata for research products, pre-calculated research performance indicators, as well as unstructured data such as scientific texts, narrative CVs and data on Open Science (OS) practices. Such sources are flanked with a set of federated monitoring and enrichment services, which can be used to handle the transformation, manipulation, and analysis of data within the infrastructure, present and visualise research assessment evidence and indicators at different levels of granularity (e.g., researcher, organisation, country) to support OS-aware RRA and to monitor the uptake and impact of OS. A Data Registry documenting the federated resources, two EOSC-interoperable catalogues (one for tools and one for services), a data interoperability and access layer (that facilitates metadata consumption) and an intuitive infrastructure front-end user interface complement the infrastructure.

Deliverable D4.1[1] described the initial design of the GraspOS architecture. In this report, implementation details and references to the respective code repositories and documentation websites of the various components of the architecture are provided.

---

[1] A. Mannocci, T. Vergoulis, S. Chatzopoulos, S. Peroni, G. Malaguarnera, S. Dumouchel, & C. Tatum. (2023). GraspOS Deliverable 4.1 "Infrastructure Architecture" (1.0). Zenodo. https://doi.org/10.5281/zenodo.8302198

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/economic/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.

Nowadays, commonly used practices for research assessment face major challenges (e.g., related to the use and abuse of often misunderstood and non-transparent quantitative indicators, the reliance on proprietary data sources, or putting too much emphasis on publications while ignoring other types of contributions in research) and there is a growing recognition that assessing research and researchers needs to be done in a more responsible way[2], which will:

- take into consideration (a) the values, scope, and context of each assessment event, (b) qualitative evidence together with supporting quantitative data and indicators, and (c) multiple aspects and merits of research work that should be acknowledged, and
- enable transparency in the assessment processes, focusing on the use of open scholarly data sources

Furthermore, since openness in science is widely recognized due to its effect in research credibility and knowledge sharing, there is a need for frameworks and technologies that will enable an *Open-Science-aware Responsible Research Assessment (OS-aware RRA)* to encourage and reward the adoption of OS practices from researchers.

GraspOS aims to alleviate the aforementioned problems by designing and delivering FOMI, an *Open and Federated Research Assessment Infrastructure* (formerly known as "Federated Open Metrics Infrastructure"). This infrastructure paves the way for the creation of an *Open Research Assessment Dataspace* (ORAD) by aggregating open resources (e.g., next-generation metrics and indicators, data, tools, services, and guidance offered by different sources) that can catalyse the implementation of policy reforms towards OS-aware RRA. A first version of the initial architecture design is described in D4.1 - "Infrastructure Architecture"[3] (a slightly updated version of it is briefly presented in Section 3 of this report).

---

[2] CoARA, 2022, The Agreement on Reforming Research Assessment. https://coara.eu/agreement/the-agreement-full-text/
[3] D4.1 "Infrastructure architecture" (draft): https://zenodo.org/records/8302198

This report contains references to the code bases of the software packages that contribute to the creation of FOMI and documents the implementation of the infrastructure. In the following sections, we offer a high-level description of the architecture to support the FOMI infrastructure and ORAD and, then, we elaborate on the implementation details of the respective software components.

3. Architecture

In this section, we offer a brief overview of FOMI's architecture. A more detailed description of the architecture, including additional information (e.g., the respective functional goals), can be found in D4.1 ``Infrastructure architecture''. It should be noted that, as explained in D4.1, the architecture is being developed following an agile, continuous design approach, hence it is subject to changes and refinements according to the input collected during the implementation of the GraspOS assessment framework (WP2), the GraspOS tools & services (WP3), and the GraspOS pilots (WP5).[4] Because of that, the current description includes small modifications compared to the initial design.

3.1. Overview & Core Components

Figure 1 offers a high-level representation of the FOMI architecture. The aim of this architecture is to enable the discovery and access of GraspOS data, tools, and services from the end-users and to facilitate the work of tool and service developers in consuming (meta)data assets from a set of supported, federated scholarly data sources. FOMI consists of the following *core components*:

- The **Data Registry**, which facilitates the discovery of the data included in the GraspOS infrastructure and serves as an inventory of the data sources and assets.
- The **Data Interoperability & Access Layer**, 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 **Services Catalogue**, which contains information for all enrichment, monitoring, and data services provided by the GraspOS infrastructure and facilitates their discovery.
- The **Tools Catalogue**, which contains information for all stand-alone GraspOS enrichment & monitoring tools and facilitates their discovery.

---

[4] It is also worth mentioning that the final architecture will be submitted in D4.2 at M20 - August 2024.

- The **Infrastructure front-end UI**, which is the entry point of the GraspOS infrastructure for the end users offering also high-level exploration functionalities for its resources.

In general, any individual accessing one of the core components (i.e., the purple boxes in Figure 1) or using a GraspOS data asset, tool or service (i.e., the orange boxes) is considered to be an *end user* of the GraspOS infrastructure. More details about each core component and its implementation are provided in Section 4. 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. Moreover, as it is indicated in Figure 1, all GraspOS catalogues and registries are designed to be connected (and synchronised) to the EOSC Marketplace offering integration of the GraspOS infrastructure with 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 catalogues of tools and services, the Data Registry, the API endpoints of the various GraspOS services and the Data Interoperability & Access Layer are all open and can be accessed directly by the end users.
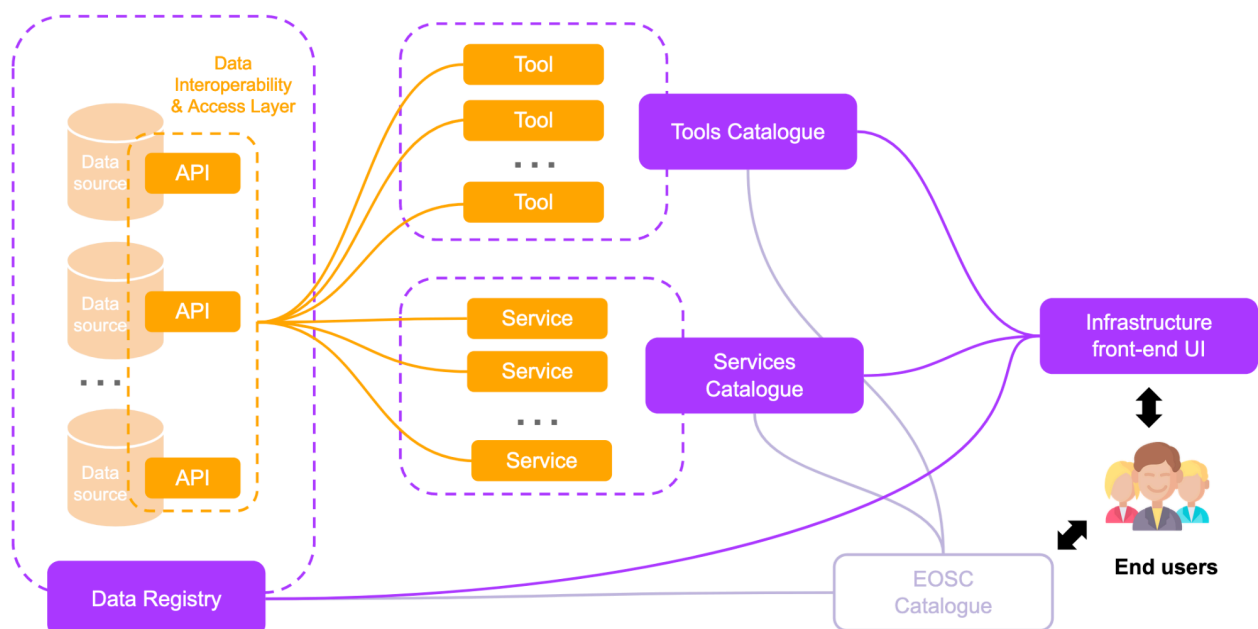

Figure 1 *Overview of the FOMI architecture*

The next sections offer valuable information on the main resources (data, tools, and services) which are hosted and exposed by the GraspOS infrastructure (Section 3.2), their access modes (Section 3.3), and their integration to EOSC (Section 3.4).

3.2. Hosted & Exposed Resources

In this section, we provide more details on the data assets (Section 3.2.1) and the tools and services (Section 3.2.2) that are hosted and exposed by FOMI. It should be noted that the list of data sources, tools, and services included in FOMI refers to the current planning, hence it is subject to changes (mostly additions), since the GraspOS infrastructure aims to act as an open infrastructure in which data and software creators can include additional sources in the future. Section 3.2.3 elaborates more on the planned data, tools, and services inclusion process.

3.2.1.   Data Assets

FOMI is designed to integrate a set of federated data asset sources, which provide both raw and structured data which can be used to either directly support the conduction of OS-aware RRA events or produce indicators or other evidence that could be valuable in such processes. The following table presents the most important data sources offered by GraspOS partners which have been decided to be part of FOMI (more may be added in the future).

Table 1 Federated data asset resources integrated in FOMI

Source	Files & API	Licence	Description
BIP! DB	Files: https://doi.org/10.5281/zenodo.4386934 API: https://bip-api.imsi.athenarc.gr/documentation	CC-BY	This data source offers citation-based impact indicators calculated for ~153M distinct DOIs that correspond to research products.
BIP! NDR	Files: https://doi.org/10.5281/zenodo.7962019	CC-BY	It contains more than 2.7M citations made by approximately 163K Computer Science conference or workshop papers that,

			according to DBLP, do not have a DOI.
OpenAIRE Graph	Files: https://zenodo.org/records/10037121 API: https://graph.openaire.eu/develop/overview.html	CC-BY (parts are also in CC0)	The OpenAIRE Graph is one of the largest open scholarly record collections worldwide, key in fostering Open Science and establishing its practices in the daily research activities.
OpenCitations data	Files: https://opencitations.net/download API: https://opencitations.net/querying	CC0	This data source includes OpenCitations Index and OpenCitations Meta, which constitute valuable resources (citations and metadata, respectively) for bibliometric analyses, such as estimating the impact of research products and enabling their reproducibility.
OPERAS Metrics	N/A	N/A	It is a usage and altmetrics platform for Open Access publishers in the Humanities and Social Sciences. It collects usage and impact metrics related to published Open Access content from many different sources (monographs, journals, repositories) and

			allows their access, display and analysis from a single access point.
ScholExplorer	Files: https://doi.org/10.5281/zenodo.1200252 API: https://api.scholexplorer.openaire.eu/v2/ui/	CC-BY	This dataset contains Scholix links exposed by the OpenAIRE ScholeXplorer service. It consists of bidirectional links between literature-to-dataset and dataset-to-dataset.
Usage Counts	The respective data are accessible through the OpenAIRE Graph.	CC-BY	Usage Counts Service gathers usage events and consolidated usage statistics reports, respectively, from a distributed network of data providers (repositories, e-journals, CRISs) by utilising open standards and protocols and exploiting reliable, consolidated and comparable usage metrics like counts of item downloads and metadata views conformant to COUNTER Code of Practice.
EOSC services accounting	N/A	N/A	A dataset that collects service metrics and

			aggregates them according to EOSC/ community rules. It would be used as a data source for impact metrics for the services.
Assessment Registry	N/A	N/A	This registry enables the publication of assessment protocols after the completion of the respective assessment events. Details will follow at a later stage.

It should be noted that some of the data sources are currently being developed (e.g., the Assessment Registry) or restructured severely, hence the respective URLs for data files and APIs are not yet finalised (but they will be provided in later versions of the deliverable).

Moreover, it should be noted that a main objective of FOMI is to facilitate the consumption of the contents of the previous data sources from the developers of GraspOS tools and services (and of third-party added-value services in the future). For this reason, FOMI contains a "Data Interoperability and Access Layer" which aims to facilitate the integration and harmonisation of (meta)data from various sources and formats within the infrastructure. At this point, this layer just comprises the endpoints of the APIs already provided by the various data sources. However, in the future, the plan is for all data sources to offer endpoints which are compatible with each other following the guidelines and specifications of widely accepted standards (e.g., those provided by RDA[5] like the "Scientific Knowledge Graph Interoperability Framework"[6] delivered by the respective RDA working group) or extending them accordingly.

3.2.2.   Tools & Services

GraspOS infrastructure is expected to host a set of tools and services which are divided in the following main *bundles*:

---

[5] Research Data Alliance (RDA): https://www.rd-alliance.org/
[6] Scientific Knowledge Graph Interoperability Framework (SKG-IF): https://skg-if.readthedocs.io/

- A bundle of **Enrichment Tools & Services**, which can be used to handle the transformation, manipulation, and analysis of data within the infrastructure to produce enrichments that can be of value in the context of OS-aware RRA. These tools and services are also expected to contribute in the further enrichment of the scholarly data sources included in the GraspOS infrastructure (see also Section 3.2.1).
- A bundle of **Monitoring Tools & Services**, which focus on (a) reporting and visualising research assessment evidence and indicators at different levels (e.g., researcher, organisation, country) to support OS-aware RRA processes, and (*b*) monitoring the uptake and impact of OS from multiple perspectives (e.g., institutional-level, national-level, scientific societal).

At this point, it is essential to explain the main difference between tools and services, according to the conventions we have adopted for the project. Both terms refer to types of software, however, each comes with a different set of purposes and characteristics. We use the term *"tool"* to refer to a stand-alone piece of software, such as a script, an executable or even a workflow, which can be used by installing and executing it locally on a computer or computational cluster owned by the end-user to conduct a particular type of analysis and or produce data. We use the term *"service"* to refer to a piece of software that provides a set of functions or features for the end user and that is typically hosted on a remote server and accessed via a Web interface or API. Essentially, the service provider aims to offer it in a 24/7 manner giving guarantees for the respective Quality of Service (QoS), such as the service availability.

Any individual using a GraspOS tool or service is, by definition, considered to be a GraspOS infrastructure end user. Of course, accessing a GraspOS service does not require any special technical skills or computational resources from the end user. On the other hand, using a GraspOS tool requires technical expertise (to install/deploy the tool) and computational resources from the end user.

Apart from the aforementioned bundles of tools and services (i.e., enrichment and monitoring), the GraspOS infrastructure is also offering a bundle of "Data Services''. These refer to the services provided based on the APIs of the data sources that are included in FOMI (see Table 1 for the available APIs of the current GraspOS data sources). Since it was evident that these APIs were, essentially, a type of service, we decided to include them in the GraspOS Services Catalogue. It is worth mentioning that the Services Catalogue and the Data Registry are both keeping metadata for the respective Data Services which are complementary: the Services Catalogue captures service-level information, while the Data Registry offers practical information on accessing the respective data assets (e.g., the location, licensing, specification and restrictions of the respective data files and APIs. It is also important to mention that the

category of data services is pretty generic and the contained services can be heterogeneous. For instance, some of the data services can serve basic & essential needs (e.g., PIDs or basic metadata, like those provided by the OpenAIRE Graph), while others are building upon such basic information to calculate indicators (e.g., the BIP DB service). Proposing appropriate classification schemes for this type of services is something that will be considered seriously by GraspOS partners in the future.

3.2.3.   Resources Inclusion in FOMI

The first list of resources (i.e., data, tools, and services) to be included in the GraspOS infrastructure was created based on the competencies of the technical partners of the project. However, FOMI is planned to be an open infrastructure. This means that the list of data sources, tools, and services currently included in FOMI is subject to changes (mostly additions). 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.

Since, at this stage, the focus of the project is to establish the basic infrastructure and to make available the initial set of resources, the details about the inclusion policy will be refined and finalised in one of the next versions of the current deliverable. 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 (e.g., compliant to the API specification of the Data Interoperability & Access layer)
- offer the minimum QoS guarantees that have been 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[7]).

---

[7] https://eosc-portal.eu/eosc-providers-hub/what-are-eosc-profiles

3.3. Accessing the Infrastructure

Accessing the data assets, tools, and services included in FOMI can be done without the need for using a centralised "entry point", since the whole infrastructure is a distributed system of federated resources. Data assets and their APIs are hosted on the premises of the respective partner (or on Cloud servers funded by them) and any files for bulk downloads are usually hosted on open data repositories (e.g., Zenodo, Figshare). GraspOS tools are based on open source and publicly available codes (hosted, for instance, on GitHub). Finally, the APIs and the front-ends of the GraspOS services are hosted on the premises of the respective partner (or on Cloud servers funded by them). As a result, all data, tools, and services from the GraspOS infrastructure are, by principle, open to access.

Nevertheless, 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. The initial version of the front-end is already available[8] and offers basic browsing functionalities. These functionalities will be improved and the UI will be enhanced with more advanced search features in the next phase of the project.

3.4. EOSC Integration

GraspOS is an EOSC-related project and, as such, it aims at integrating its infrastructure with the European Open Science Cloud (EOSC).[9] The main effort in relevance to this integration will be focused on making all provided resources (data, tools, and services) available inside the EOSC Marketplace[10]. To this end, our design ensures that the GraspOS catalogues (i.e., the Service Catalogue and the Tools Catalogue) and registries (i.e., the Data Registry) are designed to be connected and synchronised with the EOSC Marketplace.

Regarding the Tools Catalogue, our plan is to exploit a Zenodo[11] community for this purpose. This means that each GraspOS tool should be registered (as a "Software" object) to this community. Zenodo is being harvested by OpenAIRE for the creation of the OpenAIRE Graph[12], which is an EOSC core service and, more importantly, the source of the research software entries in the EOSC Marketplace. Consequently, all GraspOS tools will be part of the

---

[8] GraspOS infrastructure entry point: https://graspos-infra.athenarc.gr/
[9] European Open Science Cloud (EOSC): https://eosc.eu/
[10] EOSC Marketplace: https://eosc-portal.eu/
[11] Zenodo: https://zenodo.org/
[12] The OpenAIRE Graph: https://graph.openaire.eu/

EOSC Marketplace by default following the previous approach. More details about the implementation of the GraspOS Tools Catalogue can be found in Section 4.

As regards the Services Catalogue, our plan is to build upon the OpenAIRE Catalogue[13] and register all GraspOS services to it. The OpenAIRE Catalogue is synchronised with the EOSC Marketplace, hence by registering GraspOS services to the former, these will also be available to the latter. More details about the implementation of the GraspOS Services Catalogue can be found in Section 4.

Finally, for the implementation of the Data Registry, we have selected to adapt and deploy an instance of the CKAN open-source data management system (details in Section 4.1). Although the GraspOS Data Registry is not directly connected to EOSC, each GraspOS data source is expected to also have an entry inside the Services Catalogue (see Section 3.2.2), hence GraspOS data sources will be available in the EOSC Marketplace based on this approach. In addition, the URL to the Data Service that corresponds to each GraspOS data source (i.e., the respective Services Catalogue URL) will be one of the required metadata fields in the Data Registry together with the respective Zenodo URL (if any).

4. Implementation

This section offers more implementation details for the various components of the GraspOS open and federated research assessment infrastructure (also known as "FOMI"). The respective descriptions correspond to the current, preliminary versions of the components, which are expected to be extended, improved, refined in the next period. Section 4.1 elaborates on the technologies used to develop the Data Registry. Section 4.2 presents the GraspOS Data Interoperability & Access Layer. Sections 4.3 and 4.4 present the implementation of the Tools and Services catalogue (respectively). Section 4.5 describes the implementation of the front-end user interface, which is the entry-point of the end user in the infrastructure. Finally, Section 4.6 contains details about the code base of the infrastructure components and the respective documentation.

4.1. Data Registry

This component allows users to be informed about and understand the data assets within the infrastructure. The Data Registry collects metadata and access information for all the onboarded data sources. More specifically, apart from a set of basic metadata related to each data source (e.g., its title, owners, licence), the GraspOS data registry also contains

---

[13] The OpenAIRE Catalogue: https://catalogue.openaire.eu/

information about the location of the respective GraspOS API deployment, enabling the programmatic access to the data source contents.

The data registry harnesses the power of CKAN,[14] an open-source data management system tailored for data portals that enables easy publishing, sharing and use of data. Our component builds upon the foundation of CKAN while integrating custom extensions tailored to the needs of the GraspOS project.

Specifically, one of the standout features of our component is the introduction of new fields to datasets, enriching the metadata and providing users with a comprehensive view of the stored information. By extending CKAN's functionality, we empower users to capture and display specific details that are crucial for our project's data context (e.g., the types of indicators provided by the respective data source). As it is evident, the metadata schema of the registry is based on an extension of CKAN's data model and will be refined and finalised during the next period based on feedback received by the assessment experts, the tools and services developers, and the pilot representatives.

4.2. Data Interoperability & Access Layer

FOMI contains a "Data Interoperability and Access Layer" which aims to facilitate the integration and harmonisation of (meta)data from various sources and formats within the infrastructure. This will ease the work of developers in implementing tools and added-value services on top of the respective data sources.

The Data Interoperability & Access Layer essentially comprises the union of all the API endpoints that offer access to the contents of the (federated) data sources included in the GraspOS infrastructure. During the lifetime of the project, the plan is for all data sources to offer API endpoints which are compatible (and interoperable) with each other. Towards this, GraspOS is building upon the guidelines and specifications of widely accepted standards (e.g., those provided by RDA[15] like the "Scientific Knowledge Graph Interoperability Framework"[16] delivered by the respective RDA working group) and the EOSC Interoperability Framework.[17]

The aforementioned specifications are expected to address all the needs of the project infrastructure, and GraspOS partners will need to extend them to accommodate such aspects that are not yet captured. For instance, in its present version, the SKG-IF does not offer details

---

about the exchange of information that is conducive for GraspOS tasks (e.g., indicators associated with research products). The proposed extensions, of course, will be suggested back to the respective initiatives (e.g., to RDA) for consideration and possible inclusion in a newer iteration of their specifications.

In the context of GraspOS, the data service providers will build upon the data models and specifications of the Data Interoperability & Access Layer to make the required changes from their side and deploy the required API endpoints so that their sources will be federated. Finally, being compliant to the specifications of the GraspOS Data Interoperability & Access Layer, will be one of the requirements for onboarding additional data sources to the GraspOS infrastructure in the future (see also Section 3.2.3).

4.3. Tools Catalogue

This component aims to facilitate the discovery of GraspOS tools for enrichment and monitoring by collecting their metadata offering also an integration with EOSC by automatically synchronising its contents with the EOSC Marketplace.

The implementation of the Tools catalogue will be based on a Zenodo[18] community, called "GraspOS Tools"[19]. Zenodo is a free and open digital archive, developed by CERN and OpenAIRE, that enables researchers to share and preserve any kind of research output, such as papers, data, software, reports, and more, supporting the principles of Open Science and open access. Zenodo communities are shared areas in Zenodo where users can collaborate, curate, and manage research outputs. The representatives of each community can review and accept submissions from other Zenodo users. They can also customise the community metadata and information (e.g., the logo, the description).

Based on this implementation, each GraspOS tool will be registered to this community creating a "Software" entry. Since Zenodo is being harvested by OpenAIRE for the creation of the OpenAIRE Graph[20], which is the basis for the research software entries in the EOSC Marketplace, all GraspOS tools will be discoverable through the EOSC Marketplace. Finally, the end user will leverage the advanced search and browsing capabilities already provided by Zenodo to its users. As it is evident, after identifying a tool of interest, the end user will need a computational system to run it and some basic technical skills to install/deploy it.

---

[18] Zenodo: https://zenodo.org/
[19] GraspOS Tools Catalogue: https://zenodo.org/communities/graspos-tools
[20] The OpenAIRE Graph: https://graph.openaire.eu/

4.4. Services Catalogue

The aim of this component is to facilitate the discovery of GraspOS services (for enrichment, monitoring, and data) by collecting and indexing their metadata. This catalogue is also expected to support their integration to EOSC by incorporating these metadata into the EOSC Marketplace.

The Services Catalogue provides GraspOS end-users with a centralised and intuitive platform to explore and access the various services offered. It is based on the OpenAIRE Catalogue open-source software.[21] The backend of the OpenAIRE Catalogue is implemented in Java using the Spring Boot framework, ensuring a scalable and efficient foundation that guarantees not only reliability but also swift responses to user queries and dynamic updates. Its frontend, on the other hand, is implemented in Angular and delivers a responsive, and dynamic user interface. The Angular framework facilitates the creation of a seamless and interactive user interface, enabling intuitive navigation and exploration of the available service. Users can effortlessly search, filter, and discover pertinent information about each service, enhancing their overall engagement.

During the development phase of the GraspOS Catalogue, a local deployment of the OpenAIRE Catalogue, hosted on an ATHENA's server in Athens (Greece), will be used. At the phase when the integration to EOSC will be needed, GraspOS services are expected to be incorporated in the OpenAIRE Catalogue service.[22] However, this approach will be revisited and re-evaluated later in the project considering (a) the changes or additions needed in the metadata schema for the GraspOS services, and (b) any challenges in including some GraspOS services into the existing service portfolios of the OpenAIRE Catalogue (e.g., the reluctance of the OpenAIRE administration team to include some of the services due to various reasons).

4.5. 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 infrastructure[23], but also to explore the resources (data, tools, and services) hosted and

---

[21] The OpenAIRE Catalogue: https://github.com/madgeek-arc/resource-catalogue-ui-openaire &
https://github.com/madgeek-arc/openaire-catalogue
[22] The OpenAIRE Catalogue: https://catalogue.openaire.eu/
[23] GraspOS Infrastructure front-end UI: https://graspos-infra.athenarc.gr/

exposed by the GraspOS infrastructure. This UI is essentially the entry point to the infrastructure.

The development of this component is based on Docusaurus,[24] a framework that allows building and deploying customised and highly optimised project websites. It leverages Markdown to add the website's content, thus simplifying the authoring process and enabling content writers lacking technical background to focus on delivering high-quality, informative content without a steep learning curve.

Meanwhile, this front-end UI is highly adaptable and responsive, a direct result of Docusaurus's built-in support for mobile devices. It offers a consistent and user-friendly experience across a wide range of devices, ensuring that the content scales seamlessly to fit different screen sizes. It also allows dynamic code to be included, hence it supports implementing advanced features (for instance, to support advanced navigation functionalities). Last but not least, its built-in support for versioning and several deployment additions/plugins have streamlined our website's maintenance process, allowing us to effortlessly update content and deploy changes promptly.

4.6. Code Repositories & Documentation

In this section, we present the code bases of all important components of the FOMI architecture, as well as references to the respective documentation (where applicable). More specifically, in Section 4.6.1, we present the information related to the core components of the infrastructure (i.e., the main catalogues, repositories, and front-ends). In Sections 4.6.2 and 4.6.3, we present the same information for the enrichment and the monitoring tools and services (respectively). Finally, in Section 4.6.4, we present more information on the provision workflows of the GraspOS data sources (and the respective documentation). It is worth mentioning that more details on the tools & services (e.g., detailed functionalities) will follow in the deliverable D3.2 "Tools and Services" (to be submitted in August 2024).

4.6.1. Core Infrastructure Components

The following table lists the code repositories, the current deployments, and the documentation websites of all the core infrastructure components.

---

[24] Docusaurus: https://github.com/facebook/docusaurus

Table 2 *Core infrastructure components overview*

Component	Code repository	Current deployment	Documentation
Infrastructure front-end UI	https://github.com/athenarc/graspos-infra	https://graspos-infra.athenarc.gr/	https://github.com/athenarc/graspos-infra/blob/master/README.md
Data registry	https://github.com/athenarc/graspos-registry	https://graspos-data.athenarc.gr/	https://github.com/athenarc/graspos-registry/blob/main/README.md
Services catalogue	Front-end: https://github.com/athenarc/graspos-services-catalogue-frontend Back-end: https://github.com/athenarc/graspos-services-catalogue-backend	https://graspos-services.athenarc.gr/home	https://github.com/athenarc/graspos-services-catalogue-frontend/blob/master/README.md
Tools catalogue	N/A	https://zenodo.org/communities/graspos-tools	N/A

Regarding the Tools Catalogue, since we have selected to implement it as a Zenodo community, no development of source code and no documentation was required.

Finally, it is worth mentioning that, although the "Data interoperability & access layer", is an important component of the GraspOS infrastructure, it is implemented as the union of the API endpoints of the various data services, hence no separate code repository is provided for it (see also Section 4.2).

4.6.2.   Enrichment Tools & Services

The following table lists the code repositories, the current deployments, and the documentation websites of all the enrichment tools & services.

Table 3 Enrichment tools & services overview

Component	Code repository	Current deployment	Documentation
BIP! Citation Classifier	https://github.com/athenarc/bip-ndr-workflow	N/A	https://github.com/athenarc/bip-ndr-workflow#readme
BIP! Ranker	https://github.com/athenarc/Bip-Ranker	N/A (this is a tool)	https://github.com/athenarc/Bip-Ranker#readme
OpenAIRE Broker	https://code-repo.d4science.org/D-Net/dnet-hadoop https://code-repo.d4science.org/D-Net/dnet-applications https://github.com/openaire/broker-cmdline-client	https://provide.openaire.eu (part of the OpenAIRE PROVIDE)	https://graph.openaire.eu/develop/broker.html https://api.openaire.eu/broker/swagger-ui/index.html
OpenAIRE Text Mining	https://github.com/openaire/iis	https://mining.openaire.eu/openaireplus/analyze	https://graph.openaire.eu/docs/graph-production-workflow/enrichment-by-mining/ https://github.com/openaire/iis#readme
OpenAIRE Metadata Validator	https://code-repo.d4science.org/MaDgIK/metadata-validator-ui https://code-repo.d4science.org/MaDgIK/uoa-validator-api https://code-repo.d4science.org/MaDgIK/validator-engine	http://duffy.di.uoa.gr:5100/	Not available yet It is a tool built and based on the OpenAIRE Guidelines http://guidelines.openaire.eu/en/latest/

SCRE Pipeline	N/A	N/A	From GoTriple Ontologies: https://www.gotriple.eu/ontology/triple
Semantic Citation Classifier	https://github.com/opencitations/cec	N/A	https://github.com/opencitations/cec/blob/main/README.md

It is worth mentioning that, in addition to the previous tools and services, there is also a collection of enrichment tools and services with similar or complementary functionalities, which have been developed (or simply configured/adapted) by some of the GraspOS pilot partners and they are currently being used by them to enrich their local databases. Based on the interaction with these partners in the context of the respective piloting activities, some of these tools and services may need to be included in the bundle (and in the GraspOS Tools & Services Catalogues). Finally, it is worth mentioning that some of the tools and services are still in their alpha versions and the respective code repositories and/or documentation websites are not publicly available right now, they will become available during the next months.

4.6.3.   Monitoring Tools & Services

The following table lists the code repositories, the current deployments, and the documentation websites of all the monitoring tools & services.

Table 4 Monitoring tools & services overview

Component	Code repository	Current deployment	Documentation
BIP! Scholar	Implementation of indicators: https://github.com/athenarc/bip-scholar-indicators The front-end code will be released as open source soon	https://bip.imsi.athenarc.gr/scholar	N/A
EOSC Services Accounting	N/A	https://accounting.eosc-portal.eu/	https://argoeu.github.io/argo-accounting/

OpenAIRE Monitor	https://code-repo.d4science.org/MaDgIK/monitor https://code-repo.d4science.org/MaDgIK/monitor-dashboard	https://monitor.openaire.eu/	https://monitor.openaire.eu/about https://monitor.openaire.eu/methodology/methodological-approach https://monitor.openaire.eu/methodology/terminology https://monitor.openaire.eu/indicators/themes https://catalogue.openaire.eu/service/openaire.funder_dashboard/overview
OpenAIRE Researcher Profile	N/A	N/A	N/A
Open Science Observatory	N/A	https://osobservatory.openaire.eu/home	N/A
EOSC Observatory	N/A	https://eoscobservatory.eosc-portal.eu/home	N/A
OPERAS Metrics	N/A	https://metrics.operas-eu.org/	N/A

Similarly to the enrichment tools and services (Section 4.6.2), in addition to the tools and services listed in the previous table, there is also a collection of tools and services with similar or complementary functionalities, which have been developed (or simply configured/adapted) by some of the GraspOS pilot partners (e.g., the research.fi[25] service). Based on the interaction with these partners in the context of the respective piloting activities, some of these tools and services may need to be included in the bundle (and in the GraspOS Tools & Services Catalogues). Finally, some of the tools and services are still in their alpha versions and the respective code repositories and/or documentation websites are not publicly available right now, they will become available during the next months.

---

[25] Research.fi: https://research.fi/en

4.6.4. Data Services

The following table contains the repositories that contain the code for the production of the respective data assets and the respective documentation websites.

Table 5 *Data services overview*

Service	Data production workflow	Documentation
BIP! DB	https://github.com/athenarc/Bip-Ranker	https://github.com/athenarc/Bip-Ranker/blob/main/README.md
BIP! NDR	https://github.com/athenarc/bip-ndr-workflow	https://github.com/athenarc/bip-ndr-workflow/blob/main/README.md
OpenAIRE Graph	https://code-repo.d4science.org/D-Net/dnet-hadoop	https://graph.openaire.eu/docs/
OpenCitations data	https://github.com/opencitations/index and https://github.com/opencitations/oc_meta	https://github.com/opencitations/index/blob/master/README.md and https://github.com/opencitations/oc_meta/blob/master/README.md
OPERAS Metrics	N/A	N/A
ScholExplorer	N/A	https://api.scholexplorer.openaire.eu/v2/ui/
Usage Counts	N/A	https://graph.openaire.eu/docs/graph-production-workflow/indicators-ingestion/usage-counts
EOSC services accounting	https://argoeu.github.io/argo-accounting/	https://argoeu.github.io/argo-accounting/
Assessment Registry	N/A	N/A

5. Conclusions

In this report, we have presented implementation details regarding FOMI, the federated infrastructure to support the activities of the GraspOS project. FOMI is composed of a set of

federated Data Asset Sources providing access to and delivering research products metadata, pre-calculated research performance indicators, as well as unstructured data such as scientific texts, narrative CVs and data on OS practices.

These sources are flanked with a set of federated monitoring and enrichment services, which can be used to handle the transformation, manipulation, and analysis of data within the infrastructure, and present and visualise research assessment evidence and indicators at different levels of granularity (e.g., researcher, organisation, country) to support OS-aware RRA and to monitor the uptake and impact of OS.

The backbone of the infrastructure consists of a Data Registry documenting the federated data sources and the contained assets, a data interoperability layer that enables the easy consumption of the contained data assets for programmers, two catalogues facilitating the exploration of the tools and services which are provided by the infrastructure, and a front-end which acts as the entry point of the end-users in the infrastructure.