DISCOVER Services: Final Report on VA Activities

JUNE 2023

Scholarly Communication Services for EOSC users

D6. 3 – DISCOVER Services: Final Report on VA Activities

Version 1.0 – Final

PUBLIC

Report on the DISCOVER installations operation activity after the last reporting period, including a detailed set of indicators.

H2020-INFRAEOSC-2020-2
Grant Agreement 101017452

Document Description

D6.3 – DISCOVER Services: Final Report on VA Activities

WP6 - Virtual Access to DISCOVER portfolio	
WP participating organizations: ARC, CNR, UNIWARSAW	
Contractual Delivery Date: 06/2023	Actual Delivery Date: 06/2023
Nature: Report	Version: 1.0 Final
Public Deliverable	

Preparation Slip

	Name	Organisation	Date
From	Marek Horst	UNIWARSAW	15/06/2023
Edited by	Marek Horst	UNIBO UNIWARSAW	21/06/2023 22/06/2023
Reviewed by	Alessia Bardi Alex Ioannidis Andreas Czerniak	CNR CERN UNIBI	22/06/2023 23/06/2023 26/06/2023
Approved by	Paolo Manghi	OPENAIRE AMKE	28/06/2023
For delivery	Mike Hatzopoulos	OPENAIRE AMKE	10/07/2023

Revision History

Issue	Item	Reason for Change	Author	Organization
V0.1	Draft version	The first version of deliverable.	Marek Horst	UNIWARSAW
V0.2	Draft version	Accepting an internal review outcome.	Marek Horst	UNIWARSAW
V1.0	Final version	Minor updates related to dates and versions.	Marek Horst	UNIWARSAW

Contents

Disclaimer

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

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

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

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the OpenAIRE-Nexus consortium and can in no way be taken to reflect the views of the European Union.

OpenAIRE-Nexus is a project funded by the European Union (Grant Agreement No 101017452).

Acronyms

DOA	**Description of Action**
KPI	**Key Performance Indicator**
UoA	**Unit of Access**
VA	**Virtual Access**

Publishable Summary

This document is the final report on the DISCOVER installations operation activity after 30 months of operation, including a detailed set of indicators.

The following services were considered within the report:

- OpenAIRE EXPLORE
- OpenAIRE CONNECT
- OpenAIRE PROVIDE

The report is split into several sections covering services description, definition of KPIs, acquired KPI values, rationale for choosing Virtual Access indicators, acquisition methodology and measurement time period coverage. Any deviations regarding the KPIs definitions, their coverage and potential problems with reaching the success target were described in detail in the “Deviations of VA tables and KPIs” section. The "VA assessment report" section summarises an outcome of the work conducted by the *External Advisory Board for the VA assessment*.

1. Service Portfolio Description

DISCOVER portfolio is about a bundle of unique services that enriches the scholarly record and provides contextual and smart discovery of research output.

2. Definition of VA Activities

As written in the DOA, WP6, operates Virtual Access (VA) to all service installations in the DISCOVER portfolio, to ensure:

- onboarding to OpenAIRE service catalogue[1] and EOSC service catalogue[2],
- quality of service,
- support to EOSC users, and
- collection of usage indicators and user accounting required to extrapolate the total quantity of access to the installations provided to users beyond their usual communities.

Virtual Access will be provided to the following installations and related reimbursement schemes:

- **EXPLORE**, deployed at ICM, unit of access "number of requests", reimbursement by Actual Cost.
- **CONNECT**, deployed at ICM, unit of access "number of discovery Gateways prepared for Research Communities", reimbursement by Unit Cost.
- **PROVIDE**, deployed at ICM, unit of access "data source", reimbursement by Unit Cost.

3. Description of Services

The DISCOVER portfolio includes the following four services:

Service name: OpenAIRE EXPLORE

Description: a suite of services to provide discovery and claim functionalities, to search, navigate and enhance the OpenAIRE Research Graph.

Website: https://explore.openaire.eu/, https://graph.openaire.eu/develop/

Service name: OpenAIRE CONNECT

[1] http://catalogue.openaire.eu/search;quantity=10
[2] https://marketplace.eosc-portal.eu/

Description: a service that provides discovery Gateways for Research Communities.

Website: **https://connect.openaire.eu/**

Service name: OpenAIRE PROVIDE

Description: a registration service for data sources to join OpenAIRE, that facilitates the process of alignment, completeness and improvement of quality of the metadata they expose.

Website: **https://provide.openaire.eu/**

4. DEFINITION OF KPIS

Service name: OpenAIRE EXPLORE

KPI name	Definition
#APIrequests (Unit of Access)	Number of requests handled by publicly available OpenAIRE EXPLORE APIs.
#UserVisits	Number of unique visits registered on https://explore.openaire.eu/ site
#Uptime	Service availability, expressed as percentage value

Service name: OpenAIRE CONNECT

KPI name	Definition
#CommunityGateways (Unit of access)	Number of discovery Gateways prepared for Research Communities
#ActiveUsers	Number of all Research Communities Gateways active users
#Uptime	Service availability, expressed as percentage value

Service name: OpenAIRE PROVIDE

KPI name	Definition
#DataSources (Unit of access)	Number of data sources covering institutional repositories only
#Repositories	Total number of repositories registered in OpenAIRE PROVIDE
#BrokeringEvents	Total number of brokering events which are expected to be consumed by subscribers
#Uptime	Service availability, expressed as percentage value

5. KPIs PER SERVICE

Acquired values for indicators

This section presents all the indicator values acquired at M12, M24 and M30. The detailed description on how indicators were gathered is available in the "Indicators acquisition methodology" section. The description of the table columns is available below the KPIs table.

Service	KPI	Point of reference (Basis)	Success Target (% inc)	Initial value	Success Target value (M30)	M12 value	M24 value	M30 value
EXPLORE	**#APIrequests (UoA)**	25,000,000	120%	0	**30,000,000**	38,819,224	85,817,355	**106,915,736**
EXPLORE	**#UserVisits**	411,141	40%	0	**164,456**	474,841	1,428,937	**1,918,844**
EXPLORE	**#Uptime**				**99,98%**	99.45%	100%	**100%**
CONNECT	**#CommunityGateways (UoA)**	7	214%	0	**15**	6	8	**17**
CONNECT	**#ActiveUsers**	165	50%	0	**83**	630	773	**869**
CONNECT	**#Uptime**				**99,98%**	100%	100%	**100%**
PROVIDE	**#DataSources (UoA)**	150	153%	0	**230**	85	182	**231**
PROVIDE	**#Repositories**	189	153%	0	**289**	122	~~252~~ 279[3]	**351**
PROVIDE	**#BrokeringEvents**	29,000,000	40%	8,481,089	**20,081,089**	24,803,631	13,921,850	**14,668,384**[4]
PROVIDE	**#Uptime**				**99,98%**	100%	100%	**100%**

Column description:

Service	Service name
KPI	Key Performance Indicator, marked with "UoA" if Unit of Access
Point of reference (Basis)	Base value as of 2019 provided in the Grant Agreement document, point of reference when applying percentage increase as success target
#Uptime	Service availability, expressed as percentage value

[3] #Repositories M24 value is updated because in D6.2 report it was calculated using deprecated methodology which was not aligned with #DataSources (UoA) methodology update described in the D6.2 "Indicators acquisition methodology" section.

[4] #BrokeringEvents KPI is not cumulative but represents a state within the brokering system. This is why the M24 and M30 values can be lower than the M12 value. More details on #BrokeringEvents KPI specifics are available in the "Indicators acquisition methodology" section.

Success Target (% increase)	Success target, expressed as percentage and declared in the Grant Agreement document, to be applied on the point of reference in order to calculate success target value at M30
Initial value	Initial tracked value obtained on 2021-01-01 when the OpenAIRE-Nexus project started. Required to calculate success target value.
Success Target value (M30)	Success target value at M30. The value is either explicitly defined in the Grant Agreement document or calculated with the following formula: Initial value + (point of reference * success target)
M12 value	KPI value readout at M12 (2021-12-31)
M24 value	KPI value readout at M24 (2022-12-31)
M30 value	KPI value readout at M30 (usually 2023-06-16, numbers were gathered before the deliverable submission deadline set to 2023-06-30)

The rationale for choosing VA indicators

This section provides explanations on the VA KPIs of each service of the DISCOVER portfolio.

Service name: OpenAIRE EXPLORE

The unit of access is the number of requests handled by publicly available OpenAIRE Explore APIs starting from January 2021 as it clearly reflects an uptake of OpenAIRE EXPLORE service endpoints.

Service name: OpenAIRE CONNECT

The unit of access is the number of discovery gateways prepared for Research Communities which proportionally captures the effort required to ensure the specific portion of the OpenAIRE Graph related to the requesting EOSC user (community manager) can be identified, kept up-to-date, and finally delivered to gateway users.

Service name: OpenAIRE PROVIDE

The unit of access for OpenAIRE PROVIDE is the data source, which proportionally captures the average effort required to validate its content, produce the related notification events, provide support and training to the data source managers.

Indicators acquisition methodology

This section covers detailed acquisition methodology of all Virtual Access KPIs.

The OpenAIRE infrastructure offers a dedicated service to keep track of customised metrics over time. The metrics service is based on Prometheus (https://prometheus.io/) and Grafana (https://grafana.com/). Most of the OpenAIRE services in the DISCOVER portfolio exploit the

metrics service so that the collection of KPI values is an automated process. The type of acquisition process (manual vs automatic with Prometheus) is specified in the "other remarks" paragraph of each service.

OpenAIRE EXPLORE

KPI: #APIrequests (Unit of Access)

acquisition methodology: Statistics related to API hits handled by all the OpenAIRE Explore API endpoints (covering Search, Bulk Access, Broker, LOD) are tracked by Matomo (OpenAIRE analytics platform) as pageviews which are regularly retrieved from Matomo and persisted in metrics acquisition subsystem as VA indicators for automated reporting and further analysis.

KPI: #UserVisits

acquisition methodology: OpenAIRE-Explore website was integrated with Matomo tracking tool, so the web visits metric is translated from unique pageviews provided by analytics.openaire.eu and stored in the VA metrics acquisition system.

other remarks: 2019 base value was not explicitly denoted in the Grant Agreement document, but it was provided by OpenAIRE analytics platform (Matomo).

OpenAIRE CONNECT

KPI: #CommunityGateways (Unit of Access)

acquisition methodology: Covers EC-funded gateways which were deployed after 2021.01.01 (OpenAIRE-Nexus project start) whose visibility is not defined as "hidden". Since M15, the eligibility rules were slightly refined by covering officially requested dashboards only. This additional condition was imposed in order to have a strong proof of demand for a dashboard creation.

other remarks: Acquisition procedure is partially automated: the list of currently available Research Community Gateways is generated automatically, but individual dashboards from this list are intended to be approved by the product owner mainly due to two reasons:

- unambiguous "restricted" status which can indicate either the dashboard is not ready yet or is ready but restricted on purpose by the community manager
- to exclude gateways that have been funded in the context of other projects that ran in parallel to OpenAIRE-Nexus (e.g. the project NEANIAS - Novel EOSC services for Emerging Atmosphere, Underwater and Space Challenges - funded three community gateways).

KPI: #ActiveUsers

acquisition methodology: Total number of active gateway subscriptions is tracked by AAI and exposed via the Prometheus exporter endpoint.

other remarks: 2019 base value was not indicated explicitly in the Grant Agreement document so the only way to obtain this value was to refer to the OpenAIRE Advance KPIs internal document (*# of researchers subscribed to RCDs* in particular, covering the state as of November 2019).

OpenAIRE PROVIDE

KPI: #DataSources (Unit of Access)

acquisition methodology: The total number of data sources is calculated based on the number of institutional repositories registered in OpenAIRE and available since the start of the OpenAIRE-Nexus project. Technically, the repository is considered as eligible only if it was collected for the first time after the OpenAIRE-Nexus project started[5].

other remarks: The only difference between this indicator and #Repositories is that the Unit of Access indicator is restricted to institutional repositories only.

KPI: #Repositories

acquisition methodology: This indicator represents the number of all repositories registered in OpenAIRE and available since the start of the OpenAIRE-Nexus project. Technically the repository is considered as eligible only if it was collected after the OpenAIRE-Nexus project started.

other remarks: 2019 base value was not explicitly indicated in the Grant Agreement document but it was possible to count the number of repositories registered in 2019 by relying on an API endpoint serving metrics for metrics acquisition system.

The only difference between this indicator and the #DataSources (Unit of Access) indicator is that the Unit of Access indicator is restricted to institutional repositories only.

KPI: #BrokeringEvents

acquisition methodology: Broker service performs metadata brokering functionality by issuing events to exchange (via subscription & notification) metadata information enrichments with other sources and ensure up-to-dateness and completeness (e.g. new citations to products, new OA version, missing DOI, etc). This indicator covers the total number of such events which are expected to be consumed by subscribers.

Broker was implemented in compliance with subscriber/consumer paradigm and the number of brokering events is covered by the number of notifications. This is the number of notified events i.e. the potential events that have matched the subscriptions. A potential event can produce 0..n notifications depending on the number of the matched subscriptions.

[5] In the first VA report (D6.1) the restriction was expressed in a slightly different way and evolved around registration date. This restriction turned out to be excluding some eligible data sources which did not have registration date properly specified or which were registered before OpenAIRE-Nexus project started but were collected for the first time and entered the OpenAIRE system during the OpenAIRE-Nexus project. This remark is applicable to both *#DataSources* and *#Repositories* KPIs.

other remarks: The original #BrokeringEvents KPI name indicates the total number of brokering events and is bound to 29Mi of brokering events generated in 2019 as declared in 2019 highlights for OpenAIRE PROVIDE installation in the Grant Agreement document.

Since Broker subscription/notification mechanism was heavily refactored in 2020/2021 we cannot directly relate current readouts (over 300M of events in total) to the 2019 base number anymore. One rather relevant metric we could focus on are the notifications, which are events with at least one active subscription (over 20M). Those notifications are events that repositories are notified about and are expected to be consumed by those repositories from the public broker API.

New Broker implementation was deployed in production in August 2021 therefore it is rather difficult to relate to the initial value at the beginning of the project when the older broker implementation was still in operation. In January 2021 the number of events for which at least one active subscription was matched was 8,481,089. The first metric readouts from the new Broker implementation were gathered in October 2021 when the number of relevant events was 24,927,942.

Time period coverage

Section below describes the metrics acquisition time period. Some indicators were not tracked since the beginning of the OpenAIRE-Nexus project because the technical solutions for an automated acquisition of those indicators were introduced after the project started.

Service name: OpenAIRE EXPLORE

KPI name	From	To
#APIrequests (Unit of Access)	**01-01-2021**	**16-06-2023**
#UserVisits	**01-01-2021**	**16-06-2023**
#Uptime	**01-01-2021**	**31-05-2023**

Service name: OpenAIRE CONNECT

KPI name	From	To
#CommunityGateways (Unit of access)	**01-01-2021**	**22-06-2023**
#ActiveUsers	**01-01-2021**	**16-06-2023**
#Uptime	**01-01-2021**	**31-05-2023**

Service name: OpenAIRE PROVIDE

KPI name	From	To
#DataSources (Unit of access)	01-01-2021	23-06-2023
#Repositories	01-01-2021	16-06-2023
#BrokeringEvents	01-10-2021	16-06-2023

6. DEVIATIONS OF VA TABLES AND KPIS

Broker Service, which is a part of the OpenAIRE PROVIDE portfolio, evolved since the OpenAIRE-Nexus project submission in 2019 and was significantly refactored in 2020/2021. This fact has quite significant impact on the #BrokeringEvents (non Unit of Access) KPI current readouts and this KPI reference to its 2019 base value declared in the Grant Agreement document and the initial value at the beginning of the project. This was described in detail in the "Indicators acquisition methodology" section.

#BrokeringEvents KPI for OpenAIRE PROVIDE could not be measured since the beginning of OpenAIRE-Nexus project because it was covered with OpenAIRE measurement software after the Broker Service refactoring took place (mid 2021).

7. VA ASSESSMENT REPORT

1st VA External Board for VA assessment meeting outcome

The OpenAIRE-Nexus DISCOVER installations were assessed in the context of Virtual Access by a board composed of international experts.

The first assessment meeting took place on 29th of November, 2022 with the following External Board members participating in the meeting:

- Andrea Manzi(EGI-ACE - EGI.eu)
- Enol Fernández (C-SCALE - EGI.eu)
- Marek Horst (OpenAIRE-Nexus - ICM UW)
- Antti Pursula (DICE - EUDAT / CSC)
- Tomasz Piontek (RELIANCE - PSNC)

The following OpenAIRE-Nexus project related deliverables:

- D1.2 - Key Performance Indicators and Analysis report
- D4.1 - PUBLISH Services: Periodic Report on VA Activities I
- D4.2 – PUBLISH Services: Periodic Report on VA Activities II
- D5.1 - MONITOR Services: Periodic Report on VA Activities I
- D5.2 - MONITOR Services: Periodic Report on VA Activities II
- D6.1 - DISCOVER Services: Periodic Report on VA Activities I
- D6.2 - DISCOVER Services: Periodic Report on VA Activities II

were shared with the VA Assessment Board one week before the VA assessment meeting.

The Virtual Access related presentation was made by Marek Horst, OpenAIRE-Nexus Virtual Access Manager. The following aspects were presented and discussed during the meeting:

- Detailed description of the Unit of access for each installation
- Description of the gathered VA indicators
    - Making sure that double accounting is avoided
- An effort to enable the automatic collection of VA metrics
    - Demo of the Dashboard based on Grafana
    - When VA Automatic Collection is not available the metrics are pushed to github using CSV files

    - MoUs as an additional condition for dashboard-focused UoA eligibility
- Logs for user access to services are kept for auditing
- Integration with EOSC VA Accounting

Additionally the following topics were raised and discussed:

- Missing or miscalculated point of reference (baseline) cases
- Retention of logs and GDPR, logs anonymization

Final assessment

All the installations under VA are offered to all the users, without selecting the researchers to whom access is provided, and free of charge what fulfils the major VA requirement. In general, the VA Assessment Board finds the implementation and operation of the VA in the OpenAIRE-Nexus project to be well managed and in line with requirements.

2nd VA External Board for VA assessment meeting outcome

The second assessment meeting took place on 1st of June, 2023 with the following External Board members and project representatives participating in the meeting:

- Charis Chatzikyriakou (C-SCALE - EODC)
- Enol Fernández (C-SCALE - EGI.eu)
- Marek Horst (OpenAIRE-Nexus - ICM UW)
- Andrea Manzi (EGI-ACE - EGI.eu)
- Raul Palma (RELIANCE - PSNC)
- Tomasz Piontek (RELIANCE - PSNC)
- Antti Pursula (DICE - EUDAT / CSC)
- Debora Testi (DICE - CINECA)

During the meeting, all the project representatives shared the gained experience in the application of the VA methods and principles for the provisioning of the different offered services. A joint report on advantages and challenges related to the VA funding model was prepared and is available as Appendix A.

8. APPENDIX A

Virtual Access assessment report

The External Advisory Board for the VA assessment of the INFRAEOSC-07 projects has been established in line with the GA requirements and in agreement with the EC and is composed by one representative from each INFRAEOSC-07 project. After a first assessment, with the aim to check and discuss and provide recommendations on the VA practices and methodologies of each project (held in December 2022), a second assessment was carried out in June 2023.

The objective of the second assessment has been to share gained experiences on the application of the VA methods and principles in the provisioning of the different offered services and to define common lessons learnt.

A plenary meeting took place online on the 1st of June 2023. During the meeting, each INFRAEOSC-07 project representative has presented the advantages and challenges in the application of the VA rules and principles to the specific service offering, and commonalities/differences were discussed.

Common lessons learnt have then been agreed and are summarised as follows.

First of all, the VA method is valued as an instrument to allow the service providers to recover the costs of used resources and services. The VA method also allows to provide the services free-at-the-point of use, even in the case of usually pay per use ones, offering the possibility to users and communities to test and start the take up of the services at no costs thus reducing the adoption barrier. It is also acknowledged that it was important to have both investment and operational costs included as part of the units cost calculation, even if not all operational costs can be claimed by all providers depending on the specific site accounting rules.

However, being the VA method new in its application for some categories of services and providers part of the INFRAEOSC-07 projects, some challenges have been identified. Overcoming these issues in future EC funded projects would make the instrument even more important and effective.

Only in few cases the VA costs were based on actual costs method only, while in most of the cases unit costs or a combination of unit and actual costs was used. The right choice of unit is thus vital to allow a meaningful representation of the services usage, but in many cases the proper unit type was difficult to define at the start of the project. At the same time, the unit costs are defined at the proposal writing phase and during the projects' execution years, many costs might change (the high increase in electricity costs of the last year is just one example). However, in the current implementation of the VA model, updates in the type and cost of units imply the need to request an amendment to the GA, which is making the process time consuming. This lack of flexibility in the VA budget management makes it also difficult to adjust the VA offering to changes in the needs from the users' communities during the projects, for example by updating the quantity of

units being offered, transferring units from one installation to another, removing or adding new installations.

As already mentioned, the INFRAEOSC-07 projects collect together in their offering very diverse types of services, with different units/metrics and offered by different providers. This led to the impossibility to have a generic and automatic acquisition of the accounting data which implied in many cases to set-up manual collection of the VA accounting data; as always in manual data collection, this is time consuming and prone to errors. At the same time, the recent creation of the possibility for EOSC users to request bundles of services (from different providers but also across different projects) raised the issue on how to properly account those requests (as individual requests vs aggregated information). Another critical element related to the VA accounting is the GA requirement to maintain information needed for auditing for at least 5 years after the project ends; while logs and metrics are collected by each project, it is felt that the procedures and requirements for auditing purposes are not currently sufficiently clear (which level of details of information/logs to maintain and at which level (project level vs provider level)).

With respect to the users/communities being served, it became evident that commitment from large communities is difficult to achieve due to the short term duration of the projects and support to the VA methods. The short term VA supported offering is also perceived as discouraging for individual researchers and small groups who will not be able to pay for services via their institutions. This issue becomes particularly relevant for some categories of services like those related to data storage where the expected preservation is normally in the range of 5 to 10 years or even more. It should be also noted that having a lot of small users instead of large communities uptaking the service implies to provide much more time for users' support and engagement which might impact on the budget actual costs of the VA installations.

In summary, the INFRAEOSC-07 projects recommendations are:

- The actual cost model is in general easier to account and claim with respect to unit costs;
- More flexible procedures to update the VA offering would be advisable;
- Clearer information on the VA auditing process and requirements would be useful;
- Longer term support to the VA mechanism would reduce the barrier to adoption and make the researchers and communities more keen to update the offered services.