PUBLISH Services: Final Report on VA Activities

JUNE 2023

Scholarly Communication Services for EOSC users

D4.3 – PUBLISH Services: Final Report on VA Activities

Version 1.0 –Final
PUBLIC

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

H2020-INFRAEOSC-2020-2
Grant Agreement 101017452

Document Description

D4.3 – PUBLISH Services: Final Report on VA Activities

WP4 - Virtual Access to PUBLISH portfolio	
WP participating organizations: CERN, ARC, UNIWARSAW, CITE, CNRS	
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	Raphaël Tournoy Marek Horst	CNRS UNIWARSAW	19/06/2023 22/06/2023
Reviewed by	Alex Ioannidis Andreas Czerniak Nicolas Liampotis	CERN UNIBI GRNET	23/06/2023 27/06/2023 28/06/2023
Approved by	Paolo Manghi	OpenAIRE AMKE	28/06/2023
For delivery	Eleni Koulocheri	OpenAIRE AMKE	28/06/2023

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

DMP	Data Management Plan
DOA	Description of Action
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
KPI	Key Performance Indicator
OA	Open Access
UoA	Unit of Access
VA	Virtual Access
TBs	Terabytes

Publishable Summary

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

The following services were considered within the report:

- Amnesia
- Argos
- Episciences
- Zenodo

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. SERVICES PORTFOLIO DESCRIPTION

PUBLISH portfolio is about a bundle of unique services that allow researchers to publish FAIR and GDPR compliant research outcomes using open science practices.

2. DEFINITION OF VA ACTIVITIES

As written in the DOA, WP4, operates Virtual Access (VA) to all service installations in the PUBLISH 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:

- **Amnesia**, deployed at ICM (Poland), unit of access "user visits", reimbursement by Actual Cost.
- **ARGOS**, deployed at ICM, unit of access "managed projects", reimbursement by Actual Cost.
- **Episciences**, deployed at CNRS (France), unit of access "submitted articles", reimbursement by Unit Cost.
- **Zenodo**, deployed at CERN (Switzerland), unit of access "Terabytes (TBs) of data transferred", reimbursement by Unit Cost.

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

3. DESCRIPTION OF SERVICES

The PUBLISH portfolio includes the following four services:

Service name: Amnesia

Description: a service to anonymize sensitive research data (GDPR compliance), ready to be embedded in institutional workflows, to remove barriers and facilitate FAIRness of data.

Website: https://amnesia.openaire.eu/

Service name: ARGOS

Description: a service for machine-actionable[3] Data Management Plans, guiding researchers towards FAIR, configurable to domain discipline knowledge.

Website: https://argos.openaire.eu/splash/

Service name: EpiSciences

Description: a pan-European overlay journal platform, operating on top of OA repositories (e.g. HAL, Zenodo, arXiv), where communities can create and operate high-quality OA journals.

Website: https://www.episciences.org/

Service name: Zenodo

Description: a catch-all repository hosted by CERN (Switzerland), which researchers, communities or Research Performing Organizations, and citizens can use for FAIR sharing and long-term preserving research results.

Website: https://zenodo.org/

[3] This term refers to information that is structured in a consistent way so that machines, or computers, can be programmed against the structure. More details on principles for machine-actionable DMPs : https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006750

4. DEFINITION OF KPIS

Service name: Amnesia

KPI name	Definition
#UserVisits	Number of user visits
#DownloadsOfDesktopTool	Number of downloads of the Amnesia desktop application including various platforms.
#Uptime	Service availability, expressed as percentage value

Service name: ARGOS

KPI name	Definition
#ProjectsCreated (Unit of access)	Defined on the basis of "per managed project".
#DMPtemplates	Number of DMP templates
#LanguagesSupported	Number of languages supported
#Uptime	Service availability, expressed as percentage value

Service name: EpiSciences

KPI name	Definition
#Articles (Unit of access)	The number of article submissions performed by the users registered after 2021-01-01
#Journals	Number of journals
#Users	Number of users
#Uptime	Service availability, expressed as percentage value

Service name: Zenodo

KPI name	Definition
#TBsDataTransferred (Unit of access)	Measures the bytes uploaded and downloaded to/from Zenodo by users
#ActiveUsers	Number of active researchers registered in Zenodo
#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)	Declared Success Target (% increase)	Initial KPI value (as of Jan 1st 2021)	Success Target value (M30)	M12 value	M24 value	M30 value
Amnesia	**#UserVisits (UoA)**	12,078	77%	0	**9,300**	25,699	47,584	**59,526**
Amnesia	**#DownloadsOfDesktopTool (UoA)**	400	30%	0	**120**	993	1,900	**2,434**
Amnesia	**#Uptime**				**99.98%**	100%	98.41%	**98,68%**
Argos	**#ProjectsCreated (UoA)**	3	1667%	0	**50**	27	70	**96**
Argos	**#DMPtemplates**	1	500%	0	**5**	4	10	**10**
Argos	**#LanguagesSupported**	1	100%	3	**4**	8	8	**8**
Argos	**#Uptime**				**99.98%**	100%	100%	**100%**
Episciences	**#Articles (UoA)**	633	79%	0	**500**	603	1,357	**1,622**
Episciences	**#Journals**	14	50%	0	**7**	5	5	**8**
Episciences	**#Users**	4,471	50%	0	**2,236**	1,854	3,587	**4,240**
Episciences	**#Uptime**				**99.98%**	100%	100%	**99,85%**
Zenodo	**#TBsDataTransferred (UoA)**	19,081	110%	0	**20,989**	30,877	71,735	**80,831**
Zenodo	**#ActiveUsers**	85,000	150%	88,375	**215,875**	139,735	193,255	**224,303**
Zenodo	**#Uptime**				**99.98%**	99.25%	99.29%	**99,37%** [4]

[4] The #uptime metric exposed by Zenodo is an average of 3 different parts of the service availability (frontpage, files, search). In some cases, only one of the 3 would be down, which translates to the service being "degraded", but not "fully unavailable". Unavoidable planned maintenance operations that require some downtime are also included in this number.

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
Declared 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 KPI 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 PUBLISH portfolio.

Service name: Amnesia

The unit of access could be either "unique visits" or "downloads" of the tool as both can measure the effective interest in the desktop tool and can be used to reflect the cost-of-service maintenance and operation. Virtual Access will be provided via Actual Cost as the service offers an open API.

Service name: ARGOS

50 new "managed projects" are expected to onboard the service during the project, with 10 researchers and 5 DMPs per project as indicative average values. Supporting extra projects will require work on adapting DMP Templates to new requirements (5 new templates), supporting users to onboard the service, integration with additional 3rd party services indicated by projects, troubleshooting and enhancing language support (+1 additional language is expected to be supported during the project).

Service name: Episciences

The measurement of the number of preprints was selected because each new preprint triggers a workflow in the journal, with notifications to the users, invitations of reviewers, upload of review reports, copy-editing, automatic reminders, publication, attribution of Crossref DOI and access statistics. It is the first trackable measure of the process.

Service name: Zenodo

The measurement in TBs allows us to clearly see how much the action (e.g. uploading a dataset) of a given user impacts the overall demand on Zenodo's infrastructure. The more the actions, the bigger the total data transferred and the bigger the cost.

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 PUBLISH 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.

Amnesia

KPI: #UserVisits (Unit of Access)

acquisition methodology: Amnesia portal was covered with Matomo tracking providing information on the visits and unique visitors which is exported to VA metrics acquisition system (Prometheus).

other remarks: 2019 base value was not explicitly defined in the Grant Agreement document so it was retrieved from Matomo analytics platform.

KPI: #DownloadsOfDesktopTool (Unit of Access)

acquisition methodology: Amnesia portal was covered with Matomo tracking providing information on the number of downloads of Amnesia desktop application which is exported to VA metrics acquisition system (Prometheus). Total number includes binaries for Windows and Linux operating systems in both 32 and 64 bit versions (4 different installation files).

Argos

KPI: #ProjectsCreated (Unit of Access)

acquisition methodology: The number of managed projects is tracked by Argos and exposed via the Prometheus endpoint. The set of eligible DMPs created in Argos is restricted by:

- grant identifier association indicating EC funds
- status set to *finalised* (including *published* and *marked with DOI*)

KPI: #DMPtemplates

acquisition methodology: It indicates the number of predefined dataset templates which could be used when creating Data Management Plans. A Dataset in Argos is an editor with set up questions that support the creation of descriptions of how data are / have been handled, managed and curated throughout the research data lifecycle. The editor holds a collection of Dataset templates each one with different sets of predefined questions as per funders, institutions, research communities RDM policy requirements. Researchers and students can choose the template that corresponds to their RDM needs in order to get funding or get their degree, respectively. A DMP in Argos may consist of one or more datasets.

The number of distinctive dataset templates is tracked internally by Argos and exposed via the Prometheus endpoint.

KPI: #LanguagesSupported

acquisition methodology: Number of supported languages is tracked by Argos and exposed via the Prometheus endpoint.

Episciences

KPI: #Articles (Unit of Access)

acquisition methodology: Count of article preprint submissions (in current year) from user accounts created since 2021-01-01 00:00:00 is tracked by Episciences and exposed using the Prometheus endpoint.

KPI: #Journals

description: Number of journals.

acquisition methodology: Number of journals created since 2021-01-01 00:00:00 is tracked by Episciences and exposed using the Prometheus endpoint.

KPI: #Users

description: Number of users.

acquisition methodology: Number of user accounts created since 2021-01-01 00:00:00 is tracked by Episciences and exposed using the Prometheus endpoint.

Zenodo

KPI: #TBsDataTransferred (Unit of Access)

acquisition methodology: TBs of data transferred (counting both uploads and downloads) during the OpenAIRE-Nexus project is tracked by Zenodo and exposed using the Prometheus endpoint. Expressed in terabytes.

KPI: #ActiveUsers

acquisition methodology: Total number of registered researchers on Zenodo portal during the OpenAIRE-Nexus project is tracked by Zenodo and exposed using the Prometheus endpoint.

other remarks: Apart from the number of researchers (51k in the first 12 months of the project vs 85k defined as 2019 base) we do also track the number of unique visitors on Zenodo portal which is a slightly different way of measuring active Zenodo users. The total number of registered researchers was picked as a better and more consistent way of representing the number of active users.

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 several months after the project started.

Service name: Amnesia

KPI name	From	To
#UserVisits (Unit of access)	**20-04-2021**	**16-06-2023**
#DownloadsOfDesktopTool (Unit of access)	**20-04-2021**	**16-06-2023**
#Uptime	**01-01-2021**	**31-05-2023**

Service name: ARGOS

KPI name	From	To
#ProjectsCreated (Unit of access)	**02-06-2021**	**16-06-2023**
#DMPtemplates	**01-01-2021**	**16-06-2023**
#LanguagesSupported	**01-01-2021**	**16-06-2023**
#Uptime	**01-01-2021**	**31-05-2023**

Service name: Episciences

KPI name	From	To
#Articles (Unit of access)	**01-01-2021**	**16-06-2023**
#Journals	**01-01-2021**	**19-06-2023**
#Users	**01-01-2021**	**16-06-2023**

#Uptime	**01-07-2021**	**31-05-2023**

Service name: Zenodo

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

6. DEVIATIONS OF VA TABLES AND KPIS

Amnesia unit of access was inconsistently defined in the Grant Agreement document: #APIrequests KPI was defined in the Units of Access summary table while #UserVisits and #DownloadsOfDesktopTool indicators were defined in T4.4 ("A tool for anonymising and publishing sensitive research datasets") description of work. Since Amnesia is an anonymization tool and it could not expose any publicly available API, #UserVisits and #DownloadsOfDesktopTool indicators were finally chosen as the unit of access.

Some metrics (e.g. #UserVisits and #DownloadsOfDesktopTool for Amnesia, #ProjectsCreated for Argos, #Uptime for Episciences) could not be measured since the beginning of the OpenAIRE-Nexus project because OpenAIRE measurement software was not covering those services at that time and it was implemented and deployed several months after the project started. All such cases were indicated in the "Time period coverage" section.

#DMPtemplate KPI (part of Argos) description and its acquisition methodology was a subject for modifications in the first months of the OpenAIRE-Nexus project. Even though the value was properly reported in D4.1 the methodology description in "Indicators acquisition methodology" section was outdated and got updated in D4.2. The KPI name is misleading because in the current Argos model there is no DMP template concept. The only subject for templating is dataset encapsulating different sets of predefined questions as per funders, institutions and research communities RDM policy requirements.

7. VA ASSESSMENT REPORT

1st VA External Board for VA assessment meeting outcome

The OpenAIRE-Nexus PUBLISH 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.