D7.8
Test and validation at EOSC level

Work Package	WP7
Lead partner	MARIS
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	28 February 2023
Submission date	23 June 2023

Deliverable abstract

This deliverable focuses on capturing the current status of testing and validation of services from the RI's and ENVRI-Hub itself, that are both exposed to EOSC. This means that services are in scope that are shared via the ENVRI Catalogue of Services to the EOSC marketplace. The EOSC validation activities will check the readiness of the ENVRI-FAIR services at the sub-domain and cluster level for uptake and integration into EOSC (e.g. is there a test plan, monitoring, ..) together with their possible integrability into the ENVRI service catalogue being developed in WP5.

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Peter Thijsse	MARIS	23 June 2023
Contributing Authors	Angeliki Adamaki Ulrich Bundke Gabriel Pelouze Katerina Exter Thierry Carval Keith Jeffery et al.	ULUND / ICOS ERIC FZJ / IAGOS CMCC / LifeWatch ERIC VLIZ / LifeWatch ERIC IFREMER / EuroARGO UKRI / EPOS	
Reviewer(s)	Zhiming Zhao	UvA / LifeWatch	19 June 2023
Approver	Andreas Petzold	FZJ	23 June 2023

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	23 June 2023		Peter Thijsse

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

GLOSSARY

A relevant project glossary is included in Appendix A. The latest version of the master list of the glossary is available at http://doi.org/10.5281/zenodo.4471374.

PROJECT SUMMARY

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions, and connects the ENVRI Cluster to the EOSC.

This goal is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent, and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

D7.8 – Test and validation at EOSC level

1 Introduction

Task 7.4 focuses on testing and validating resulting ENVRI-FAIR services at different levels, including individual RI, subdomain, cluster and EOSC. WP7 will support the RIs in the subdomains in testing and validating their specific service developments.

In D7.7 there has been an analysis of the services developed in the subdomains, and means have been provided on how the services can be tested, via unit tests and integration tests.

This deliverable D7.8 focuses on capturing the current status of testing and validation of **services from the RI's and ENVRI-Hub itself, that are both exposed to EOSC.** This means that services are in scope that are shared via the ENVRI Catalogue of Services to the EOSC marketplace. The EOSC validation activities will check the readiness of the ENVRI-FAIR services at the sub-domain and cluster level for uptake and integration into EOSC (e.g. is there a test plan, monitoring, ..) together with their possible integrability into the ENVRI service catalogue being developed in WP5.

This deliverable will therefore document status of tests and validations done to:

- ENVRI services:
  - ENVRI-Hub itself
  - Knowledge base
  - Science demonstrators
- Services in the ENVRI-Hub:
  - RI services published in the ENVRI Catalogue of Services

As a result of the analysis, the overview of test and validation activities will provide feedback to the RI's and subdomains on their status of testing and validation, and WP7 will make recommendations for improvement.

Documenting the status should be seen as a step in awareness and towards maturity of services present in the EOSC sphere.

2 Methodology

The ENVRI Catalogue of Services has two major uses: (a) to catalogue ENVRI digital assets for use by researchers in the ENVRI domain (disciplinary or multi/interdisciplinary research), and in this way allow for a central validation and harmonisation action; (b) to provide the opportunity for a centralised channel for uploading ENVRI digital asset metadata to EOSC for wider use. The "central catalogue' provides means for checks on FAIRness, integrity, governance (including potential alignment of licence conditions, acknowledgement/citation) and interoperation capability of the ENVRI digital assets. The alternative also exists, and that is directly onboarding of services onto e.g. the EOSC platform if they comply with the EOSC requirements for providers and services, when needed for project reasons.

Looking at the services in the ENVRI-FAIR project the first question will be:

- Is the service present in the ENVRI Catalogue of Services, currently available via https://envri-hub.envri.eu/cservicesmain?

If the answer is positive, then the next set of questions is valid to check if testing and validation of the services is sufficient.

Validation and testing for uptake at EOSC level has two sides:

- Validation: Is the service described as it should, and are the assets published (via the service) FAIR ? => see also ENVRI-FAIR D5.5: Guidelines for validation of ENVRI-FAIR services (1.1). Zenodo (https://doi.org/10.5281/zenodo.6758244), in which the criteria for validation are listed
  - *1st Criterion for integration within the ENVRI-EOSC catalogue of services*

All services included in the catalogue need to be described by means of rich metadata. This refers most importantly to the full description of the metadata elements needed for machine actionability (e.g., description of web-service query parameters).

- *2nd Criterion for integration within the ENVRI-EOSC catalogue of services*
  Provide a clear metadata definition of the types of assets to be included in the catalogue. Applying this criterion ensures that any potential consumer of the catalogue will be able to understand whether the assets consumed are machine-actionable or human-readable.
- *3rd Criterion for integration within the ENVRI-EOSC catalogue of services*
  Ensure that the assets provided and described in the catalogue are compliant to the FAIR principles as much as possible. This includes in particular reference to the status of assignment of Persistent identifiers (PIDs), completeness of the Authentication and Authorization Infrastructure (AAI) system

- Testing: Is the service, or an update of it available and performing as it should without errors? As documented in D7.7 the WP7 team has offered subdomains and WP5 a feasible plan to test and validate the developed ENVRI-FAIR services and provided some best practices from several of the RI's.
- Monitoring: Services published as operational need to be monitored for their uptime, and in case of problems fixed.

Validation as described above is a relatively qualitative assessment on the aspects as mentioned in D5.5.

Whether there is a sufficient testing and if monitoring process is available, this can be checked by the fact that certain systematic processes are in place:

- A development environment, staging environment and operational environment
- A test plan containing unit tests and integration tests for new software releases to guide the process from development to operation
- Technical monitoring of uptime and disturbances.

An additional question to the RI's is: Is the service registered directly in EOSC Marketplace? Although the promoted solution for publication of RI services to EOSC is in ENVRI-FAIR via the catalogue of services, there are various circumstances that an RI also publishes services in the EOSC marketplace directly. This is e.g., done when onboarding a service as part of EOSC related projects (like EOSC-Future). Although this is a potential duplicate entry/route, it is not necessarily a problem since the user will decide which route to follow from the marketplace. Onboarded services via the marketplace at the moment have committed during the process to monitoring from EOSC side, and EGI checkin federation if necessary. Additionally, the marketplace benefits from the work done in ENVRI-FAIR to provide access to more FAIR assets and available services.

In this deliverable we will collect the input to the aspects mentioned above from the ENVRI and RI services that are published as part of the ENVRI Catalogue of Services (hereafter ENVRI CoS) which is onboarded onto the EOSC. To streamline the process of collection a table has been constructed which collects the above aspects for each service published at EOSC level. The results have been integrated in this document.

3 Test and validation of services at RI level

Dependency with: D7.7, D5.4, D5.5, D5.7

3.1 Inventory

Following the methodology described in chapter 2 a Google spreadsheet has been created to collect the information from the RI representatives for the services that are published per 9 June 2023 as part of the ENVRI Catalogue of Services (more services from e.g., ICOS and CREA have been added to the CoS after 9 June 2023!).

A slightly modified version of the spreadsheet with responses is shown in table 1 on the next page to be able to fit it into the document. Each column in the table corresponds with one of the questions that are listed here, for which the related shortened column header is provided in italic.

The series of questions:

- Is the service listed in the ENVRI catalogue of services? y/n
  - *ENVRI CoS*
- Is rich service metadata present including those for machine actionability? y/n
  - *Rich service metadata*
- Does the metadata include a description of the type of assets available via the service? y/n
  - *Metadata includes an asset description*
- Is a FAIR metadata description of the assets (e.g. datasets) accessible via the service available? y/n
  - *FAIR metadata description of the service's assets*
- Is a formal test plan available for new releases? (unit and integration test) y/n
  - *Test plan*
- Are test/staging/live environments available for the service? y/n
  - *Test/staging/live environments*
- Is operational monitoring in place in the operational environment? y/n
  - *Monitoring*

Additional when service is onboarded directly in marketplace:

- Is the service also directly onboarded in the EOSC marketplace? y/n
  - *On EOSC marketplace*
- Link to EOSC marketplace
  - See the original spreadsheet for the links
- Is EOSC monitoring active? y/n
  - *EOSC monitoring*
- Is the service integrated within the EOSC AAI federation (e.g., via EGI check-in)?
  - *EOSC AAI integration*

Remarks

- Some extra general remarks from the RI side are given in the original spreadsheet.

Table 1. Test and Validation results per 9 June 2023

Service name	ENVRI CoS	Rich service meta- data	Metadata includes an asset description	FAIR metadata description of the service's assets	Test plan	Test/ staging/ live environments	Moni-toring	On EOSC market-place	EOSC monitoring	EOSC AAI integration
ENVRI										
ENVRI catalogue of services - GUI	y	y	y	y	n	y	y	y	n	-
ENVRI catalogue of services - API	y	y	y	y	n	y	y	y	n	-
Knowledge base - GUI	n	n	n	n	y	y	n	n		
Knowledge base - API					y					
Marine										
CDI SPARQL endpoint	y	y	y	y	y	y	y	y	y	n
EDMO SPARQL endpoint	y	y	y	y	y	y	y	y	y	-
EDMERP SPARQL endpoint	y	y	y	y	y	y	y	y	y	-
Seadatanet - visualisation services (Erddap)	y	y	y	y	n	y	y	y	y	-
Seadatanet - Products Catalogue (CSW)	y	n	y	y	n	y	n	y	y	-
Seadatanet - Sextant SPARQL Endpoint	y	n	y	y	n	y	n	y	y	-
Ifremer OpenSearch service	y	y	y	y	y	y	y	y	y	-

Service name	ENVRI CoS	Rich service meta- data	Metadata includes an asset description	FAIR metadata description of the service's assets	Test plan	Test/ staging/ live environments	Moni-toring	On EOSC market-place	EOSC monitoring	EOSC AAI integration
Argo Data subsetting API	y	y	y	y	n	y	y	y	y	-
Argo SPARQL Endpoint	y	n	y	y	n	y	n	y	y	-
Argo Metadata API	y	y	y	y	n	y	y	y	y	-
Ifremer ERDAPP server	y	y	y	y	n	y	y	y	y	-
The NERC Vocabulary Server	y	y	y	y	y	y	y	y	y	-
Marine Regions gazetteer	y	y	y	y	n	y	y	n	n	-
WoRMS (World Register of Marine Species)	y	y	y	y	n	y	y	y		
Atmosphere										
SIOS observation facility catalogue - REST API	y									
Machine readable metadata endpoint (OGC-CSW) for the Svalbard Integrated Arctic Earth Observing System (SIOS)	y									
FAIR ENVRI atmospheric data demonstrator	y	n	n	n	n	n	n	n	n	n
Atmospheric colocation service	y	n	n	n	n	n	n	n	n	n

Service name	ENVRI CoS	Rich service meta- data	Metadata includes an asset description	FAIR metadata description of the service's assets	Test plan	Test/ staging/ live environments	Moni-toring	On EOSC market-place	EOSC monitoring	EOSC AAI integration
IAGOS Data Portal - REST API for data and metadata access	y	y	y	y	y	y	n	n	n	n
IAGOS Data Portal - Human readable interface	y	n	y	y	y	y	n	y	n	n
ACTRIS REST API	y									
ACTRIS Vocabulary Server	y									
Solid Earth										
LOS Displacement Time Series	y	y	y	y	y	y	y	n	n	y
Unwrapped Interferograms	y	y	y	y	y	y	y	n	n	y
DEM in radar geometry	y	y	y	y	y	y	y	n	n	y
Lookup table from radar coordinates to ground coordinates	y	y	y	y	y	y	y	n	n	y
Spatial Coherence	y	y	y	y	y	y	y	n	n	y
Interferogram Atmospheric Phase Screen from Global Atmospheric Model	y	y	y	y	y	y	y	n	n	y
Map of LOS Vector	y	y	y	y	y	y	y	n	n	y
Wrapped Interferograms	y	y	y	y	y	y	y	n	n	y
Waveform quality and availability metrics distributed by the ORFEUS Data Center at KNMI	y	y	y	y	y	y	y	n	n	y

Service name	ENVRI CoS	Rich service meta- data	Metadata includes an asset description	FAIR metadata description of the service's assets	Test plan	Test/ staging/ live environments	Moni-toring	On EOSC market-place	EOSC monitoring	EOSC AAI integration
Seismic waveforms distributed by the ORFEUS Data Center at KNMI	y	y	y	y	y	y	y	n	n	y
Station metadata distributed by the ORFEUS Data Center at KNMI	y	y	y	y	y	y	y	n	n	y
Biodiversity										
EcoPortal Metadata	y	y	y	y				y		-
LifeWatch ERIC Metadata Catalogue Web Service	y	y	y	y						-
RESTful endpoint to access our historical weather data.	y									

3.2 Analysis of the results at RI level

Using the responses of the RI representatives we can make the following analysis of the status of tests, validation and monitoring of services.

- *Rich service metadata present including those for machine actionability?*
  Although there is room for improvement, the RI's indicate in most cases their service metadata is well provided. This is a result of the aligned work via the Catalogue of services, and quality validation process (see D5.4 for a description of the process).

- *Does the metadata include a description of the type of assets available via the service?*
  This is the case for almost all RI's and is without doubt due to the work in ENVRI-FAIR where a lot of work has gone into increasing the FAIRness of metadata and datasets accessible via the services listed.

- *FAIR metadata description of the assets (e.g. datasets) accessible via the service available?*
  Assets are mostly metadata records, datasets or information objects (like an organisation or project description) and these are all available via the published service. Only in the case of 2 more generic human interface services, the assets are not directly accessible, but require a dialogue via the browser.
- *Test Plan available for new releases? (unit and integration test)*
  In some of the more mature and operational RIs a test plan for unit testing and/or integration is available, but not for all. There is room for improvement.

- *Test/staging/live environments available for the service?*
  For almost all services (except for the very new cross-RI Atmosphere demonstrators) these environments exist, which is a very valuable condition to overcome the missing test plans.

- *Monitoring in place on the operational environment?*
  Not all services are monitored yet while they are part of the catalogue. Monitoring of uptime, response times, etc. is critical when a service is marked as operational and published for end-users in order not to lose them because of running into errors.

Additional information, relevant when a service is also onboarded directly in the EOSC marketplace.

- *Service directly onboarded in EOSC marketplace?*
  As a result of work in other EOSC related projects some services are onboarded directly to demonstrate integration with EOSC core services and support project services. These services are directly findable in the Marketplace via the EOSC search engine.

- *Link to EOSC marketplace*
  When available the links are shared in the original inventory spreadsheet, see chapter 3.1 footnote.

- *EOSC monitoring active?*
  When services are onboarded, they will be officially monitored from the side of EOSC (apart from internal RI monitoring). This monitoring is in place for most, or underway. Both users and providers have access to the results of the monitoring service.

- *Is the service integrated within the EOSC AAI federation (e.g. via EGI check-in)?*
  As part of onboarding a seamless access to services is meant to be provided by federated authentication. For this the EGI check-in system can be used and federated, but only if the service actually requires authentication.

4 Test and validation at ENVRI services level

Dependency with: D7.8, D5.5 and MS20

This set of services concerns "overarching" services that are developed at the ENVRI-FAIR level to support the ENVRI community in reaching the end-user in general via the ENVRI-Hub services.

Table 2. Overview of main ENVRI services

#	Service name
1	ENVRI catalogue of services - ENVRI-Hub GUI
2	ENVRI catalogue of services - API
3	ENVRI Knowledge base - GUI
4	ENVRI Knowledge base - API
5	ENVRI Knowledge Base search engine

4.1 Inventory of test and validation status

For collecting test, validation, and monitoring information from the ENVRI(-Hub) level services themselves the Google spreadsheet table with questions was less fit, since they concern a central service or were not meant to be onboarded/published to EOSC (like e.g., the ENVRI Knowledge Base). Therefore, the information has been collected in a qualitative way, for the points that are relevant.

4.1.1 ENVRI-Hub GUI and API

Validation

In principle the ENVRI catalogue of services GUI back-end is currently "consuming" the catalogue. The content of the GUI depends on what is provided there by the API which collects all service descriptions. The central catalogue service checks the metadata provided by the service providers as turtle files during the "merge" step. The turtle files are generated via:

- the SHAPEness editor
- the service providers themself
- codes directly without an editor

In this process there is thus a first syntax check (validation) implemented.

As a second step in the validation there is a manual/visual check by the central ENVRI-Hub team and the providers of the metadata using the ENVRI-Hub GUI. Feedback then flows to the providers for change, e.g., if fields are missing or textual content looks odd or is incomplete.

On the ENVRI-Hub side there is an automatic update notification implemented if there are changes/additions on the catalogue side. Also the Turtle files used for the two SPARQL endpoints files are synchronised automatically. At the moment both sync scripts are not scheduled regularly to prevent server downtime as long as the catalogue is a moving target and automated checks (if exist) are not really operational.

Monitoring

To help people finding issues we have setup a regularly triggered service check:

- https://hubtest.envri-fair.eu/api_check

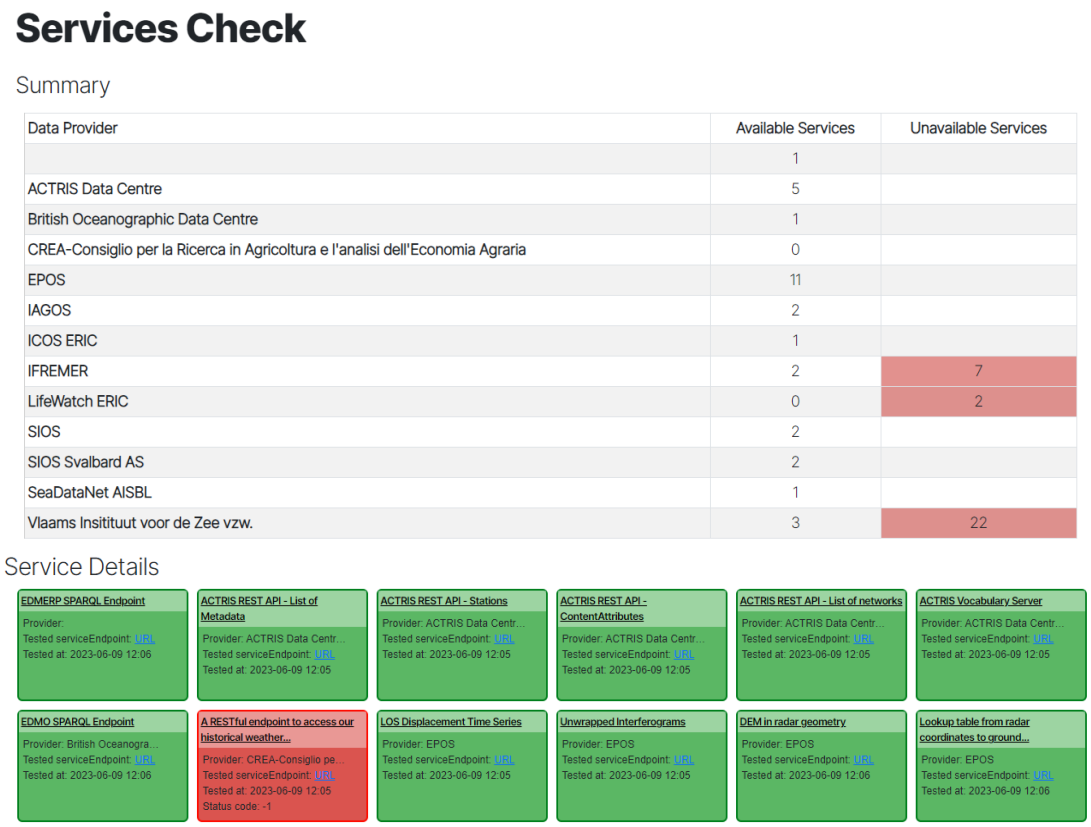

Figure 1: Overview of services check at ENVRI-Hub level

Here providers can see if a server responds within time including the status code or report otherwise a "timeout". Also, the last timestamp of the check is reported.

At the moment there is no "check if alive" field defined in the DCAT-AP profile providing an URL to check the service. At ENVRI-Hub level there is currently work ongoing on a temporary "work around" using the "endpoint" syntax information, however this does not help for monitoring services with authentication.

Testing process
At ENVRI-Hub level for the hub services there is a test procedure implemented to check the Django framework application during merge/commitment steps.
Currently a development server https://hubtest.envri-fair.eu/ is used to check the new release before it is finally deployed on the EGI Server https://envri-hub.envri.eu/. The deployment itself is script controlled.

4.1.2 Knowledge base GUI and API

The GUI and API can be seen as one service to analyse. The core functionalities are now part of the search engine (https://search.envri.eu), for which there is no public API. There is also a RDF management system (based on Ontowiki: https://envri-fair.lab.uvalight.net/OntoWiki/index.php) with a SPARQL endpoint. However, it is independent from the ENVRI Knowledge Basesearch engine, and isn't actively developed or monitored yet.

4.1.3 Search engine

For the ENVRI Knowledge Base search engine, there are integration tests and a staging/production deployment strategy in place. Currently, the integration tests are run manually, but there are plans to automate them in the near future. In summer 2023 monitoring of the operational environment will be implemented.

4.2 Analysis of the results at ENVRI level

The ENVRI catalogue of services is a critical component/deliverable of the project and has sufficient monitoring in place, both on itself as well as on its content. Automated content monitoring would be ideal, but this is not possible yet due to the free text nature of many service metadata fields.
The Knowledge base and search engine are demonstrators of what is possible and more as a research output of the project. Therefore, testing and monitoring is not applied yet.

5 Conclusions and recommendations

The conclusions and recommendations are given separately for the:

- The services published in the ENVRI-Hub
- High level ENVRI services, e.g; ENVRI-Hub, ENVRI Knowledge Base and Use cases

5.1 The services published in the ENVRI-Hub

Using the information gathered in table 1 we can give a current status of validation, testing including monitoring and integration with EOSC marketplace of the services from the RI's.

Validation
Regarding the validation it is found that most RI's have their service metadata as well as the asset metadata provided well. Furthermore, the assets that the services provide follow the FAIR principles and are in most cases directly accessible. There are no big differences between the different sub-domains, apart from the fact that there are two atmosphere services that lack behind the others.

Testing
The marine and atmosphere services lack behind the solid earth services with regards to an available test plan used for new releases. Even though most of these services do have a test/staging/live environment available, a test plan would still be a big improvement as it contains unit tests and integration tests for new software releases to guide the process from development to operation.

Monitoring
Not all services that are part of the catalogue are being monitored yet, while the monitoring of uptime, response times, etc. is critical when a service is marked as operational and published as the end-users will not return to a service, if it is down often.

EOSC Marketplace and Integration
The marine domain has most of its services listed on the EOSC marketplace as a result of the EOSC-Future project where partners are involved in different Science Projects. These services are mostly integrated with EOSC core services, such as EOSC monitoring as well as horizontal services such as EGI-check-in federation. The marketplace offers increased visibility to the individual ENVRI services in parallel to them also becoming visible via the ENVRI catalogue of services as a whole on the marketplace. More in detail: If a user searches for a specific marine service he/she will not get the ENVRI catalogue content, because there is no integration with the service metadata records of the ENVRI catalogue. But if he/she search for "marine" in services she/he will get the ENVRI catalogue in the results and from there an additional entry point to the ENVRI marine services
This "duplication of access to services" as such is not a bad thing as onboarding of the services individually also enables the use of the EOSC core and horizontal services that can help the user get more smooth access to the ENVRI services.

Therefore, a recommendation would be to the services from the Atmosphere, Biodiversity and Solid Earth domain to onboard their services to the EOSC marketplace, to increase visibility and make use of the EOSC core or horizontal services they see fit.

5.2 High level ENVRI services

The ENVRI catalogue of services performs a syntax check and a manual check on the metadata of the services in the catalogue. It would, however, be recommended to have an automated content monitoring in place for the content of the catalogue of services. This would require a different approach to the metadata fields of the services that are currently in free text, which makes that difficult.

Furthermore, the catalogue of services has sufficient monitoring in place, both on itself as well as on its content via a regularly triggered service check from the ENVRI-Hub. An improvement would be to include a “check if alive” field in the DCAT-AP profile providing an URL to check the service.

At ENVRI-Hub level there is a development server available that allows for checking a new release of the catalogue of services before it goes into production.

The Knowledge base and search engine do not have monitoring and testing applied as they are approached more as a research output of the project. However, if the catalogue of services will be sustained after the project ends, it would be very beneficial to also sustain the Knowledge base, to get a better idea of the landscape in which these services operate. This would then require having also monitoring and testing in place.