D7.7
Test and validation at subdomain and cluster report

Work Package	D7
Lead partner	MARIS
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	31-12-2021
Submission date	18-01-2022

Deliverable abstract

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 subdomains by testing and validating their specific service developments. Moreover, WP7 will test and validate the resulting ENVRI-FAIR services portfolio at the subdomain, cluster and EOSC level, following the validation guidelines as provided by WP5. In D7.7 the cluster validation activities have been specified and each will test real scientific use cases to check the quality and validate the applicability of (ENVRI-)FAIR services. The EOSC validation activities will test the readiness of the ENVRI-FAIR services at the sub-domain and cluster level for uptake and integration into EOSC together with their possible integrability into the ENVRI service catalogue being developed in WP5.

As a result, the test and validation activities will provide feedback to the RI's and subdomains as to the results of testing and validation and make recommendations to help steer development effectively throughout the project. Additionally, it will provide a report on validation results to WP5 (T5.4) for inclusion in synthesis analysis on the readiness of ENVRI-FAIR services for uptake in EOSC.

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	15 January 2022
Contributing Authors	Keith Jeffery Markus Fiebig Nicola Fiore Zhiming Zhao	EPOS NILU LifeWatch UvA	15 January 2022
Reviewer(s)	Spiros Koulouzis Alex Vermeulen Zhiming Zhao	Lifewatch/UvA ICOS/ULUND UvA	
Approver	Andreas Petzold	FZJ	18 January 2022

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	3-11-2021	First draft with document structure and collected use cases from subdomain implementation plan documents. Distributed to subdomain representatives.	Peter Thijsse (MARIS)
V 0.2	30-11-2021	Feedback by subdomain representatives	Keith Jeffery (EPOS), Markus Fiebig (NILU) Nicola Fiore (LW) Thierry Carval (EuroArgo/Marine), Peter Thijsse (SDN)
V 0.3	15-12-2021	Review of feedback and consolidation	Peter Thijsse (MARIS), Zhiming Zhao (UVA)
V 0.4	22-12-2021	Technical, editorial and scientific review of the document – done by an internal expert.	Spiros Koulouzis (Lifewatch/UvA)
V 0.5	11-1-2022	Review as WP5 lead and user community (RI) representative, analysed the usability of the deliverable for Stakeholders	Alex Vermeulen (ICOS/ULUND)
V 0.6	12-1-2022	Review of feedback and consolidation	Peter Thijsse (MARIS), Zhiming Zhao (UVA)
V 0.7	13-1-2022	Decision that the reviewer comments are sufficiently well taken care and the deliverable quality is acceptable	Zhiming Zhao (UVA)
V1.0	14-1-2022	Official release to coordination	Peter Thijsse (MARIS)

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

D7.7 Test and validation at subdomain and cluster report 5
1 Objectives 5
2 Methodology 6
3 Test plan for use cases at the Subdomain level 7
3.1 3.1 Atmosphere 7
3.2 Marine 8
3.3 Solid Earth 9
3.4 Biodiversity and terrestrial ecosystems 11
4 Test plan for use cases at the ENVRI services level 13
5 Test procedure and steps 14
6 Annex 1 A sample user story template for an Integration Test 15
7 Annex 2 A sample template for a Unit Test Plan 15

D7.7 Test and validation at subdomain and cluster report

1 Objectives

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 subdomains by testing and validating their specific service developments. Moreover, WP7 will test and validate the resulting ENVRI-FAIR services portfolio at the subdomain, cluster and EOSC level, following the validation guidelines as provided by WP5.

In the following, the cluster validation activities have been specified and each will test real scientific use cases to check the quality and validate the applicability of (ENVRI-)FAIR services. The EOSC validation activities will test the readiness of the ENVRI-FAIR services at the sub-domain and cluster level for uptake and integration into EOSC together with their possible integrability into the ENVRI service catalog being developed in WP5.

This deliverable will document:

- use cases from RIs and subdomains (guided by requirements and gaps identified by T5.1)
- common services that can be tested and supported;
- an indicative schedule for released services, which will be supported by test activities to be performed with RIs for validating their service developments;
- an overview of test and validation activities of ENVRI-FAIR services at the subdomain and cluster level that can be aligned with the development schedules of the RIs and following WP5 guidelines;
- a template for the user stories (related to the selected use cases) to be defined by the WP7 team with the RI/Subdomain representatives at the start of the test activities.

As a result, the test and validation activities will provide feedback to the RI's and subdomains as to the results of testing and validation and make recommendations to help steer development effectively throughout the project. Additionally, it will provide a report on validation results to WP5 (T5.4) for inclusion in synthesis analysis on the readiness of ENVRI-FAIR services for uptake in EOSC.

2 Methodology

WP 7 aims to provide support to the developments in the ENVRI-FAIR community, and specifically to the common developments in subdomains and RIs involved. Secondly, WP7 aims to support the developments of the services exposed at the ENVRI-HUB level which could be published to serve as the interface of the ENVRI community to EOSC.

This deliverable offers a feasible plan for test and validation activities of these ENVRI-FAIR services. It is important to align with the development schedules of the RIs, follow the WP5 guidelines in D5.5, as well as take into account the analysis of the ENVRI-HUB related services defined by task force 6.

Taking into account the broadness of the FAIRness improvement and development activities in the project and the different requirements, T7.4 aims to develop the test plan in close cooperation with WP8-11 and supported by WP5 and TF6. It is assumed that the RI's themselves (as part of WP8-11 work) do the **unit tests** for the relevant existing (supposed to be FAIR) and the newly developed software and solutions. For this work a template as provided in Annex 2 should be used. The supported tests as described in this document are **integration tests** with a focus on priority services and common developments and will use the template in Annex 1.

The process to come to a structured plan has been the following:

1. To create an overview of developed software and the supported use cases (involving services being developed and from a third party)
    a. At the subdomain level - **with a focus on the main technical solutions**.
    b. This overview will be based on subdomain implementation plan documents produced in WP8-11.
    c. At the ENVRI-HUB level - based on the specified developments by the ENVRI-HUB working group (mainly TF1 and TF6)

2. Set the priorities for testing and a first-time planning based on the release of the developments;
    a. The focus will be on the common/shared developments in the subdomains.
    b. Other priorities are the developments that have a direct link to ENVRI-HUB components: Knowledge base, (data) service catalogue and demonstrators.
    c. The priority for testing will also consider the available resources in WP7.

3. Define the steps in the test process:
    It is important to define the steps in the test process including collecting information about the latest status of development (e.g., URLs), defining the tests for the test round in a user story, and providing expected feedback. This will be further explained in chapter 5.

3 Test plan for use cases at the Subdomain level

Use cases have been collected from the implementation plans provided by WP8-11. **We mainly focus on the mature technical developments, preferably with overlap to other subdomains; we do not intend to collect the domain specific (less mature) developments.**

3.1 3.1 Atmosphere

Subdomain contact: Markus Fiebig mf@nilu.no

Table 1. Overview of main WP8 Atmosphere use cases

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for test of release
3.1.1a	Standard interfaces for metadata access	Query the catalogue interface endpoint, developed with Restful interface standardisation. In other words: Search and determine Atmospheric RIs data availability for the period and area of interest; and in some cases Essential Climate Variables (ECVs)	ACTRIS: https://prod-actris-md.nilu.no/index.html https://prod-actris-md.nilu.no/Metadata https://prod-actris-md.nilu.no/Metadata/content/ethane	April 2022
3.1.1b	Standard interfaces for data access	Perform dataset search supported by semantic elements (based on vocabularies) Check the data access to OpenDAP services. => Fetch the selected data including automatic previews of the datasets	ACTRIS: "dataset_url":"https://thredds.nilu.no/thredds/dodsC/ebas/CV0001G.20151231235128.20170725000000.online_gc..air.1y.3262s.CV01L_Agilent_GC-FID_7890A_G3440A.CV01L_Manual_AIR_only.lev2.nc"	April 2022
3.1.2	Indexing of data resources in GEOSS	Query the OAI PMH catalogue interface endpoint, developed with Restful interface standardisation??	https://ebas-oai-pmh.nilu.no/oai/provider?verb=ListIdentifiers&set=actris-dc&metadataPrefix=iso19139	April 2022
3.1.3	Endpoint for providing service metadata to ENVRI-Hub	Request the service metadata endpoint to feed the ENVRI-Hub	n.a.	April 2022
3.1.4	Documenting provenance	Request the provenance information in (validated) PROV-O format => (Download data as a bundle with automated compilation of provenance information)	Example ACTRIS Cloud Remote Sensing to which PROV-O format will need to be added: https://cloudnet.fmi.fi/file/522d0505-a298-4c71-8043-cdad9f4e81f7 https://cloudnet.fmi.fi/quality/522d0505-a298-4c71-8043-cdad9f4e81f7	December 2022

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for test of release
3.1.5	Execute Jupyter notebook processing	Execute a Jupyter notebook that defines a workflow to process data from various Atmosphere datasets (linked to co-location service demonstrator)	https://github.com/damienboulanger/envri-wp8-demonstrator	December 2022 (High priority - this ties all together)

3.2 Marine

Used document: D9.6[1]
Subdomain contact: Thierry Carval Thierry.Carval@ifremer.fr

The marine subdomain is working on a demonstrator in two versions in which a central component will query services in each of the involved RI's:

- V1 (2021): A brokered query for EOV datasets from available ERDDAP instances in the marine RIs
- V2 (2022): A brokered query for EOV datasets from all available API's and SPARQL endpoints in the marine RIs

Table 2. Overview of main WP9 Marine use cases - Release 1 (end 2021/early 2022)

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for test of release
3.2.1	Broker EOV term to P01 terms	- Query the broker API to retrieve a list of P01 codes for 1 EOV (e.g., oxygen) - check response	n.a.	April 2022
3.2.2	Request RI ERDDAP services with P01 terms	- Query for each RI the ERDDAP service with the P01 terms - check ERDDAP response	n.a., will need URL per RI	April 2022
3.2.3.	Feed list of data URL's back to the user	- Check how the list of URLs is provided back to the user.	n.a.	April 2022

Table 3. Overview of main WP9 Marine use cases - Release 2 (end 2022)

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for test of release priority (short/mid/long term)
3.3.1	Broker EOV term to P01 terms	- Query the broker API to retrieve a list of P01 codes for 1 EOV (e.g., oxygen) - check response	n.a.	December 2022
3.3.2	Request RI API's and Sparql endpoints with P01 terms	- Query the RI API's and Sparql endpoints with P01 terms received from broker - check the responses, e.g., is the service up, do the responses provide back dataset links, is there m2m	n.a., will need URLs of all RI services	December 2022

[1] https://zenodo.org/record/4766796

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for test of release priority (short/mid/long term)
		metadata available for each dataset		
3.3.3	Feed list of data URL's back to user	- Check how the list of URLs is provided back to the user.	n.a.	December 2022

3.3 Solid Earth

Used document: D10.4[2]
Subdomain contact: Keith Jeffery <Keith.Jeffery@keithgjefferyconsultants.co.uk>

Table 4. Overview of main WP10 Solid Earth use cases

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for of release priority (short/mid/long term)
3.3.1	ICS-C (EPOS Core): Query the ICS-C vocabulary service to retrieve available vocabularies and mappings	Web API + a GUI for creating, managing, publishing, and querying (via LD-API or SPARQL) registers.	Basic vocabulary: http://data.geoscience.earth/ncl/ Registry tool: UkgovLD: https://github.com/UKGovLD/registry-core	April 2022
3.3.2	ICS-C: Search API	Query the new Search API and retrieve the data as DCAT-AP/json, geojson, json	https://www.ics-c.epos-eu.org/	December 2022
3.3.3	ICS-D: check the FAIRness of the VRE	VRE user interface, allows access to the Jupyter Notebook to process data - What are the FAIRness improvements? (metadata for the services, authentication/authorisation handling, …?)	https://docs.google.com/document/d/1kP4Uf7jb5t_KgDWFMgwORo3tsWcaqNRc/edit https://docs.google.com/spreadsheets/d/1v7_joZZpLcXoWKMfSvJ8JHHP-W_rxFxB/edit?usp=drive_web&ouid=113866711328656043951&rtpof=true	December 2022
3.3.4	Self defined use case for testing ICS-C and ICS-D	Collect datasets into ICS-C workspace. • Add ICS-D services (e.g., Jupyter Hub with pre-installed ObsPy, Enlighten visualisation, etc.) into ICS-C workspace. • Build a workflow in ICS-C. ICS-D services have predefined options/methods for how they can be used within the workflow.	https://gitlab.com/KNMI-OSS/swirrl/swirrl-api	December 2022 (In parallel a prototype workflow system developed in seismology is being tested and further developed)

[2] https://zenodo.org/record/4418953

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for of release priority (short/mid/long term)
		• Deploy workflow to a dedicated cluster (yet another ICS-D) and run. o fetch data for each TCS service based on user defined parameters o convert TCS payload to specific format (FEATHER ?) • Demonstrate how the SWIRRL API works as a generic service for fetching data • Combine and stage the files onto a dedicated environment, launch Jupyter Notebook through SWIRRL API • Prepare data for visualisation in notebook • Launch Enlighten through the SWIRRL API. • Create snapshot of the environment on Github for later use/sharing.		

3.4 Biodiversity and terrestrial ecosystems

Document used: D11.2[3], section 4.2 "User-oriented cross-RIs demonstration cases".
Subdomain contact: Nicola Fiore <nicola.fiore@lifewatch.eu>

Table 5. Overview of main WP11 Biodiversity and terrestrial ecosystems use cases

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for of release priority (short/mid/long term)
3.4.1	Soil Water content - check similar metadata model applied across sites [1]	In this use case we have defined a metadata schema to be used to describe Soil Water Content datasets, based on EML2.2.0, that includes about 40 fields. The agreement on the schema allowed it to perform a harvesting exercise by taking metadata records from the different involved Ris. An XSLT transformation has been developed in order to allow the mapping from ISO19139 to EML2.2.0. Metadata records can be exported in RDF to store them into a triple store and perform data discovery and analysis.	https://www.google.com/url?q=https://drive.google.com/drive/folders/1TEC5urVChdvLeZPZ8fJ1OQYzlmlUdWO2&sa=D&source=docs&ust=1639151508085000&usg=AOvVaw16y6GugBu_wmpUUmGuG1-q https://github.com/luciaV86/ISO19139_to_eml220	Test prototype April 2022 Test Operational December 2022
3.4.2	Species Scientific Names Identification (SNI): data query for scientific names	In this use case, we explored the usability and interoperability of species names and related data across the different RIs within the ENVRI cluster (such as ICOS, eLTER, ANAEE, LifeWatch) We run some datasets including scientific names against the LifeWatch taxon match service and Catalog of Life name-match tool. As a long-term result, we decided to turn the report into a RIO paper. A notebook was created comparing a few other services (this will be polished and cleaned up and moved to a different repo and will be included as a supplement to the paper).	https://docs.google.com/document/d/1pAjkqq9Kg8rWcI2pKs_uuHaDWeQ8WzKFjgc278PUnXk/edit# https://github.com/sharifX/random/blob/main/names-tool-comparison.ipynb	Test prototype April 2022 Test Operational December 2022
3.4.3	Site Documentation Interoperability (SDI)	The used catalogues have different terminologies and definitions for their research facilities. In this use case, we agreed on common	https://docs.google.com/spreadsheets/d/1dCSQeWQKEnnwagSoDfNHcNxFT3WMquKR6zs382rP56c/edit#gid=0	Test prototype April 2022 Test Operational December 2022

[3] https://zenodo.org/record/4682826

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for of release priority (short/mid/long term)
		definitions to facilitate data aggregation and allow us to search site information across Ris. In particular, a minimum set of information describing research sites has been developed ("core fields" acceptable for all RIs and reused also for the Soil Water Content Use Case). The use case is focused on ensuring seamless feeding of the ENVRI-hub through the generation of DCAT formatted metadata based on the site documentation data provided by the RIs.	https://github.com/stopopol/deims_dcat	
3.4.4	Common layer of Core MetaData (CMD): Testing the availability of m2m accessible RDF (DCAT-AP) describing the data and services as developed and used in the other use cases.	UC will provide guidelines tools, and documentation to foster DCAT compliance over datasets developed within Use cases as prescribed by TF1	• DCAT validation API • Notebook example DCAT usage handbook	December 2022

4 Test plan for use cases at the ENVRI services level

Dependency with: D5.5 (M39) (not yet available): Guidelines for validation of ENVRI-FAIR services

This set of use cases 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.

Document used: D5.2

Table 6. Overview of main ENVRI services use cases

#	Use case name	Description with services required/involved	URLs involved (a starting point - not finalised)	Timeline for of release prio (short/mid/long term)
4.1.1	ENVRI-Hub catalogue of services - GUI	- Search for "..." via the user interface - Check browse list of content - access details page (sample set) - check content of details pages		April 2022
4.1.2	ENVRI-Hub catalogue of services - API	- Request API get-capabilities - check content - Query the API for "..." - Check response listing - Request the details of one service (sample set) - Check content of response		December 2022
4.1.3	Knowledge base - GUI	- Search interface for different types assets, e.g., web pages, data sets, API - Interface to other KB tools, e.g., consult interface, Provenance template - Interface for managing indexing pipeline at the backend.	search.envri.eu	April 2022
4.1.4	Knowledge base - API	- Define the exact user story first. - Is the API up? - Submit query - Check response		April 2022

5 Test procedure and steps

In previous chapters the use cases and services have been defined where WP7 will provide test support to the subdomains. When the time has come to start (following progress and timeline), the integration tests will start with the following (repetitive) sequence of actions:

1. WP7 team to contact the responsible developers to make a concrete test plan for the software release
    a. Set the exact release date (regular updates!).
    b. Check the status of developments at the start for the use cases defined in chapter 3.
    c. Collect the required and most recent version of the URLs involved in the use case.
    d. Based on use cases in chapter 3 and the actual state of developments, define the user stories, and document them according to the template in Annex 1, for the test together with the subdomain/RI representatives.
    e. Check: Review of the documentation (do they provide the required information for tests). Are the services well described regarding input, standards and services used, and expected output?
2. Make an integration test following the design requirements and user stories. Test the integrated parts of the services if it concerns more than one. This will be done by WP7 experts together with RI's and subdomain representatives. Of course, during development, the RI's/Subdomains should perform similar tests already.
3. Document the results.
4. Provide feedback and directions for improvement.
5. Agree on a time frame to repeat the test (steps can be repeated multiple times - maximum 3 iterations).

6 Annex 1 A sample user story template for an Integration Test

Title:	Priority:	Estimate:
As a <type of user> I want to <perform some task> so that I can <achieve some goal>		
Acceptance criteria Given <some context> When <some action is carried out> Then <a set of observable outcomes should occur>		

7 Annex 2 A sample template for a Unit Test Plan

Nr.	Requirements	Typical Components	Detailed Description
1)	Introduction	a) Test Strategy and Approach	
		b) Test Scope	
		c) Test Assumptions	
2)	Walkthrough (Static Testing)	a) Defects Discovered and Corrected	
		b) Improvement Ideas	
		c) Structured Programming Compliance	
		d) Language Standards	
		e) Development Documentation Standards	
3)	Test Cases (Dynamic Testing)	a) Input Test Data	
		b) Initial Conditions	
		c) Expected Results	
		d) Test Log Status	
4)	Environment Requirements	a) Test Strategy and Approach	
		b) Platform	
		c) Libraries	
		d) Tools	
		e) Test Procedures	
		f) Status Reporting	

source ref: The Art of Software Testing – Myers, Glenford J. / Software Testing & Continuous Quality Improvement - W E Lewis