D4.5: Interfacing EOSC Report (Release 2)

Work Package	WP4 - Developing and operating the Blue Cloud VRE, its services and Virtual Labs
Lead Partner	DKRZ (EUDAT)
Lead Author (Org)	Merret Buurman (DKRZ), Pasquale Pagano (CNR), Leonardo Candela (CNR), Heinrich Widmann (DKRZ)
Contributing Author(s)	
Reviewers	Dick M.A. Schaap (MARIS)
Due Date	30.04.2022, M31
Submission Date	08.07.2022
Version	1.0

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Blue-Cloud - Piloting Innovative services for Marine Research & the Blue Economy - has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n. 862409.

DISCLAIMER

"Blue-Cloud, Piloting Innovative services for Marine Research & the Blue Economy" has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n.862409.

This document contains information on Blue-Cloud core activities. Any reference to content in this document should clearly indicate the authors, source, organisation, and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the Blue-Cloud Consortium, and it cannot be considered to reflect the views of the European Commission. 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.

COPYRIGHT NOTICE

VERSIONING AND CONTRIBUTION HISTORY

Version	*Date*	*Authors*	*Notes*
0.1	01.11.2021	Heinrich Widmann (DKRZ)	Initial draft
0.2	03.01.2022	Heinrich Widmann (DKRZ)	Proposal for step by step Onboarding
0.3	13.04.2022	Merret Buurman (DKRZ)	Overhaul of chapter 4
	17.04.2022	Merret Buurman (DKRZ)	Continued
	27.06.2022	Leonardo Candela (CNR)	Added sections on Blue-Cloud catalogue.
	29.06.2022	Merret Buurman (DKRZ)	Rewrote introduction, summary, conclusion. Updated section numbers and figure numbers. Updated list of abbreviations and list of references.
	30.06.2022	Leonardo Candela (CNR)	Edited section 3.4.1
	30.06.2022	Dick M.A. Schaap (MARIS)	Reviewed as Blue-Cloud Technical Coordinator
1.0	01.07.2022	Sara Pittonet, TRUST-IT	Finalised for EC submission

Table of Contents

Executive summary 4
List of abbreviations and acronyms 6
1. Introduction 7
2. The Blue-Cloud Catalogue 9
2.1. Description of the Catalogue 9
2.2. The Blue-Cloud onboarding procedure 14
2.3. The Blue-Cloud Catalogue organisational model 16
3. The EOSC Catalogue and onboarding procedure 17
3.1. Description 17
3.2. Definition of EOSC Providers and EOSC Resources 17
3.3. EOSC Profiles 18
3.4. Preconditions for onboarding 18
3.5. The onboarding procedure 19
3.5.1. Initial onboarding of resource providers and services 19
3.5.2. Onboarding/updating via EOSC Catalogue HTTP API 19
3.5.3. Alternative: Bulk-onboarding the entire catalogue 21
4. Onboarding Blue-Cloud services into the EOSC Catalogue 22
4.1. Integration of the service providers 22
4.2. Onboarding the provider “Blue-Cloud” 25
4.3. Onboarding of the services 26
4.3.1. Service description and Blue-Cloud catalogue integration 27
4.3.2. Retrieval of service metadata from the Blue-Cloud catalogue 28
4.3.3. Mapping of the service metadata 29
4.3.4. Validation using the EOSC validation API 29
4.3.5: Onboarding of first service via GUI 31
4.3.6. Onboarding of the other services via HTTP API 34
4.3.7. Authentication and Authorization 35
4.3.8. Results 36
5. Summary, conclusions and next steps 39
References 41

Table of figures

Figure 1 Screenshot of the Blue-Cloud service catalogue 9
Figure 2 Catalogue Data Model 10
Figure 3 Publish an item - Common metadata. 15
Figure 4 Publish an item - Type specific metadata. 15
Figure 5 D4Science is listed as a provider in the Blue-Cloud catalogue 22
Figure 6 Information about the provider D4Science in the Blue-Cloud catalogue. 23
Figure 7 D4Science is already a Provider in the EOSC Catalogue 24
Figure 8 Resources already provided by the Provider D4Science in the EOSC Catalogue. 25
Figure 9 The Blue-Cloud service was approved as a provider in EOSC. 26
Figure 10 Catalogue view of Services onboarded into the Zoo-Phytoplankton EOV 27
Figure 11 Blue-Cloud Service Onboarding Form 28
Figure 12 Screenshot of the first page of the onboarding GUI with the data of the first service filled in. 31
Figure 13 The possible subcategory options depend on the selected category. This is something we had to implement in the python tool as well. 32
Figure 14 The First Resource Status was pending while some questions and clarifications were being solved by the Blue-Cloud and EOSC teams via email. 33
Figure 15 The first resource was successfully onboarded and can now be found in the marketplace. 33
Figure 16 This is what users see in the marketplace. 34
Figure 17 Services in the Blue-Cloud catalogue. 38

Table of tables

Table 1 Provider Profile classes and elements 12
Table 2 Service Profile classes and elements 12
Table 3 Dataset Profile classes and elements 13
Table 4 Method Profile classes and elements 14
Table 5 Deliverable Profile classes and elements 14

Executive summary

The Blue-Cloud Service platform features a variety of services that can be used for undertaking world-class science via the **European Open Science Cloud** (EOSC) framework, by featuring leading operational marine research infrastructures and e-infrastructures. As part of its service offer, the Blue-Cloud service catalogue gathers all the services enabled in the Blue-Cloud Virtual Research Environment (VRE) that will be onboarded in the EOSC Catalogue. This onboarding, once finished, will allow easy access to all EOSC users to Blue-Cloud services via the EOSC Portal Marketplace[1].

This deliverable describes the onboarding of Blue-Cloud services available in the Blue-Cloud Virtual Research Environment (VRE) into the EOSC Catalogue. It includes details on the activities performed to ensure that the configuration, formats and metadata of resources are compatible with the requirements of the EOSC Catalogue and Marketplace. The document also introduces how providers have been/are being imported and referenced into the EOSC Catalogue and Marketplace.

While the different procedures for onboarding 'providers' and 'services' are described in detail in the first release of the deliverable (D4.3), this document goes a step further by describing the actual onboarding within the EOSC Catalogue and should be considered as updated and extended version of D4.3.

[1] https://marketplace.eosc-portal.eu/

List of abbreviations and acronyms

Abbreviation	Signification
EOSC	European Open Science Cloud
VRE	Virtual Research Environment
API	Application programming interface
HTTP API	API that uses Hypertext Transfer Protocol
REST API	HTTP API that follows specific concepts called Representational State Transfer
GUI	Graphical User Interface
DKRZ	Deutsches Klimarechenzentrum GmbH / German Climate Computing Centre, Blue-Cloud partner leading this deliverable.
EOV	Essential Ocean Variables
CNR	National Research Council of Italy, Blue-Cloud partner co-leading this deliverable.
VLab	Virtual Laboratory
Virtual Laboratory	A web-based environment providing its designated community of practices with the services and facilities needed to accomplish a goal.

1. Introduction

Blue-Cloud is a large pan-European project federating thematic data infrastructures and services from marine research institutions across Europe. This is achieved thanks to the Blue-Cloud platform, whose architecture (Schaap et al. 2020) consists of two major components: (a) the *Blue-Cloud Data Discovery and Access* service to serve federated discovery and access to 'blue data' infrastructures; and (b) the *Blue-Cloud Virtual Research Environment (VRE)* to provide a Blue-Cloud VRE as a federation of computing platforms and analytical services. The resulting Blue-Cloud VRE is made available via the Blue-Cloud gateway at https://blue-cloud.d4science.org. This gateway provides access to the set of services and Virtual Labs developed by the project (Assante et al, 2021).

This deliverable illustrates the thematic catalogue of services available in the Blue-Cloud Virtual Research Environment (VRE) and reports on the activities carried out to interface them with the European Open Science Cloud (EOSC), so as to make the Blue-Cloud services findable and accessible via the EOSC Catalogue and Marketplace. This will make the Blue-Cloud services more visible to potential users which are not part of the consortium, benefitting scientists from various scientific disciplines.

The general workflow to "onboard" the Blue-Cloud services into the EOSC Catalogue is to register the service providers according to the EOSC guidelines[2], and to register each service in the EOSC Catalogue. Before this, providers and services were already identified, described and onboarded into the Blue-Cloud Thematic Service Catalogue with appropriate comprehensive metadata in an EOSC-compatible format. The overall workflow includes:

- registration of the provider 'Blue-Cloud' via the EOSC website and the Service Description Template provided by EOSC on request. This is currently the only way for providers to register.
- registration of services via the Web GUI provided by EOSC. This way is mandatory for a provider's first onboarded service.
- registration of services via the HTTP REST API interface provided by EOSC. This method is used for all but the first service. Due to technical problems, this process is not entirely finished yet. However, the tooling needed to perform this has been developed and it is expected that the onboarding process will be carried out in the upcoming weeks.

Section 2 describes the Blue-Cloud Service Catalogue and its onboarding process. The EOSC Service Catalogue with its preconditions is described in Section 3, while its onboarding process is described in detail in Section 4.

The onboarding of Blue-Cloud VRE services into EOSC, and consequently the deliverable reporting about it, was performed over two phases, one from 11/2020 to 04/2021 (reported in deliverable D4.3), and a second phase started in 11/2021 and still ongoing (reported in the current deliverable D4.5).

[2] https://providers.eosc-portal.eu/becomeAProvider

The first release of the deliverable (D4.3) focussed mainly on the identification of services to onboard into the Blue-Cloud catalogue. In this second release (D4.5) the onboarding procedure being carried out is described, including semantic and technical issues that had to be fixed. Due to technical and interoperability difficulties caused by the interaction with the evolving EOSC Catalogue and Marketplace, the onboarding process is still ongoing.

At the time of writing, out of the services that fulfilled the EOSC onboarding conditions (namely, a Technology Readiness Level of 7 or more), one is currently onboarded and can be viewed at https://marketplace.eosc-portal.eu/services/phytoplankton_eovs. The onboarding of the other services with TRL7 or more (nine so far, as listed in section 4.3.8) will follow in the upcoming weeks, once the tool developed and described in this deliverable is adapted to the recent, new EOSC metadata profile version #4 which was released in 06/2022. The remaining the Blue-Cloud services can be onboarded as soon as they reach TRL 7 (two so far are not ready yet).

While working in this deliverable, the Blue-Cloud team worked close with the EOSC onboarding team, testing and using the Onboarding Web GUI and corresponding HTTP API and the EOSC Providers Documentation, pointing out those aspects of the process where documentation could be clearer to the providers. This way, the activities described in this deliverable and feedback shared with the EOSC onboarding team turned out to be beneficial for both parties, Blue-Cloud and EOSC.

2. The Blue-Cloud Catalogue

2.1. Description of the Catalogue

The Blue-Cloud Service Catalogue[3] lists all the resources and analytic services developed and provided in the scope of the Blue-Cloud project. It is reachable from the main Blue-Cloud website and it is hosted on D4Science at the link https://blue-cloud.d4science.org/catalogue-bluecloud. A screenshot is available in Fig. 1.

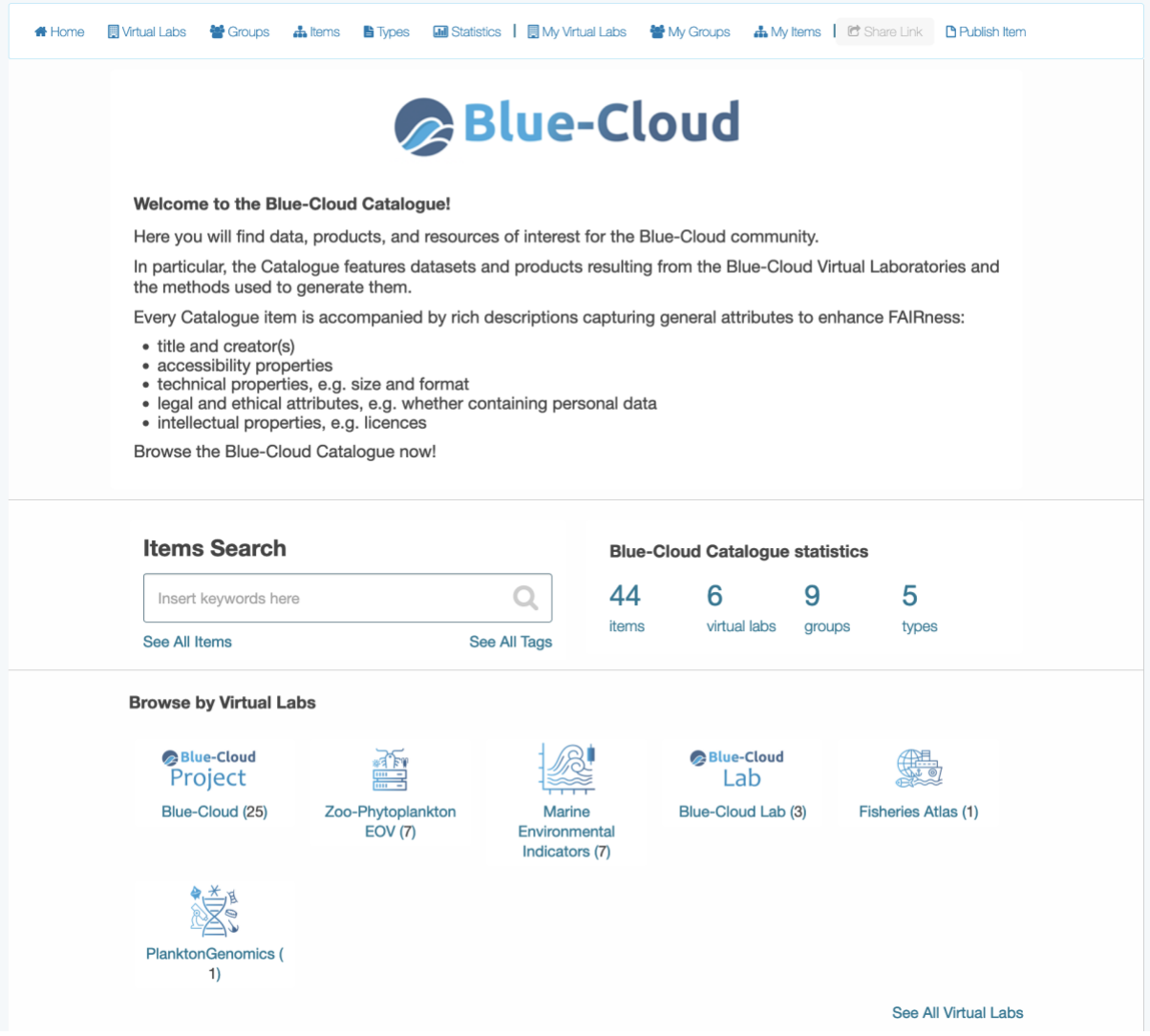


Figure 1 Screenshot of the Blue-Cloud service catalogue

.

[3] https://blue-cloud.d4science.org/group/bluecloud-gateway/data-catalogue

The Blue-Cloud catalogue has been developed by instantiating the catalogue service technology (Assante et al, 2020). The catalogue is supported by the data model presented in Fig. 2 and has been exploited to serve the needs of the Blue-Cloud project and community.

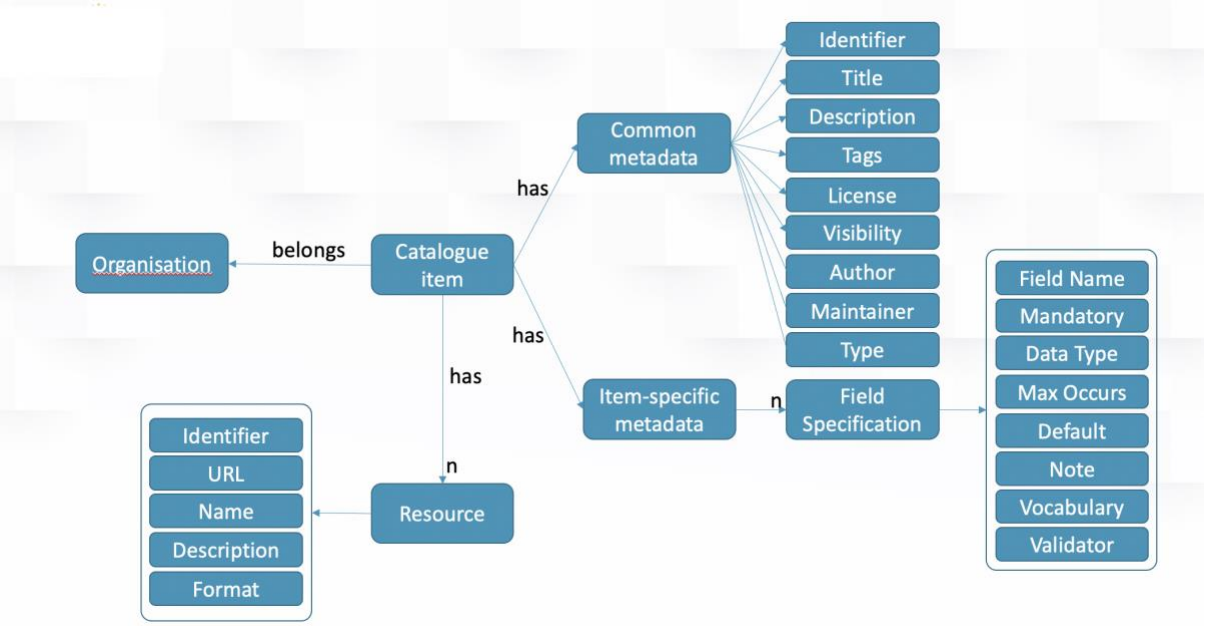


Figure 2 Catalogue Data Model

Catalogue items are published by "organisations". An ***organisation*** represents the context/authority responsible for the publishing of the item. Organisations are usually paired with VLabs for the items stemming from them or with other existing contexts in the case of pre-existing VLabs.

This catalogue serves all the virtual laboratories the Blue-Cloud VRE provides support for. In this initial phase, 6 organisations have been created:

- Blue-Cloud
- Blue-Cloud Lab
- Fisheries Atlas
- Marine Environmental Indicators
- Plankton genomics
- Zoo-Phytoplankton EOV

In particular, the **Blue-Cloud organisation** is the organisation responsible for the publication of resources produced by the consortium for its operation, such as Deliverables and Training Materials. The **Blue-Cloud Lab organisation** is the organisation delivering general purpose services that are offered in open-access mode to researchers and scientists beyond the boundaries of the Blue-Cloud consortium. Demonstrators have their own environment to publish resources directly in the Blue-Cloud catalogue, too.

The Blue-Cloud catalogue supports five types of **items:** Provider, Dataset, Service, Deliverable and Method. In particular, *Provider* and *Service* are compliant with the profiles promoted by EOSC: the Provider catalogue item is compliant with the EOSC Provider profile[4] while Service catalogue item is compliant with the *EOSC Resource* profile[5].

A *Service* describes an online technology that is used to transmit, process, store, create, display, share or exchange information. A *Service* is provisioned by a *Provider* that is the entity responsible for its operation. For example, the Blue-Cloud Lab is the *Organisation* providing access to the RStudio *Service* operated by the D4Science Infrastructure *Provider*.

A *Method* describes a computational script that requires a *Service* to be executed. A *Dataset* describes a collection of data either imported or generated via a Service. A *Deliverable* describes a report.

Every item of the catalogue is characterised by a set of ***common metadata,*** including a unique identifier, a title, a description, a list of tags (e.g., keywords, subjects), a licence, visibility (whether the item is publicly available or visible only to the members of a VLab), an author, a maintainer, and a type. Every Type is characterised by a specific set of attributes, controlled vocabularies and formats carefully describing the specific class of items.

Service and *Provider* typologies are inherited by EOSC; *Dataset* and *Method* are inherited by the SoBigData Infrastructure[6]; *Deliverable* has been defined by Blue-Cloud.

The common metadata for each of the catalogue items are documented in the following tables. Each table reports the classes of attributes characterising each metadata profile and the specific elements belonging to the class.

The **Provider** metadata profile (cf. Tab. 1) consists of 41 elements (15 of which are mandatory) organised in 8 classes. Moreover, 14 elements have a controlled vocabulary to ease their filling, e.g. scientific domains and subdomains are compiled by relying on the Frascati framework.

Metadata classes	**Metadata elements**
Basic Information	ID*, Name*, Description, Abbreviation*, Website*, Legal Entity*, Legal Status,
Marketing Information	Logo*, Multimedia
Classification Information	Scientific domains & subdomains, Tags
Location Information	Address, City, Region, Country
Contact Information	Main contact, public contact
Maturity Information	Lifecycle status

[4] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-provider-profile
[5] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile
[6] SoBigData: European Research Infrastructure for Big Data and Social Mining, www.sobigdata.eu

Other Information	Hosting legal entity, Participating countries, Affiliations, Networks, Structure type, ESFRI domain and type, MERIL scientific domain & subdomain, Areas of activity, Societal grand challenge, National roadmaps

Table 1 Provider Profile classes and elements

The **Service** metadata profile (cf. Tab. 2) consists of 61 elements (22 of which are mandatory) organised in 11 classes. Moreover, 15 elements have a controlled vocabulary to ease their filling, e.g. categories and subcategories are compiled by relying on the list of possible values identified by EOSC to identify typologies of services.

Metadata classes	**Metadata elements**
Basic Information	ID*, Name*, Description, Resource Organization*, Resource Provider, Webpage*
Marketing Information	Tagline*, Logo*, Multimedia, Use cases,
Classification Information	Scientific domains* & subdomains*, Categories* and subcategories*, Target users*, Access type, Access mode,
Availability Information	Geographical availability*, Language Availability*, Resource Location
Contact Information	Main contact*, public contact*, helpdesk email*, security contact email*
Maturity Information	TRL*, Lifecycle status, Certifications, Standards, Open source technologies, Last update, Change log,
Dependencies Information	Related resources, Related platforms
Attribution Information	Funding bodies, Funding programs, Projects
Management Information	Helpdesk, User manual, ToU, Privacy policy, Access policy, Service level, Training information, Status monitoring, Maintenance
Access & Order Information	Order type*, Order
Financial Information	Payment model, Pricing

Table 2 Service Profile classes and elements

The **Dataset** metadata profile (cf. Tab. 3) consists of 35 attributes (10 of which are mandatory) organised in 7 classes. Moreover, 14 fields have a controlled vocabulary to ease their filling, e.g. availability is either “on-line” or “on site”, processing degree is either “primary” or “secondary”, language is in ISO 639-3.

Metadata classes	**Metadata elements**
Basic Information	ID*, Name*, Description*
Identity Information	External Identifiers, Creators*, Creation date*, Owners, Related papers
Coverage Information	Semantic coverage, Time coverage, Spatial coverage
Access Modality	Accessibility, Availability, Accessibility mode
Technical Details	Processing degree*, Manifestation type*, Language, Size, Disk size, Format, Format schema
Rights Details	A list of 8 elements to report detailed information about the policies governing the exploitation of the dataset, e.g., territory of use, license term
Attribution Details	A list of 3 elements to report detailed information about the text to use to acknowledge the usage of the dataset
Data Protection Details	A list of 6 elements to report the characteristics of the dataset with respect to personal and sensitive characteristics and management policies

Table 3 Dataset Profile classes and elements

The **Method** metadata profile (cf. Tab. 4) consists of 22 attributes (5 of which are mandatory) organised in 7 classes. Moreover, 6 fields have a controlled vocabulary to ease their filling, e.g. access mode is either “as-a-Application via Blue-Cloud Infrastructure”, “as-a-Application via third-party Infrastructure”, “as-a-Service via Blue-Cloud Infrastructure”, or “as-a-Service via third-party Infrastructure.

Metadata classes	**Metadata elements**
Basic Information	ID*, Name*, Description*
Identity Information	External Identifiers, Creators*, Creation date*, Owners, Related papers

Coverage Information	Semantic coverage
Access Modality	Usage mode*, Availability
Technical Details	Hosting environment, Programming language, Dependencies, input, output
Rights Details	A list of 7 elements to report detailed information about the policies governing the exploitation of the method, e.g., field of use, embargo period
Attribution Details	A list of 2 elements to report detailed information about the text to use to acknowledge the usage of the method

Table 4 Method Profile classes and elements

The **Deliverable** metadata profile (cf. Tab. 5) consists of 11 attributes (9 of which are mandatory) organised in 3 classes. Moreover, 2 fields have a controlled vocabulary to ease their filling, e.g., dissemination level is either "PU: Public", "PP: Restricted to other programme participants (including the Commission)", "RE: Restricted to a group specified by the consortium (including the Commission)" or "CO: Confidential, only for members of the consortium (including the Commission)".

Metadata classes	**Metadata elements**
Basic Information	ID*, Title*, Description*
Identity Information	Deliverable Identifier*, Deliverable Authors*, Deliverable Contributors*, Publication date*, Owners, Related papers
Dissemination Information	Dissemination level*, Status*

Table 5 Deliverable Profile classes and elements

2.2. The Blue-Cloud onboarding procedure

This section describes in detail the procedure to register items into the Blue-Cloud catalogue.

Users having the 'Catalogue Editor' role are entitled to publish items into the catalogue. This role is assigned per VLab. For instance, one user could be authorised to publish items exclusively to the virtual laboratory "Fisheries Atlas", whereas another user could be authorised to publish items to the two virtual laboratories "Blue-Cloud" and "Blue-Cloud Lab".

Users having the 'Catalogue Admin' role are notified about new published items and they can review, edit and either approve or reject them. Also, this role is assigned per VLab ensuring complete autonomy in the management of the catalogue of each virtual environment.

The publication of the items can be performed via either the Workspace user interface, the Catalogue user interface, or the gCat REST service APIs[7].

The first two modalities open a wizard organised in three sections.

The first section of the Wizard is common to all different types and allows the user to specify the common metadata (see Fig. 3).

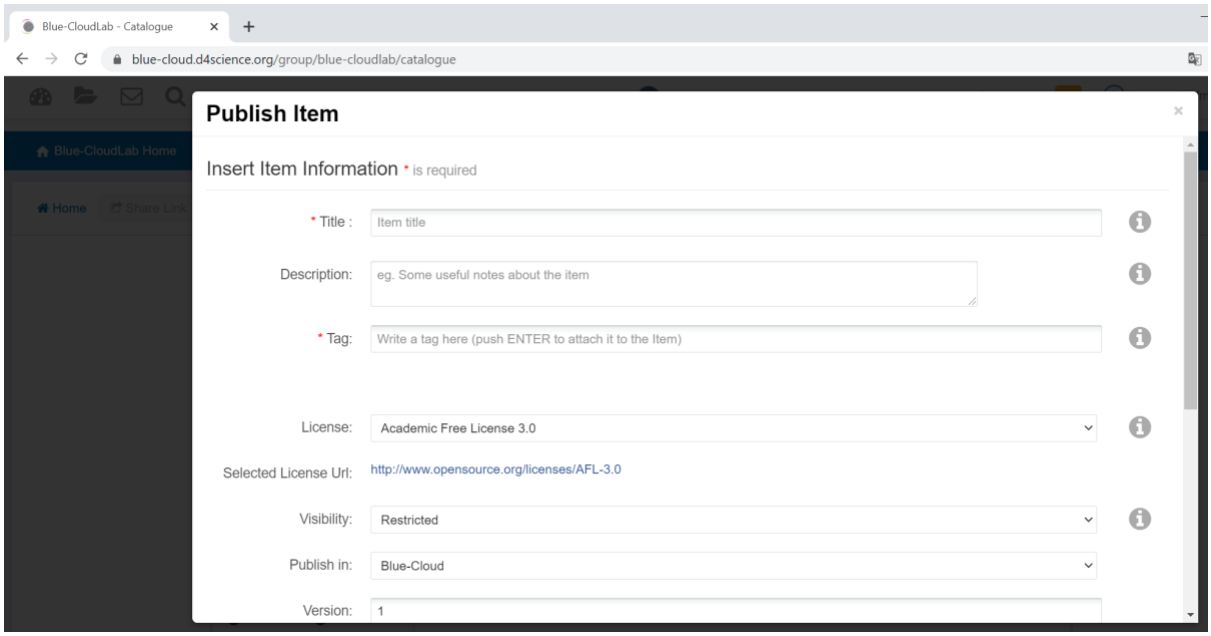


Figure 3 Publish an item - Common metadata.

The second section is specific to the selected Type (see Fig. 4).

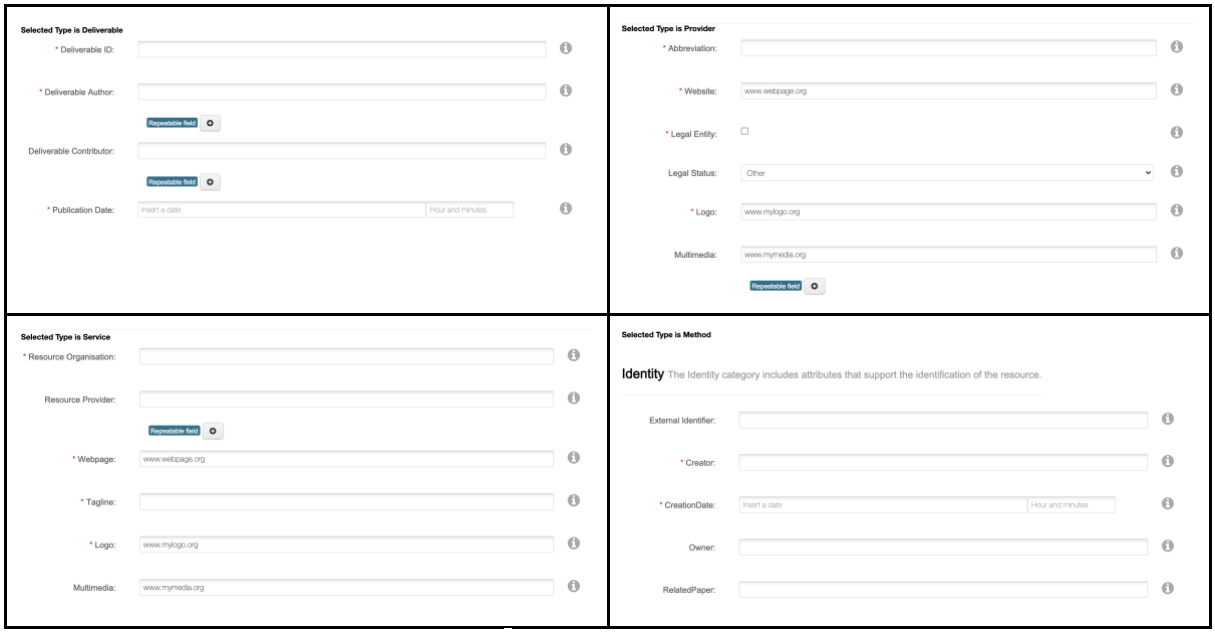


Figure 4 Publish an item - Type specific metadata.

[7] gCat REST APIs - https://wiki.gcube-system.org/gcube/GCat_Service

The third section of the Publication wizard allows the user to add one or more manifestations of the resource. A manifestation is the payload of the item and it is characterised by the identifier, a name, a description, a format, and, most importantly, the URL pointing to the content.

2.3. The Blue-Cloud Catalogue organisational model

The Blue-Cloud project federates and pilots innovative services for the marine research and the Blue economy communities. Its catalogue includes services and resources that existed before Blue-Cloud and are operated by third-party e-infrastructures, as well as new services and resources that are generated by the Blue-Cloud project consortium. To model this dichotomy in the Blue-Cloud catalogue, different Providers are being registered starting from the following two:

- the **Blue-Cloud Provider** is the Resource Organisation, i.e., "*the entity that manages or delivers the services and resources generated by the project, or that coordinates the delivery of those services and resources in a federated scenario*"[8];
- the **D4Science Provider** is the Resource Provider, i.e., the entity that "*that manage[s] or deliver[s] the Resource in federated scenarios*", which in Blue-Cloud context means that it manages and delivers the enabling services and technology exploited by the virtual laboratories. For example, the services offered by the Blue-Cloud Lab - namely RStudio, JupyterNotebook, Analytics Engine and the Software Importer – existed before the Blue-Cloud project but their delivery is coordinated by the project.

Finally, each VLab registers the services implemented as project demonstrator by specifying Blue-Cloud as Resource Organisation and D4Science as Resource Provider. Currently, no other e-infrastructure (such as WEkEO or EUDAT) provides services in the Blue-Cloud VLabs, so this deliverable is restricted to the two providers mentioned above.

[8] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile

3. The EOSC Catalogue and onboarding procedure

3.1. Description

The EOSC Service Catalogue is a listing of services targeted towards open research. These resources can be browsed in the EOSC Marketplace, an "*integrated platform that allows easy access to lots of resources for various research domains along with integrated data analytics tools*"[9].
This section describes the onboarding procedure into the EOSC Catalogue. In section 4 it is described how the actual execution of this onboarding was carried out in the scope of Blue-Cloud task 4.5.[10]

3.2. Definition of EOSC Providers and EOSC Resources

Central concepts in this and the next sections are those of *EOSC Provider* and *EOSC Resource*. These concepts are defined as follows:

- *An EOSC Provider*[11] *is an EOSC System User responsible for the provisioning of one or more Resources to the EOSC. EOSC Providers are organisations, a part of an organisation or a federation that manages and delivers Resources to End-Users. EOSC Providers can be: Resource Providers, Service Providers, Data (Source) Providers, Service Developers, Research Infrastructures, Distributed Research Infrastructures, Resource Aggregators, Thematic Clouds, Regional Clouds, etc.*
- *An EOSC Resource*[12] *is an asset made available by means of the EOSC system and according to the EOSC Rules of Participation to EOSC End-Users to perform a process useful to deliver value in the context of the EOSC. EOSC Resources include Services, Data Sources and any other asset. A Resource Profile describes the information requested to onboard Resources into the EOSC Provider Portal.*

In the EOSC Catalogue, providers can have two **roles** in relation to resources (e.g. services):

- the role of a *Resource Organisation* or
- the role of a *Resource Provider*.

Every resource has one *Resource Organisation* and none to several *Resource Providers*. All of them are *Providers* in EOSC terminology, are described according to the *Provider Profile* (see next section), and have a *Provider ID* by which they are linked to the resources. The *Resource Organisation* is "*the*

[9] https://marketplace.eosc-portal.eu/

[10] Although EOSC also aims at making research data accessible across Europe, datasets may not be found directly in the Marketplace. However, the Marketplace lists services that allow the discovery of research datasets. As this deliverable focuses on the interfacing of services and methods, the discovery of Blue-Cloud data is covered in a different task (2.3) and deliverable (D2.5)

[11] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-provider-profile

[12] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile

organisation that manages or delivers the Resource, or that coordinates Resource delivery in a federated scenario" while *Resource Providers* are those entities *"that manage or deliver the Resource in federated scenarios"*[13] .

3.3. EOSC Profiles

The services listed in the EOSC Service Catalogue as well as their providers are described by structured and well-defined metadata. The metadata items are listed and defined in the EOSC Profiles[14]. Their documentation provides examples and guidance on every metadata item.

For describing a service provider, the EOSC Provider Data Model must be used:
https://eosc-portal.eu/providers-documentation/eosc-provider-portal-provider-profile
It lists the mandatory and optional metadata, as well as controlled vocabularies (code lists, taxonomies, classifications) for items such as legal statuses, areas of activity, countries, scientific domains, consortia, etc. Currently, 15 items are listed as mandatory and 36 are optional.

For describing a service, the EOSC Resource Data model must be used:
https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile
Currently, 22 items are listed as mandatory and 39 are listed as optional.

Note: The EOSC catalogue was in the process of switching from profile version 3 to profile version 4 while this deliverable was being written. In the meantime, profile 3, according to which the actions described in this deliverable were implemented, has become outdated, which implies a required upgrading in the coming weeks.

3.4. Preconditions for onboarding

There are a number of preconditions that services must comply with before they can be part of the EOSC Catalogue and Marketplace. First of all, as Open Science is the primary goal of the EOSC, all onboarded services have to enable, enhance and comply with the principle of Open Science.

EOSC defined a set of Rules of Participation that the services and their providers have to comply with[15,16]. These rules mainly address high-level general concepts such as openness, FAIR-ness, policies and legislation. The services also have to be technically mature and be classified at least Technology Readiness Level 7 (as defined by the European Commission[17]), which means that a

[13] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile
[14] https://eosc-portal.eu/sites/default/files/EOSC-Profiles-v3.00.pdf
[15] https://www.eosc-hub.eu/key-exploitable-results/eosc-rules-participation-rop
[16] https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184432576
[17] https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

prototype demonstrator of the system must be running in an operational environment. Also, the services have to be properly and correctly described with service metadata. More details on the preconditions can be found in the EOSC-Hub Integration Handbook (Sipos et al. 2020)[18], on the EOSC documentation for providers[19] and also in the EOSC wiki[20].

3.5. The onboarding procedure

There are currently two ways of onboarding providers and services into the EOSC Catalogue and for updating the information once it is integrated: Using a form/GUI or using an HTTP API. As mentioned above, bulk-onboarding an existing compatible catalogue is not available yet.

3.5.1. Initial onboarding of resource providers and services

At first, a provider has to apply for becoming an EOSC Provider. This process is documented in detail at https://eosc-portal.eu/for-providers (see EOSC-Hub Integration Handbook, Sipos et al. 2020)[21]. A new resource provider can be registered using the form at https://providers.eosc-portal.eu/provider/add. Once this information is reviewed by the EOSC onboarding team, detailed *Service Description Templates* are created and have to be filled and handed in by the Blue-Cloud service providers. The information contained in these Service Descriptions is used by EOSC to determine whether the provider's services offered fulfil the preconditions, such as the Rules of Participation (RoP)[22]. It is also used to populate the EOSC Service Portfolio. After review and approval by both sides - EOSC and the service providers - the descriptions are published and are now part of the public EOSC Service Catalogue.

3.5.2. Onboarding/updating via EOSC Catalogue HTTP API

EOSC provides an open HTTP API to manage the provider and service descriptions in the Service Catalogue. The API is documented here: https://providers.eosc-portal.eu/openapi. In this deliverable, reference is made to the version 3.0.0 of the API (as of 2022-04-06).

The API provides endpoints to update provider information, and to create and update service information. In the following subsections, the various actions that can be taken are described.

[18] https://www.eosc-portal.eu/sites/default/files/EOSC-hub%20Integration%20Handbook%20for%20Service%20Providers.pdf

[19] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-inclusion-criteria

[20] https://wiki.eosc-hub.eu/display/EOSC/Criteria+for+possible+inclusion+in+the+EOSC+Service+Portfolio

[21] https://www.eosc-portal.eu/sites/default/files/EOSC-hub%20Integration%20Handbook%20for%20Service%20Providers.pdf

[22] https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184432576

The providers still have to be onboarded beforehand, as their integration has to be reviewed and approved by the EOSC onboarding team. However, integrating the individual services and keeping their descriptions up-to-date in the EOSC Catalogue becomes easier and more automatic with the HTTP API. An automatic tool can be deployed that feeds changes in the metadata automatically into the EOSC Catalogue in near-real time. Obviously, the metadata about the services still has to be created by the service providers and it has to be stored someplace - ideally the provider's own service catalogue - for such a tool to pick it up and push it into EOSC Catalogue.

For any of the operations described below, the client performing them has to be authorized. The authorization mechanism adopted in the EOSC Catalogue API is the API key mechanism[23], where the client provides a token in the request header of an HTTP request. This token is supplied by the operator of the API, i.e. by EOSC.

- **Updating a provider**

  The HTTP endpoint `/provider` allows the client to update the information about the provider, by uploading a JSON document containing the updated information via `HTTP PUT`. The JSON document must contain the updated metadata in a specific format. For examples and documentation, please refer to https://providers.eosc-portal.eu/openapi. The types and values are described in the "Provider Data Model", presented in section on EOSC Profile (section 3.3). The initial integration of a provider is not possible using this API and has to be done as described in section 3.5.1.

- **Creating a resource**

  The HTTP endpoint `/resource` allows the providers to create a new resource, e.g. a service, by uploading a JSON document containing the new information via `HTTP POST`. The JSON document must contain the metadata in a specific format. For examples and documentation, please refer to https://providers.eosc-portal.eu/openapi. The types and values are described in the "Resource Data Model", presented in section on EOSC Profile (section 3.3).

- **Updating a resource**

  The HTTP endpoint `/resource` allows the providers to update an existing resource/service based on its id, by uploading a JSON document containing the updated information via `HTTP PUT`.

- **Validating a resource**

  The endpoint `/resource/validate` allows the providers to validate the metadata via `HTTP POST`. This means that the service metadata is posted, and the response will contain an error message if the metadata does not conform to the EOSC profiles, e.g. "*The URL 'http://www.dummy.org' you provided is not valid. Found in field 'useCases'*".

- **Vocabulary endpoint**

[23] https://swagger.io/docs/specification/2-0/authentication/api-keys/

The endpoint `/vocabulary` allows to retrieve controlled vocabularies, such as category at https://beta.providers.eosc-portal.eu/api/vocabulary/byType/CATEGORY. This is useful for mapping, because the Blue-Cloud catalogue contains the full names of items, while the EOSC catalogue expects the ids. The endpoint provides both, which can be used for mapping:


```
[
    {
        "id":"funding_body-ec",
        "name":"European Comission (EC)",  [sic!]
        "description":null,
        "parentId":null,
        "type":"Funding body”,
        "extras":{}
    },
    {
        "id":"funding_body-bmbf",
        "name":"Federal Ministry of Education and Research (BMBF)",
        "description":"Germany",
        "parentId":null,
        "type":"Funding body",
        "extras":{}
    },
    … etc. …
]
```


3.5.3. Alternative: Bulk-onboarding the entire catalogue

There are plans to enable the onboarding of entire Service Catalogues if their resources are compatible with the EOSC Service Catalogue:

> *"The EOSC Portal team (...) is implementing a method to import and share whole service catalogues from in one-go [sic!], simplifying the service onboarding for provider consortia."*
>
> (EOSC-Hub Integration handbook for service providers, Sipos et al. 2020).

The resource provider still has to be onboarded first, as its integration has to be reviewed and approved by the EOSC onboarding team. However, the integration of the individual services becomes easier and more automatic. Of course, the services have to be fully described in the original catalogue for the descriptions to be imported into the EOSC Catalogue.

At the time of drafting this deliverable, this functionality was not available yet, but very recently it has been implemented with the new Provider and Service profile (v4.0) (end of June 2022). Blue-Cloud might make use of this new functionality in the near future to allow onboarding of whole resource catalogues in one go.

4. Onboarding Blue-Cloud services into the EOSC Catalogue

The creation of the EOSC-compatible Blue-Cloud catalogue was an important basis for the EOSC integration. In this context, “compatible” means that the Blue-Cloud catalogue creators made sure that every mandatory (and many optional) metadata items are included in the Blue-Cloud catalogue, with similar names and matching definitions. Based on it, it is mostly straightforward to transfer the created resources (i.e. the descriptions of the services) into the EOSC Catalogue. First, the service providers are registered in the EOSC Catalogue. Once the providers are integrated, the next steps performed in this task focus on the integration of the services.

4.1. Integration of the service providers

As mentioned in Section 2.3, the services generated in the Blue-Cloud project will be provided by the Resource Organisation *Blue-Cloud*, while *D4Science* is the Resource Provider that manages and delivers the enabling services and technology exploited by the virtual laboratories.

D4Science is already present as a Provider in the Blue-Cloud catalogue as can be seen in Figures 5 and 6. Moreover, in Fig. 7, D4Science is also already integrated into the EOSC Catalogue as a provider for various non-Blue-Cloud services[24], so there is no need to integrate D4Science. Blue-Cloud itself was not yet integrated as a Provider in EOSC before the beginning of this task. So, the first step was to create and integrate the Blue-Cloud service provider (see next section), which will be the main service provider in the Blue-Cloud project besides D4Science.

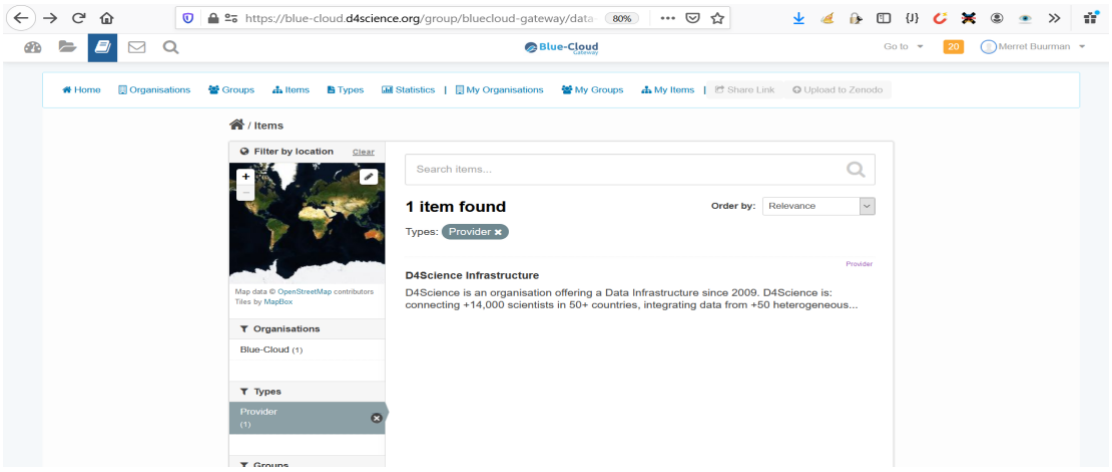


Figure 5 D4Science is listed as a provider in the Blue-Cloud catalogue

24 https://marketplace.eosc-portal.eu/services?geographical_availabilities=&order_type=&providers%5B%5D=30&providers-all=&providers-filter=&rating=&related_platforms-all=&related_platforms-filter=&scientific_domains-all=&scientific_domains-filter=&target_users-all=

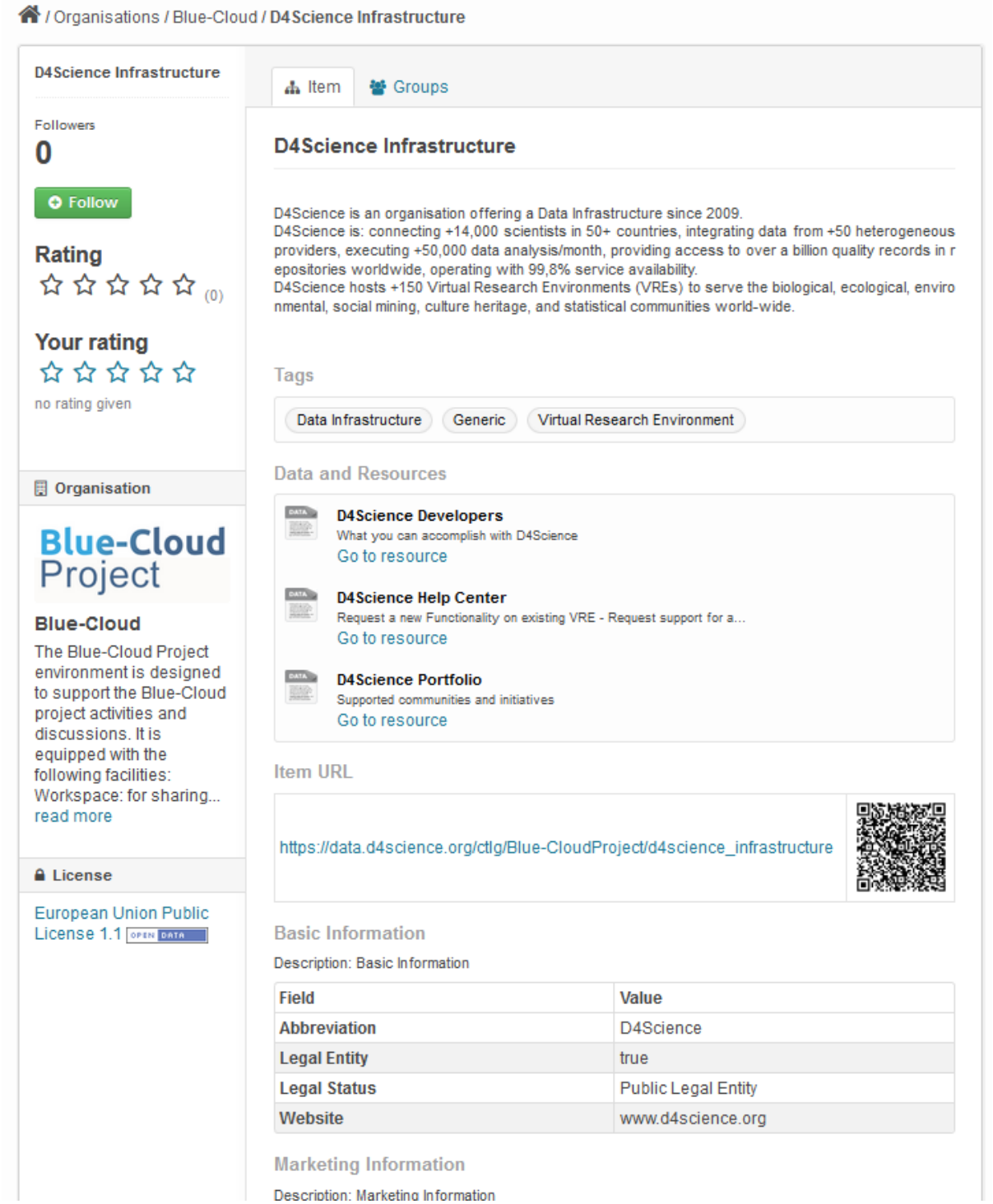


Figure 6 Information about the provider D4Science in the Blue-Cloud catalogue.

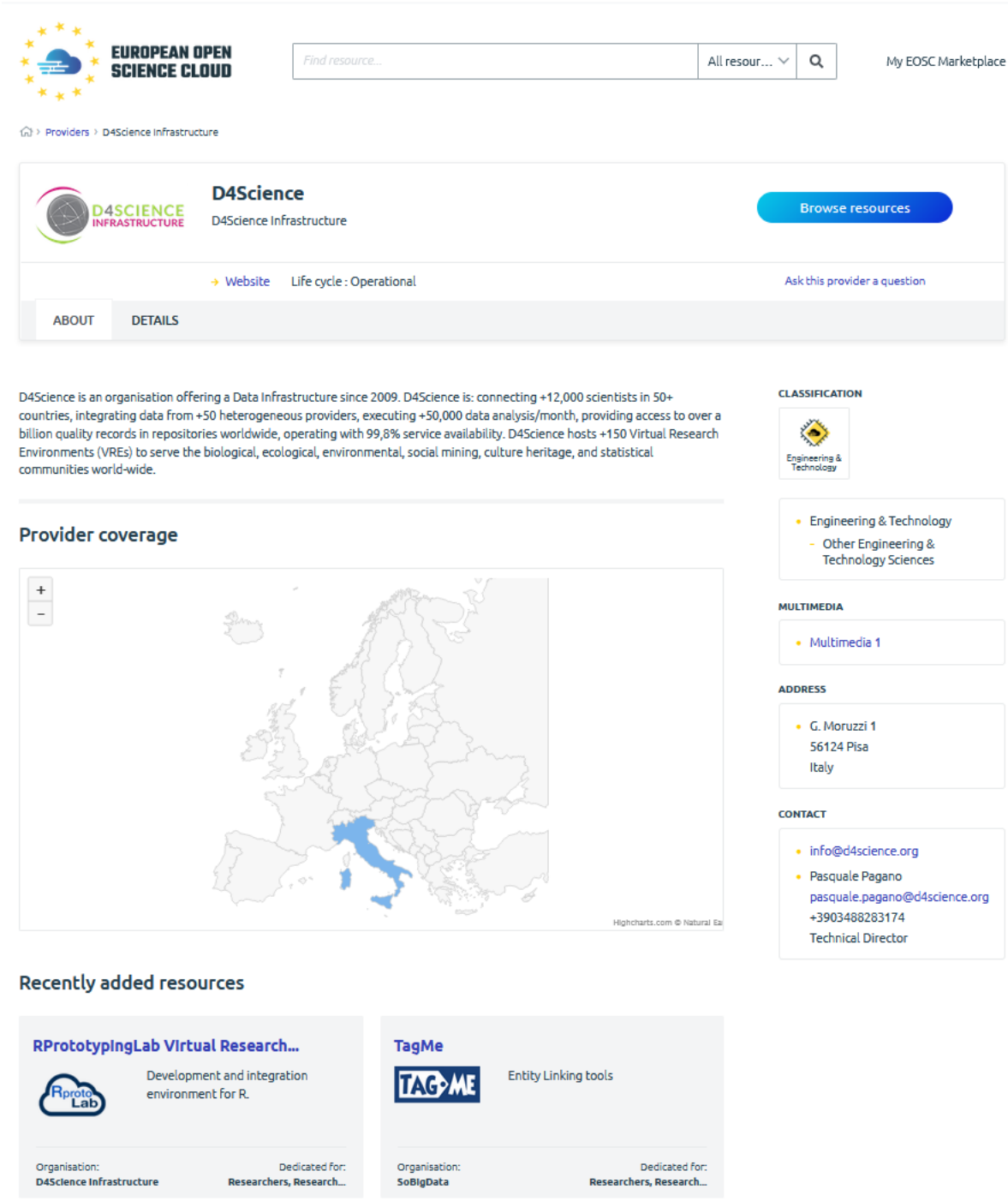


Figure 7 D4Science is already a Provider in the EOSC Catalogue

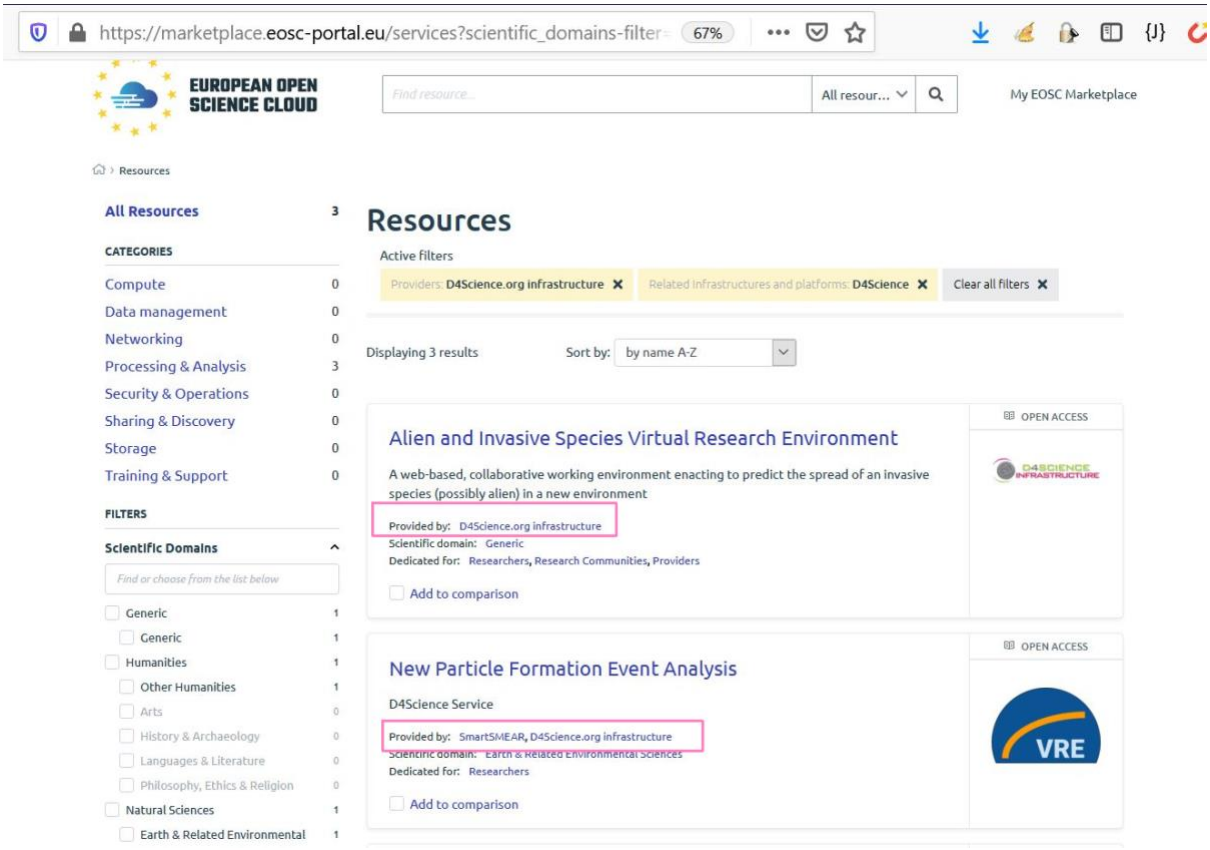


Figure 8 Resources already provided by the Provider D4Science in the EOSC Catalogue.

4.2. Onboarding the provider “Blue-Cloud”

The onboarding of an EOSC Provider is based on a process consisting of:

- the registration/login of an authorised representative of the Provider into the EOSC Portal;
- the authorised representative applies to onboard the provider by filling in a form with all the information characterising the provider (including other users playing the role of authorised representatives);
- The EOSC Portal Onboarding Team reviews the newly onboarded provider and when the provided information is suitable the provider profile enters into the “approved” status.

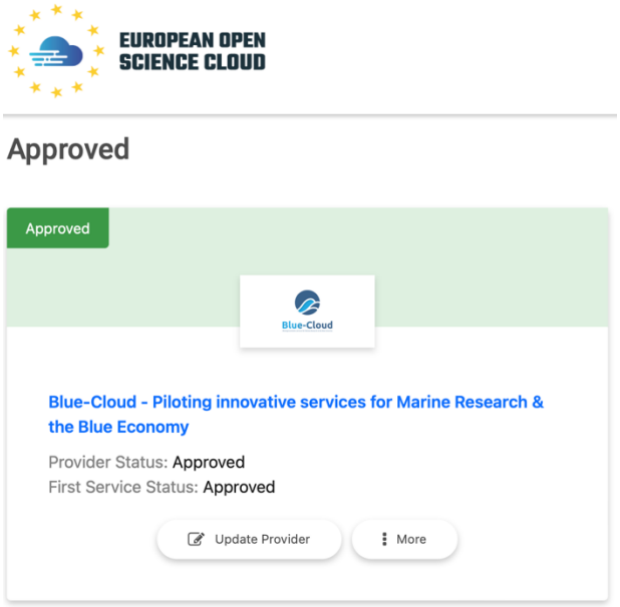


Figure 9 The Blue-Cloud service was approved as a provider in EOSC.

4.3. Onboarding of the services

The main goal of the EOSC integration is to make the Blue-Cloud processing services visible and usable to a wider audience. Thus, the integration of the services into EOSC is the main activity to be completed in this task (task 4.5).

First, the services were thoroughly described by the service owners and included in the Blue-Cloud service catalogue (see section 2). These descriptions can be viewed at the Blue-Cloud catalogue at https://blue-cloud.d4science.org/catalogue-bluecloud (scroll to "Browse by type" and click "Services"). This work heavily depends on the readiness of the services developed in the Blue-Cloud consortium as well as of the service owners' cooperation in providing service metadata.

Once the first set of services were described and included in the Blue-Cloud catalogue, the onboarding process into the EOSC Catalogue could be started. The descriptions of those services that are ready to be offered to a broader user base and comply to the EOSC preconditions (described in section 3.4, i.e. classified at least Technology Readiness Level 7, etc.) are transferred into the EOSC Catalogue. This is ongoing work, due to the recent upgrading of the EOSC catalogue which surpassed the finalisation of the onboarding process, as mentioned above. However, it will be finalised in the coming weeks, while the remainder of this deliverable describes the preparations that were already made and which require upgrading.

In the future, the preferred method for this transfer will be the bulk-import described in section 3.5.3. However, this option was not yet available at the time of writing this deliverable. Furthermore, even though as the metadata standard used by Blue-Cloud catalogue is closely related to using the EOSC Profiles to describe their provider and services, it is uncertain whether the required level of compatibility is provided. Thus, activities will be undertaken by the Blue-Cloud to

evaluate if this very recently added is suited. In that case, the Blue-Cloud would like to make use of it and directly onboard the entire Blue-Cloud catalogue (or the subset of services which are considered mature enough). Obviously, as mentioned before, the services have to be fully described in the Blue-Cloud catalogue for the descriptions to be imported into EOSC.

As the content of the Blue-Cloud service catalogue could not yet be onboarded in bulk-mode at the time of writing, it was decided to set up a small service program that will leverages the EOSC HTTP API (see section 3.5.2) and the Blue-Cloud catalogue's HTTP API to regularly retrieve and push changes to service descriptions that were made in the Blue-Cloud catalogue to the EOSC Catalogue.

The whole process is described below.

4.3.1. Service description and Blue-Cloud catalogue integration

Onboarding of services into the Blue-Cloud catalogue was facilitated by the VLab Catalogue instances. Service owners were provided with the role of "catalogue editor" enabling the publishing of service descriptions in the specific application context. In practice, in every VLab the service owners onboarded the services offered by the VLab including the VLab on its own. Services that were non-specific of a VLab, e.g. the Data Discovery and Access Service, were onboarded by using the VLab supporting the project activities, namely the Blue-CloudProject VLab. The figure below showcases the services onboarded into a specific VLab. Services onboarded in a specific VLab are discoverable and available by the overall Blue-Cloud catalogue made available by the Blue-Cloud VRE.

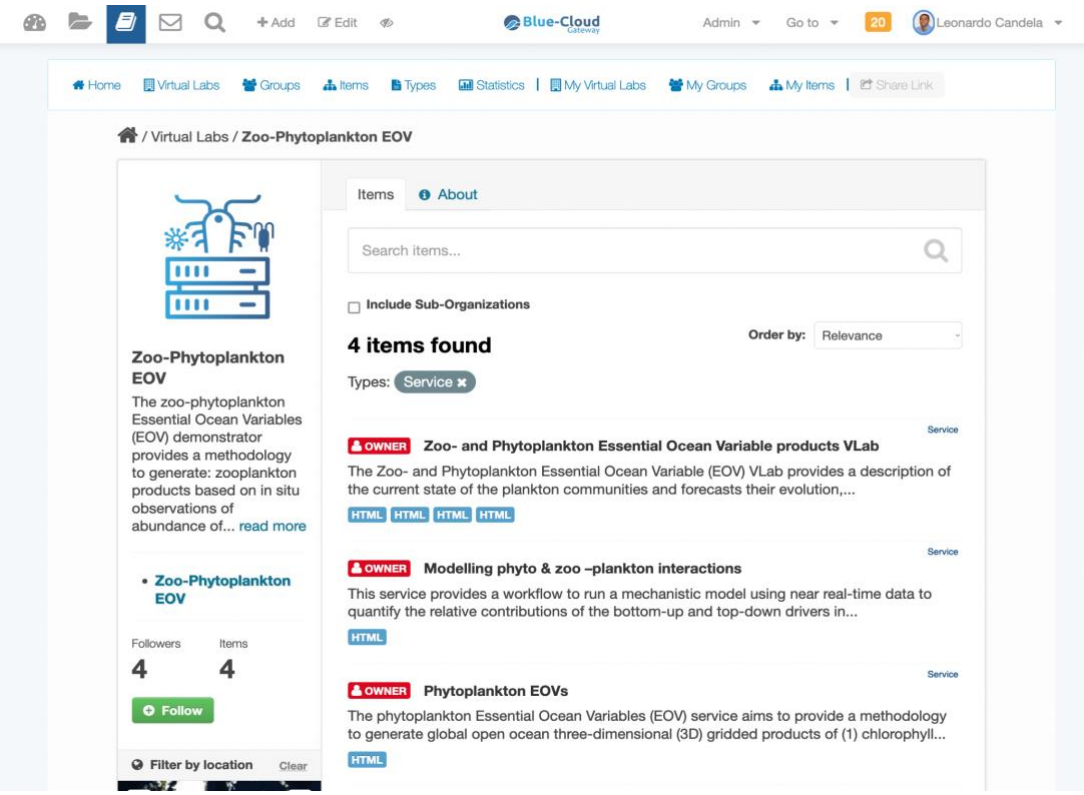


Figure 10 Catalogue view of Services onboarded into the Zoo-Phytoplankton EOV

In order to onboard services into the EOSC Catalogue, service owners are requested to fill in a form enabling to specify the expected information characterising each service ranging from generic information like name of the service and description up to the information requested to characterise services from the EOSC point of view, namely: marketing information including service logo and tagline; classification information including scientific domain, category, and target users; contact information including main contact and helpdesk email; maturity information including TRLness and life cycle status; availability information including language and geographic availability; attribution information including funding body and project supporting the development and operation of the service; management information including links to service terms of use and privacy policy; access and order information including details on how the service is made available. Whenever expected and envisaged the form provides the service owner with allowed values (e.g. scientific domain) to fill in certain fields as well as the form controls that the inserted values match the constraints imposed by the EOSC profiles (e.g. regarding the maximum allowed length for certain texts). Figure below showcases the Blue-Cloud catalogue form.

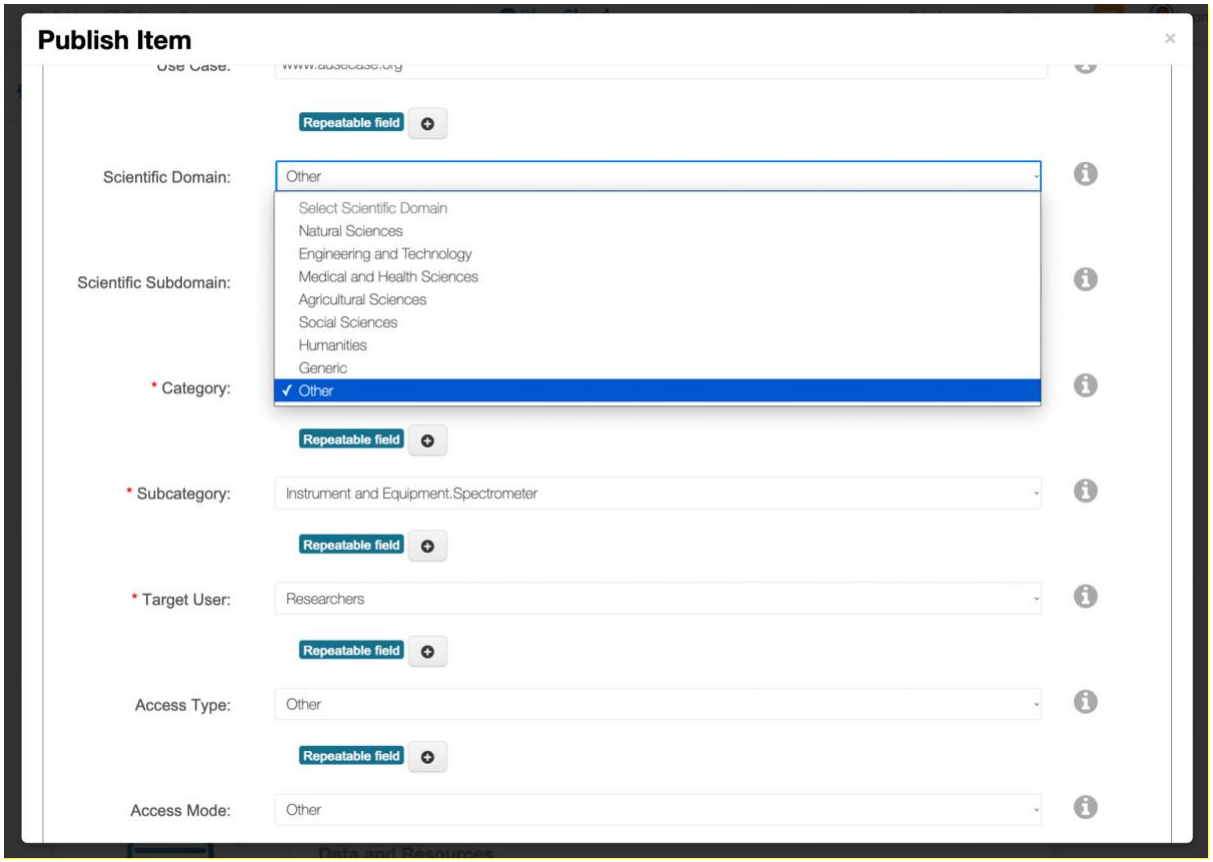


Figure 11 Blue-Cloud Service Onboarding Form

4.3.2. Retrieval of service metadata from the Blue-Cloud catalogue

As explained above, it was decided to set up a small service program that leverages the EOSC HTTP API (see section 3.5.2) and the Blue-Cloud catalogue's HTTP API to regularly retrieve and push metadata changes that were made in the Blue-Cloud catalogue to the EOSC Catalogue.

The program uses the protocol HTTP to retrieve a list of services for each VLab from the Blue-Cloud catalogue. Each of these service names are then used to retrieve the service's metadata in JSON format. Due to the good usability and documentation of the Blue-Cloud API, and the support by the Blue-Cloud team at CNR, this task is straightforward.

4.3.3. Mapping of the service metadata

As the metadata items have slightly different names, values and format, and as the EOSC resource profile only requires a subset of the available metadata items, the Blue-Cloud Metadata then has to be carefully mapped to the EOSC profile.

For example, the mandatory item "Technology Readiness Level" is named "`MaturityInformation:Technology Readiness Level`" in Blue-Cloud catalogue, while it has to be named "`trl`" in EOSC JSON metadata. The value, in case of TRL 7, is "`TRL7 System prototype demonstration in operational environment`" in the Blue-Cloud catalogue, which is easily converted to simply "`TRL7`" to match the EOSC profile.

For this mapping process, a python module was developed that finds the appropriate names and values for each EOSC metadata item, verifies whether all mandatory items are present and whether they have a correct length and format, and then constructs the EOSC metadata, also in JSON format. For some simple and systematic cases, the mapping module can apply modifications, as in the case of technology readiness level mentioned above.

For various other items, the appropriate EOSC values have to be retrieved from controlled vocabularies. As an example, let's take the funding body. In Blue-Cloud this item has the key "`AttributionInformation:Funding Body`" with example values "`European Comission (EC)`". This translates to key "`fundingBody`" and value "`funding_body-ec`" in EOSC.

Further items where the valid values have to be fetched from controlled vocabularies include "fundingProgram", "targetUsers", "accessTypes", "accessModes", "lifeCycleStatus", "orderTypes", "scientificDomains" and "scientificSubdomains", "categories" and "subcategories", and "resourceOrganisation" and "resourceProviders".

Developing the mapping and validating the mapped metadata at the EOSC validation API was an iterative process in the course of which problematic on incompatible values were discovered (e.g. typos, exceeding lengths, special characters). Examples are described the next section on validation.

4.3.4. Validation using the EOSC validation API

The next step was to validate this mapped metadata using the EOSC validation API (see section 5.3.2). This verifies the adherence to the EOSC metadata standard in more depth than our preliminary checks and also helped us refine our mapping in the development stage.

Examples for malformed values that were identified in this stage were "`Category`" and "`Subcategory`". These two items must contain values that match predefined controlled

vocabularies[25], but in addition to that, the possible values of subcategory depend on the chosen category. In those cases where this had not been respected, the error message *"Subcategory is not contained in category: Data Analysis.Other not in Data Analysis"* was returned. To avoid this kind of problem in future, the problem was reported back to the Blue-Cloud catalogue team at CNR, who then envisaged implementing a constraint to enforce proper categorization.

Another example is the message *"Provider with ID 'xyz' does not exist. Found in field 'resourceProviders'"*. The fields "resourceOrganisation" and "resourceProviders" only accept valid EOSC Provider IDs of already onboarded Provider entities, which would be out of scope for the Blue-Cloud catalogue to verify beforehand. As we have a limited number of providers, our program can easily replace their names with the official EOSC provider IDs, retrieved from the EOSC API (https://api.eosc-portal.eu/provider/).

A very simple yet very time-consuming issue that was identified during validation is missing "http…" in front of URLs in metadata items such as "useCases", "webpage", "multimedia" and others. The validation failed, but the API did not return any error message in this case, so identifying the cause of the failure had to be done by trial and error.

For items which allow free text entries, is it more complicated to map them using a program. Thus, it was agreed together with the Blue-Cloud team at CNR that it is preferable if the values are already compatible in the originating catalogue. Whenever values did not match the EOSC profile, the corresponding values were modified in the Blue-Cloud catalogue. This was the case, for example, for too long taglines (EOSC allows up to 100 characters for "tagline") or inconsistent contact person names. While Blue-Cloud asks for the name as a single string (e.g. in "ContactInformation:Main Contact Name"), EOSC requires two separate strings ("firstName" and "lastName"). In general, the first option is preferable, as it allows for various types of names, while requiring a first and last name may not fit certain Non-European naming schemes. However, some service owners filled the name with "*John Doe*" and others with "*Doe, John*". As the comma-separated one is easier to parse into a firstname-lastname scheme, it was adopted for the Blue-Cloud catalogue and deviating values were adjusted.

During the development and testing of the mapping, some missing or malformed values were reported back to the service owners and subsequently added or modified. For example, some services had no information on "public contact email", "helpdeskEmail" and "securityContactEmail", which caused the service to fail any validation attempts. After opening a ticket in the Blue-Cloud ticketing system, this was taken care of. Some services had filled the "useCases" and "multimedia" fields with dummy URLs, which had to be corrected too.

Some typos in the Blue-Cloud catalogue (such as a missing S in the "subcategory" value "Application.Collaboration" - it has to be "Applications.Collaboration") are still pending. Other problematic EOSC vocabulary values include those that contain special characters, which the Blue-Cloud catalogue could not deal with. For example, the "scientificSubdomain" value

[25] https://eosc-portal.eu/providers-documentation/eosc-provider-portal-resource-profile#Resource%20Category,%20Subcategory%20(and%20Supercategory)

"`Natural Sciences.Earth and related environmental sciences`" expects a "`&`" instead of "`and`" in EOSC, and thus the mapping failed.

These cases are currently handled by the mapping script, which has been developed in a way to recognize and fix several known issues.

In other cases, mapping problems occurred that were due to inconsistent or contradictory documentation on EOSC side. For example, the "`orderType`" value "`Request/Order required`" fails validation, despite being listed as correct by the EOSC Profile description. The EOSC vocabulary endpoint, on the other hand, only lists the possible values "`Fully open access`", "`Open Access`", "`Order required`" and "`Other`". In this case, the service provider did not make a mistake, as they followed the Profile description, but the value to pass validation still has to be converted. These cases were reported to the EOSC onboarding team and it is hoped that this will be addressed in future Profile versions or in the documentation.

4.3.5: Onboarding of first service via GUI

As mentioned above, the first service must be onboarded using the GUI on the EOSC website. For this, the onboarding person must be logged in at the EOSC Portal and its account has to be assigned administration permissions for the Blue-Cloud provider.

A well-designed and easy-to-use GUI guides the onboarding person step by step through the various areas of metadata. Where values are restricted to Controlled Vocabularies, drop-down menus assist to find the appropriate values. The process can be paused, the filled fields saved as a draft for later. This allows various persons to log in from various sites and work on / review the draft.

When all values are filled, the metadata can be submitted to the EOSC onboarding team by clicking a button. Then, its status changes to "pending". The EOSC team reviewed the resource, suggested a few improvements or clarifications, and then approved the resource.

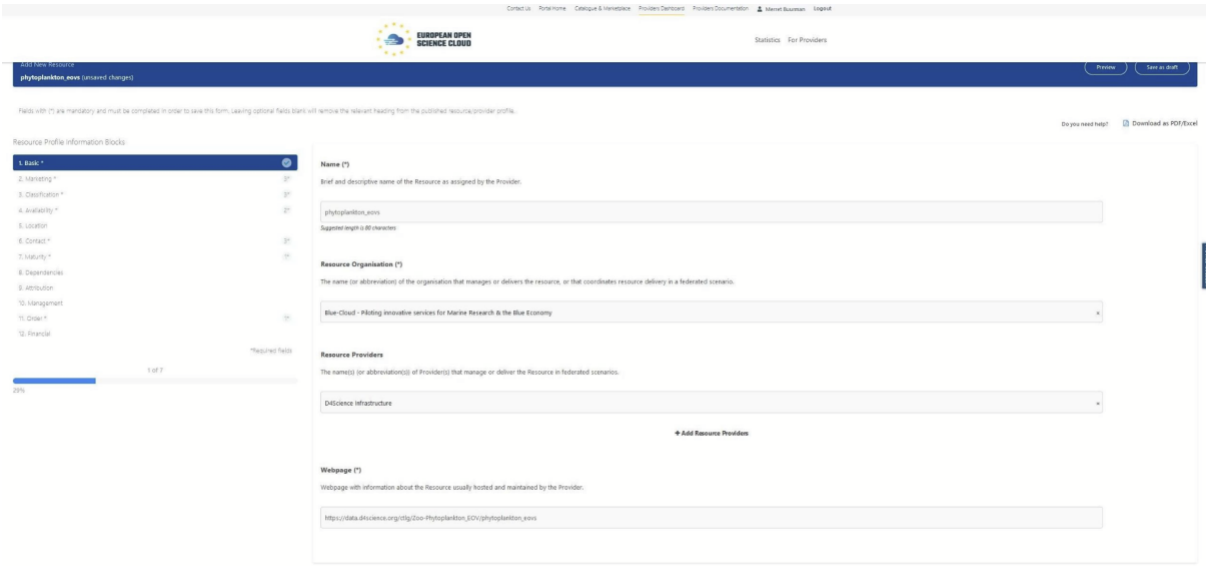


Figure 12 Screenshot of the first page of the onboarding GUI with the data of the first service filled in.

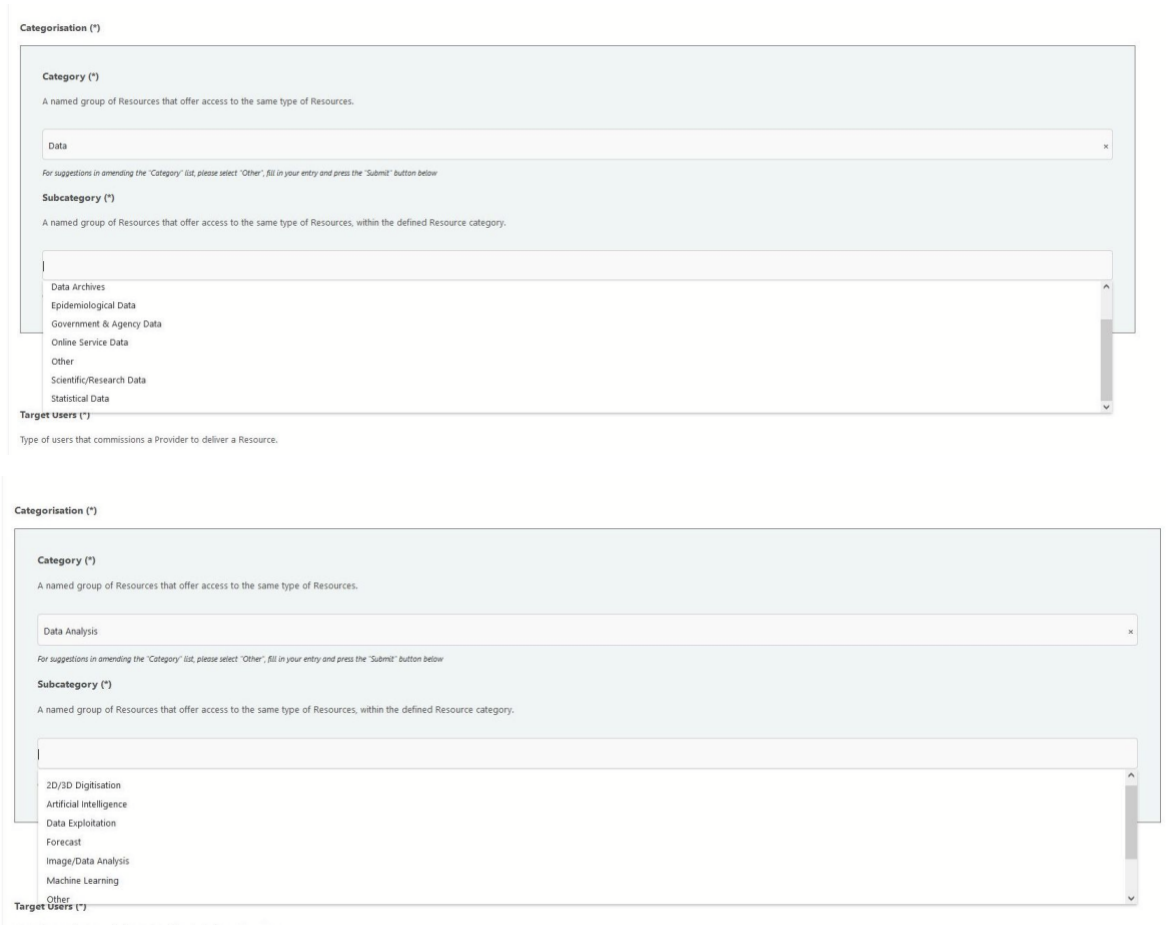


Figure 13 The possible subcategory options depend on the selected category. This is something we had to implement in the python tool as well.

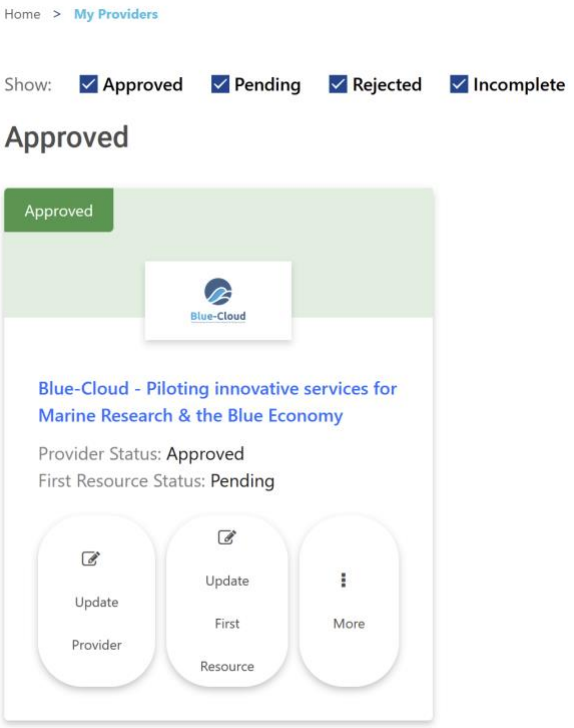


Figure 14 The First Resource Status was pending while some questions and clarifications were being solved by the Blue-Cloud and EOSC teams via email.

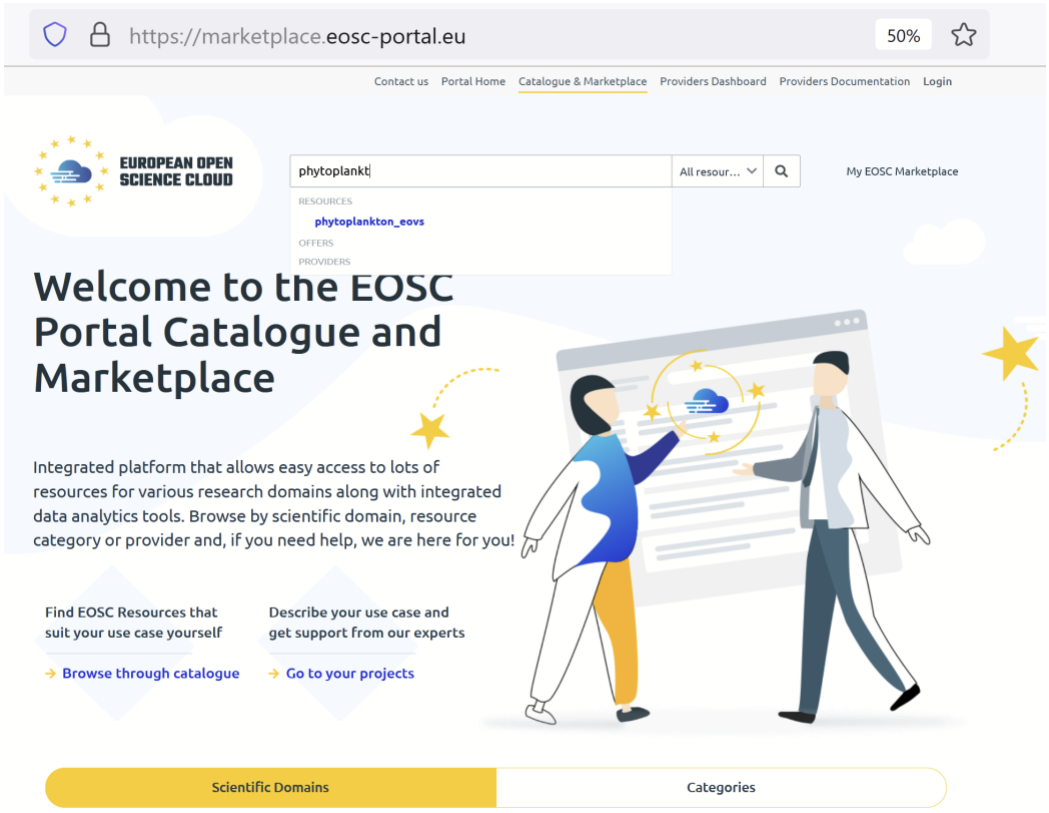


Figure 15 The first resource was successfully onboarded and can now be found in the marketplace.

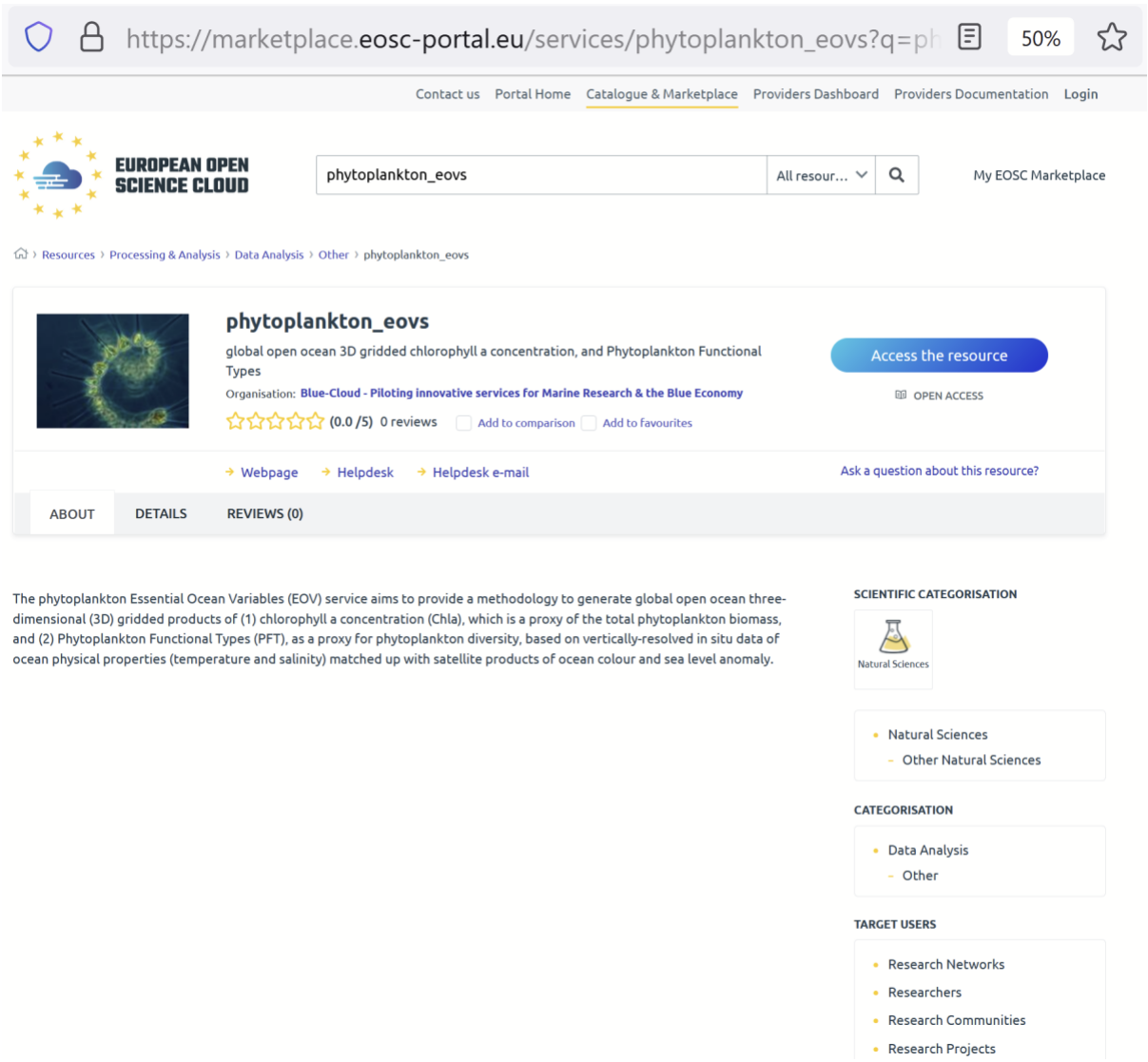


Figure 16 This is what users see in the marketplace.

4.3.6. Onboarding of the other services via HTTP API

As onboarding via GUI is a time-consuming process and prone to human errors (e.g. filling in wrong values into wrong fields, mixing up values of various services, overlooking a changed value that has to be updated etc.), it is not a solution for all services and even less suitable for regular metadata updates. Those have to be automated by the aforementioned program.

During this task, a program was developed in the programming language Python, which is able to retrieve the metadata from the Blue-Cloud Thematic Catalogue and translate it into the EOSC metadata standard. The compliance with the standard is ensured by validating via the EOSC validation API.

After mapping and validating, the program uploads the service metadata to the EOSC Catalogue. Different methods have to be used for a first-time creation and a metadata update. Thus, the

program must be able to check whether a service already exists in the catalogue or not. For this, a unique identifier (primary key) is required.

EOSC assigns a unique identifier to each resource (in this case, representing a service) upon creation. It is thus paramount to keep track of these identifiers on Blue-Cloud side and map them to the identifiers that Blue-Cloud assigns in their catalogue. To minimise installation and maintenance effort, the identifier mapping is written into a simple text file which obviously has to be kept safe.

A different option, which does not require keeping track of anything, is to search all existing services in the EOSC catalogue for the correct one, and retrieve its identifier. This is inefficient and it requires that the program is able to identify a service even in the case of modified metadata. For example, if services are matched using the service name (assigned by Blue-Cloud), it is not unthinkable that another service by another provider might have the same name. This method would also break if the service name was ever changed on Blue-Cloud side. This method could be an option if a metadata value (or combination thereof) can be guaranteed that is unchangeable and unique. For the time being, keeping track of the identifiers is the preferred way.

The program was developed for EOSC Metadata profile version 3. Since end of June 2022, profile version 4 is rolled out. The program thus has to undergo an upgrade before being able to be used further.

Once all services have been onboarded initially, it is planned to deploy it at CNR premises so that it will run automatically on a regular basis. Any updates on the Blue-Cloud side will be transferred to the EOSC catalogue, and new services will be onboarded once they reach TRL 7.

4.3.7. Authentication and Authorization

In order to interact with the Blue-Cloud HTTP API, one needs to be an authenticated and authorized user. In normal websites, this is realized through logging in with a username and password. For HTTP APIs, such a log-in is realized via a so-called token. For every VRE, a user gets a separate token, which can then be used to interact via the HTTP protocol.

As it is planned to let the program do the regular updates automatically and autonomously, a personal user token is not an ideal choice. Thus, CNR has provided a generic token to be used by the program, which has access to each VRE.

For EOSC, the read-only API endpoints (e.g. validation and vocabulary queries) do not require authentication/authorization. The creation and upload of service and provider description obviously do. As no service token was available, this is currently done using the personal user account of the Blue-Cloud task leader and co-author Merret Buurman (DKRZ), who was given administrator privileges for the Blue-Cloud provider.

4.3.8. Results

As described above, the service “Phytoplankton EOVs” of the VLab “Zoo-Phytoplankton EOV” was onboarded into EOSC catalogue using the Web GUI and is available on the EOSC marketplace (https://marketplace.eosc-portal.eu/services/phytoplankton_eovs).

At the time of writing, nine more services were listed as TRL 7 in the Blue-Cloud catalogue, and were thus ready to be onboarded:

- Zooplankton EOVs (VLab Zoo-Phytoplankton EOV)
- Modelling phyto & zoo –plankton interactions (VLab Zoo-Phytoplankton EOV)
- Zoo- and Phytoplankton Essential Ocean Variable products VLab (VLab Zoo-Phytoplankton EOV)
- Marine Environmental Indicators VLab (VLab Marine Environmental Indicators)
- OceanPatterns notebooks
- OceanRegimes notebooks
- Carbon Data notebooks
- Storm Severity Index (SSI) notebook
- MEI Generator

Personal communication with the service owners revealed that at least some of the services’ TRL had increased since filling in the metadata. This increase will soon be reflected in the Blue-Cloud catalogue too. Once they are listed as TRL 7, the Blue-Cloud program will pick them up for onboarding.

Two more services were in preliminary state at the time of writing, so no metadata could be harvested from the Blue-Cloud catalogue:

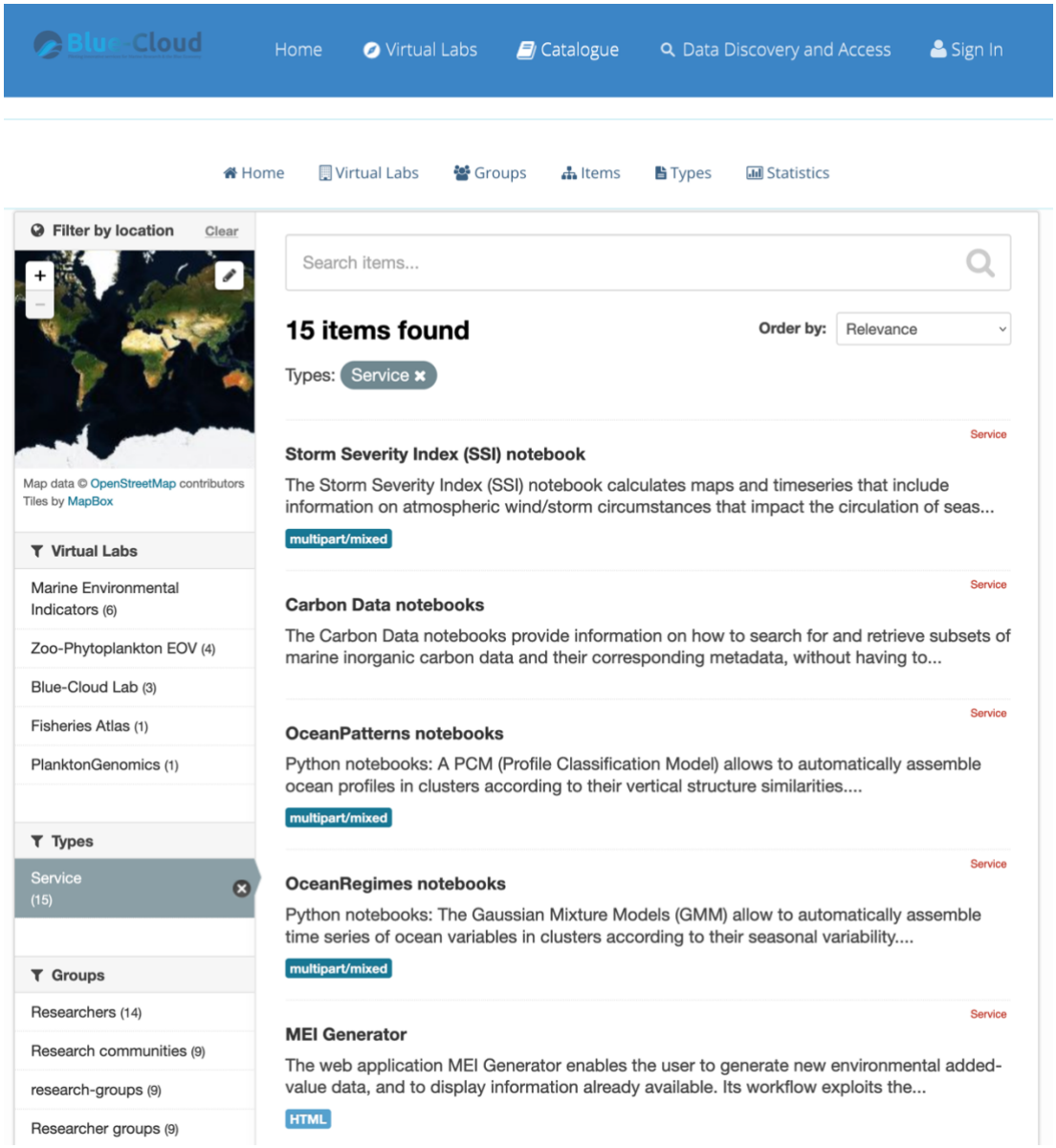
- Fisheries Atlas VLab
- Plankton Genomics VLab.

It is expected that these are ready for mapping soon.

Except for the one onboarded via Web GUI, none of the services listed above is available in the EOSC marketplace as of today. The reason why these services are not onboarded yet is that by the time most of the mapping problems described in sections 4.3.3 and 4.3.4 were solved, the EOSC catalogue went into a scheduled upgrade downtime. This lasted until the end of June 2022. During this time, technical support was reduced so remaining technical issues could not be solved.

When the downtime was over, the marketplace had switched their metadata profile from version 3 to version 4, increasing the number of mandatory metadata fields. So, the validation of the metadata that passed before now fails. The mapping process has to be adapted to match the new profile before we can onboard the services.

In order to not delay this deliverable further, it was decided to describe the results as they are. The upgrade to profile version 4 followed by the onboarding of the services, as well as the operationalisation of the onboarding procedure, will thus be done in the upcoming weeks.


Blue-Cloud
Home
Virtual Labs
Catalogue
Data Discovery and Access
Sign In
Home
Virtual Labs
Groups
Items
Types
Statistics
Filter by location
Clear
Map data © OpenStreetMap contributors
Tiles by MapBox
Virtual Labs
Marine Environmental Indicators (6)
Zoo-Phytoplankton EOV (4)
Blue-Cloud Lab (3)
Fisheries Atlas (1)
PlanktonGenomics (1)
Types
Service (15)
Groups
Researchers (14)
Research communities (9)
research-groups (9)
Researcher groups (9)
Search items...
15 items found
Order by: Relevance
Types: Service
Service
Storm Severity Index (SSI) notebook
The Storm Severity Index (SSI) notebook calculates maps and timeseries that include information on atmospheric wind/storm circumstances that impact the circulation of seas...
multipart/mixed
Service
Carbon Data notebooks
The Carbon Data notebooks provide information on how to search for and retrieve subsets of marine inorganic carbon data and their corresponding metadata, without having to...
Service
OceanPatterns notebooks
Python notebooks: A PCM (Profile Classification Model) allows to automatically assemble ocean profiles in clusters according to their vertical structure similarities....
multipart/mixed
Service
OceanRegimes notebooks
Python notebooks: The Gaussian Mixture Models (GMM) allow to automatically assemble time series of ocean variables in clusters according to their seasonal variability....
multipart/mixed
Service
MEI Generator
The web application MEI Generator enables the user to generate new environmental added-value data, and to display information already available. Its workflow exploits the...
HTML

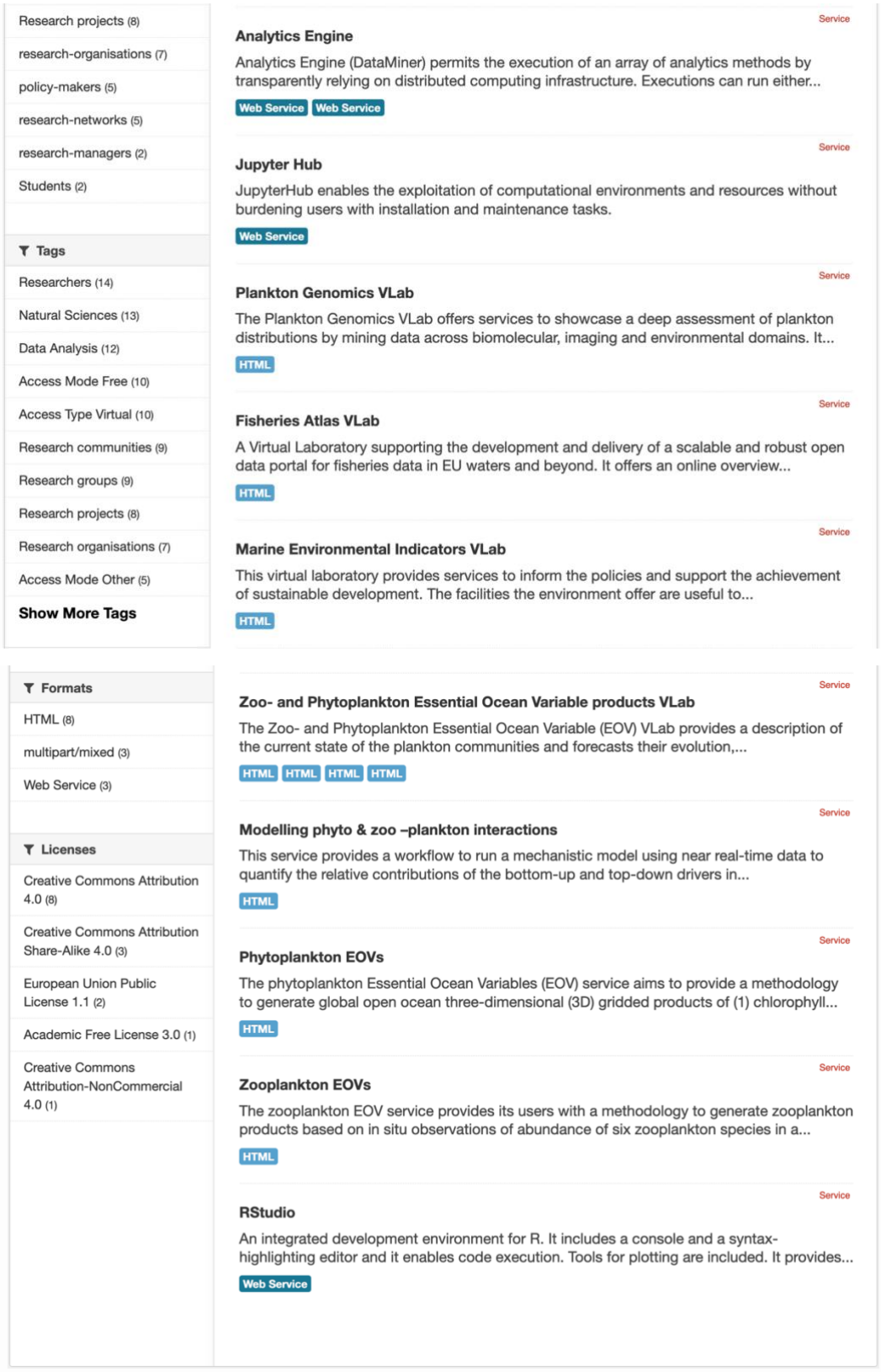


Figure 17 Services in the Blue-Cloud catalogue.

5. Summary, conclusions and next steps

In this deliverable, the Blue-Cloud Service Catalogue and the EOSC Portal Catalogue were described, and for both catalogues, the onboarding process for new providers and services.
Providers and services can be integrated into the catalogues in two ways:

- By manual integration using a GUI (via the EOSC Portal website). For providers, this is currently the only way to register initially.
- Via the HTTP REST API interface provided by the EOSC Portal.

For both methods, the services and providers must be thoroughly described by detailed metadata following the EOSC Data Model. Furthermore, to be featured in EOSC, services must have at least a Technology Readiness Level (TRL) of 7 ("System prototype demonstration in operational environment").

In this deliverable, it is described how the service provider Blue-Cloud was integrated into EOSC and how the services were described with metadata and included in the Blue-Cloud Thematic Catalogue. In detail the mapping process is described that converts the Blue-Cloud metadata into EOSC-compatible metadata. For this, a mapping module has been developed in the programming language Python. Then the metadata was validated using the EOSC validation API, also using a Python tool developed during this activity.

The next step is to upload the validated metadata into the EOSC Catalogue. This was carried out for one of the services that matched the TRL requirement at the time of writing, using the Web GUI. The remaining services will be onboarded using the EOSC HTTP API, which was undergoing a scheduled upgrade downtime in June, so it is expected to complete the onboarding process during July 2022. This is likely to be a continued work as services mature, reach higher Technology Readiness Level, slightly change scope or improve descriptions. It is also possible that more services are added as time passes.
The EOSC upgrade in June 2022 rolled out a new metadata profile (version 4), which introduces few new mandatory values, among other minor changes. Thus, the mapping tool has to be adapted to reflect the new metadata standard. Once this has been done, the services will be uploaded to the catalogue. Then, the developed tool will be deployed on CNR premises, so that it can pick up changes to the metadata in the Blue-Cloud catalogue and validate and transfer them to the EOSC catalogue automatically, on a regular basis. This will also ensure that those services that are now too premature to be onboarded will be integrated upon reaching TRL-7.

It may be a good idea to streamline the Blue-Cloud and EOSC catalogue contents slightly, so that less fields have to be adapted by the mapping tool. This would improve the stability and maintainability of the tooling. However, it is not strictly necessary for the functioning of the tool.

However, there is a non-technical dimension - as in every undertaking that involves metadata -, which is making sure that the metadata is well-described and that those descriptions are kept up-to-date in the first place. This challenge remains and may have to be addressed in the long run by the Blue-Cloud sustainability working group (WP6). However, this mainly concerns the Blue-Cloud catalogue, as metadata which is updated there will automatically be propagated to the EOSC Catalogue.

References

M. Assante, L. Candela, P. Pagano, A. Dell'Amico, L. Frosini, L. Lelii, M. Lettere, F. Mangiacrapa, G. Panichi, F. Sinibaldi (2020) ***Blue Cloud VREs Common Facilities (Release 1)***. D4.2 Blue-Cloud Deliverable, November 2020

M. Assante, L. Candela, P. Pagano, R. Cirillo, A. Dell'Amico, L. Frosini, L. Lelii, M. Lettere, F. Mangiacrapa, G. Panichi, F. Sinibaldi (2021) ***Blue Cloud VREs Operation Report (Release 1)***. D4.1 Blue-Cloud Deliverable, February 2021

D. M. A. Schaap, P. Thijsse, P. Pagano, M. Assante, E. Boldrini, M. Buurman, M. D'Antonio, C. Ariyo, G. Maudire, C. Nys (2020) ***Blue Cloud Architecture (Release 1)***. D2.6 Blue-Cloud Deliverable, July 2020

G. Sipos, T. Zamani, K. Koumantaros, A. Coveney, I. D. Álvarez, N. Tonello, G. Koloventzos, O. Appleton, N. Liampotis (2020) **EOSC-hub Integration Handbook for Service Providers.** 10.5281/zenodo.3826907