



## *D2.5: Interfacing Blue Cloud Data Discovery and Access with EOSC*

<b>Work Package</b>	WP2 - Developing the Blue Cloud discovery and access service and overall Blue Cloud architecture
<b>Lead Partner</b>	DKRZ (EUDAT)
<b>Lead Author (Org)</b>	Merret Buurman (DKRZ), Anna-Lena Flügel (DKRZ), Heinrich Widmann (DKRZ), Dick M. A. Schaap (MARIS)
<b>Contributing Author(s)</b>	Claudia Martens (DKRZ)
<b>Reviewers</b>	Pasquale Pagano (CNR) *
<b>Due Date</b>	31.05.2021, M20
<b>Submission Date</b>	31.05.2021
<b>Version</b>	1.0

### Dissemination Level

- 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)



## 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



This work by Parties of the Blue-Cloud Consortium is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). “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.

## VERSIONING AND CONTRIBUTION HISTORY

<b>Version</b>	<b>Date</b>	<b>Authors</b>	<b>Notes</b>
0.1	20.04.2021	Merretuurman (DKRZ), Heinrich Widmann (DKRZ)	Initial draft.
0.2	13.05.2021	Merretuurman (DKRZ), Anna-Lena Flügel (DKRZ), Heinrich Widmann	Finished first draft.
0.3	16.05.2021	Merretuurman (DKRZ), Claudia Martens (DKRZ), Anna-Lena Flügel (DKRZ)	Further contributions
0.4	25.05.2021	Heinrich Widmann (DKRZ)	Restructured and cleaned up
0.5	28.05.2021	Claudia Martens (DKRZ)	Example record description
0.6	30.05.2021	Dick Schaap (MARIS)	Edits and completed
1.0	31.05.2021	Heinrich Widmann (DKRZ)	Finalised draft, ready for internal review*
1.1	04.08.2021	Merretuurman (DKRZ)	Incorporated changes addressing the comments by the mid-term review report.

## Contents

1. Executive summary .....	4
1. Introduction and overarching goal .....	7
2. The Concept: Multi-tiered Interdisciplinary Data Discovery .....	8
3. The Implementation .....	11
3.1. The Blue-Cloud Data Discovery and Access Service	11
3.2. The Blue-Cloud Common Metadata Model	12
3.3. The EUDAT-B2FIND metadata model and mapping of BlueCloud onto B2FIND schema	15
3.4. Harvesting the Blue-Cloud metadata by B2FIND	17
6. 4. Outlook and Open Issues .....	20

# 1. Executive summary

The Blue-Cloud Service platform will enable, among other services, discovery and access of digital objects from the marine science domain via the European Open Science Cloud (EOSC) framework. For that purpose, the Discovery and Data Access Service which Blue-Cloud is developing will be made accessible through the EOSC Portal, by harvesting metadata from the domain-specific Blue-Cloud Data Discovery and Access Service and publishing in EUDAT-B2FIND, EOSC's generic and interdisciplinary discovery service. This allows end-users to search on a cross-domain level in B2FIND, easily accessible to all EOSC users via the EOSC Portal Catalogue and Marketplace<sup>1</sup>. It goes without saying that the more detailed and fine-grained search capabilities of the Blue-Cloud Data Discovery and Access Service – and of each Blue Data Infrastructures – are not replaced or affected by this undertaking, which is an enhancement of the visibility and findability of marine data also towards other research domains, such as climate sciences or food and agriculture sciences.

This deliverable illustrates conceptually the interfacing of data discovery and access of Blue-Cloud data resources with EOSC. The technical implementation comprises the exposure of metadata of the Blue-Cloud Data Discovery and Access service and the harvesting by B2FIND and will be carried out as soon as the Blue-Cloud Data Discovery and Access Service has been launched and is operational.

---

<sup>1</sup> <https://marketplace.eosc-portal.eu/>

## List of abbreviations and acronyms

Abbreviation	Signification
EOSC	European Open Science Cloud
B2FIND	EUDAT's interdisciplinary discovery portal for research data
EUDAT	A European collaborative data infrastructure, one of the largest infrastructures of integrated data services and resources supporting research in Europe.
EurOBIS	European Ocean Biodiversity Information System
Euro-Argo	ERIC coordinating the European contribution to the international Argo program, deploying and operating a global array of Argo floats
Argo GDAC	Argo Global Data Assembly Centre
EMODnet	European Marine Observation and Data Network
ELIXIR-ENA	European Nucleotide Archive
EcoTaxa	Web application dedicated to visual exploration and taxonomic annotation of planktonic biodiversity images
SOCAT	Surface Ocean CO <sub>2</sub> Atlas
ICOS-Marine	Integrated Carbon Observation System
SeaDataNet	Pan-European infrastructure for managing, indexing and providing access to marine data sets and data products, acquired by European organisations from research cruises and other observational activities in European coastal marine waters, regional seas and the global ocean
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.

MARIS	Mariene Informatie Service “MARIS” BV, Blue-Cloud partner leading the operation of the Blue-Cloud Data Discovery & Access service
CNR	National Research Council of Italy, Blue-Cloud partner co-leading this deliverable

# 1. Introduction and overarching goal

Blue-Cloud is a large pan-European project federating marine and ocean data infrastructures and services involving marine research institutions across Europe. This is achieved thanks to the Blue-Cloud platform, the architecture of which (Schaap et al. 2020) consists of two major families of components: (a) the *Blue-Cloud Data Discovery and Access* service component to serve federated discovery and access to ‘blue data’ infrastructures; and (b) the *Blue-Cloud Virtual Research Environment (VRE)* component to provide a web-based environment as a federation of computing platforms and analytical services. This deliverable focuses on the former component, addressing findability and accessibility of research data – two crucial pillars for FAIR data (Wilkinson et al., 2016).

An abundance of marine research data exists today: raw data from remote and in-situ sensors, quality-checked, processed and aggregated datasets, and derived data products. However, these are typically scattered across diverse research institutions around the world, making the discovery and access of the data a tedious endeavour for the user. While experienced marine scientists often know where and how to look for the data they need, scientists from other domains (e.g. climate sciences, food and agriculture) or those who are new to the marine data landscape are often discouraged and deterred by its complexities.

This is where generic, cross-disciplinary discovery tools come into play. Providing an aggregated, coarse-grained view over datasets from various domains of research with a common description format, they realise an easy entry point for users from a variety of disciplines who look for research data on a particular topic. Once a dataset of interest is found, further details, if required, can then be retrieved at their repository of origin. This simplifies interdisciplinary studies which have so far been thwarted by not knowing which data exists or where to find it.

The European Open Data Cloud (EOSC) provides such a generic discovery tool called B2FIND. The overall goal of the EOSC is to provide scientists of all disciplines with all kinds of data and services they may need to conduct their research. As such, EOSC invites – and depends on – the various scientific disciplines to provide their services to a wider audience, and describe them in a generic, understandable way. The marine sciences being central and crucial to a range of scientific questions, the integration of marine data and services into the EOSC is an important milestone towards interdisciplinary scientific research. This integration will be performed by Task 2.3 in Blue-Cloud and is described in this deliverable. By providing its data in an aggregated, harmonized way to the users of all disciplines via the EOSC, Blue-Cloud acknowledges its important role and responsibility of opening up its community’s data treasures to the broader scientific world.

This document is structured in four sections: Following this introduction (section 1), the concept of multi-tiered interdisciplinary data discovery is presented in section 2. Section 3 elaborates how this will be put into practice in Blue-Cloud. The proof of concept is described in section 4. Eventually, we will give a summary and outlook in the concluding section.

## 2. The Concept: Multi-tiered Interdisciplinary Data Discovery

The goal of the activities described in this deliverable is to allow users not only from the marine scientific domain – who may already be familiar with the marine data infrastructures making up Blue-Cloud – but all researchers of any discipline to find marine research data.

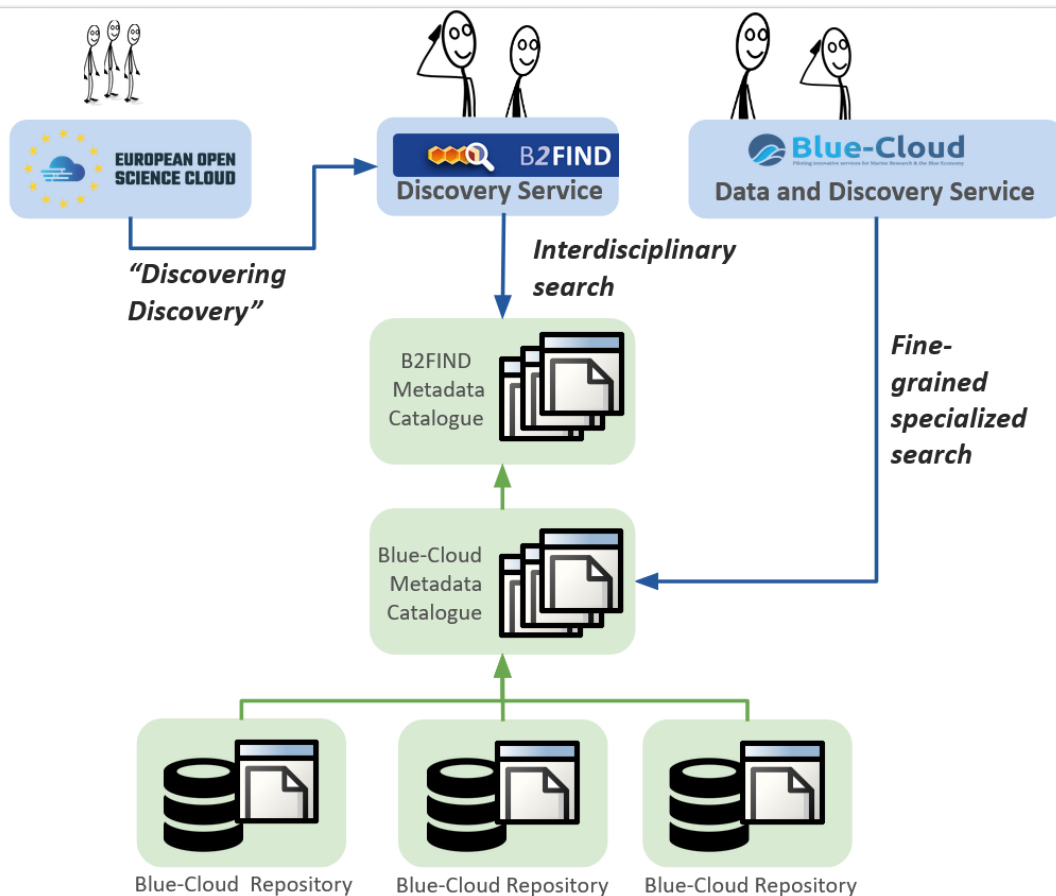
To achieve this goal, we will put in place four levels or *tiers*, which are:

2. The *Blue Data Infrastructures* and their repositories, where the Blue-Cloud partner infrastructures publish their research data.
3. The *Blue-Cloud Data Discovery and Access Service*, which collects and aggregates metadata from these individual repositories and makes it available – for users, but also for harvesting – in the Blue-Cloud metadata catalogue.
4. *EUDAT-B2FIND*, EOSC's generic Discovery and Access Service which harvests the metadata from the Blue-Cloud metadata catalogue, thereby making the referenced data discoverable and accessible on a cross-domain level.
5. The *EOSC-Portal*, which allows all end-users to discover marine data resources by referring users to the B2FIND portal.

The metadata gets more and more aggregated with each further level. The first three tiers each have a specific target group in mind: from the marine researcher with very specific data needs and detailed knowledge of the marine data landscape, passing by oceanographers familiar with Blue-Cloud but not with each individual data provider, up to a cross-disciplinary researcher, potentially from another area of science. The last tier allows users from each group to find the other tiers.

This bottom-up harvest workflow is schematically shown in figure 1:





**Figure 1:** Schema illustrating the various discovery options (blue) and the bottom-up metadata uptake along the tiers (green): The data and metadata resides in the Data Infrastructures (bottom), from where it is harvested by the Blue-Cloud Data and Access service, which in turn is harvested by B2FIND, which is featured in the EOSC-Portal.

This deliverable focuses on tier 3 of the above list: Harvesting the metadata from the Blue-Cloud Data and Access service into B2FIND and thus making it available via the EOSC Portal. The main challenge is the establishment of synchronous, incremental and comprehensive harvesting and an elaborate mapping on base of the data models and standards of Blue-Cloud and B2FIND. How this is managed and implemented is in detail described in section 3.

The final goal is to enable each type of end-user to find and access the data as directly and easily as possible to their specific search criteria and use case. To illustrate this with three types of users:

- Someone who is just starting their PhD and/or is not familiar with the infrastructures of marine research. This person would choose the EOSC portal as an entry point and be pointed to EUDAT-B2FIND as a cross-domain search portal for research data, from where the person will be directed towards the Blue-Cloud Data Discovery and Access service for deeper search and data retrieval.

- Someone who has been working in marine research for some time and is familiar with the relevant data providers and with Blue-Cloud. This person will directly use the Blue-Cloud Data Discovery and Access service as an entry point to access the desired data.
- A user from one of the Blue-Cloud data providers or a researcher with a very narrow-scoped or fine-grained research question may directly use the search-and-access facilities at the individual data provider without any aggregated search facility.

## 3. The Implementation

After having described the concept of the various tiers and resolutions of data discovery, it is time to describe how this relates to Blue-Cloud on a technical level. The main ingredients to realize the EOSC integration are the Blue-Cloud Data Discovery and Access Service, the B2FIND Discovery Service, their respective metadata schemas and the mapping between them, and finally the harvesting of the metadata.

### 3.1. The Blue-Cloud Data Discovery and Access Service

The Blue-Cloud Data Discovery and Access Service (described in more detail in deliverable D2.4) is a service that aggregates and integrates metadata harvested from the various Blue Data Infrastructures comprising the Blue-Cloud project:

- SeaDataNet,
- EuroBIS,
- Euro-Argo - Argo GDAC,
- EMODnet,
- ELIXIR-ENA,
- EuroTaxa,
- SOCAT,
- ICOS-Marine.

As the central service where all the metadata from very diverse sources comes together, it is here that the crucial harmonization of metadata takes place: All the different metadata schemas from the above infrastructures are mapped to a single central Blue-Cloud Metadata Model, described in the next section. Furthermore, very fine-grained data sets are being aggregated to coarser collections in order not to overwhelm the users searching for data at this more generic level.

The Blue-Cloud Data Discovery and Access service has a 2-level approach for searching. At the first level, data sets are aggregated at collection level and described with a common metadata model for all blue data infrastructures. These metadata are then also exchanged with B2FIND for publication in EOSC. There is also a second level, the granule level, which facilitates users to drill down their search, find individual data sets, and retrieve these data sets, using a shopping basket mechanism.

To exchange the Blue-Cloud collection metadata, an endpoint is set-up at:

<https://data.blue-cloud.org/api/collections>

This endpoint provides a summary and harvesting of all Blue-Cloud collection metadata records as JSON files. Currently, there are circa 25.000 Blue-Cloud collection metadata records. The JSON summary and individual record files are generated on-the-fly from the Blue-Cloud Data Discovery and Access service, each time a harvest is initiated by B2FIND.

The Blue-Cloud JSON metadata summary is formatted like:

```
{
  "name": "Blue-Cloud-discovery-records",
  "blue-cloud-urls": [
    "https://data.blue-cloud.org/report/collections/103b40c0-c07b-46ba-b0e3-f4d5293edd13",
    "https://data.blue-cloud.org/report/collections/f080166b-0632-4de2-85df-97829d56eabf",
    "https://data.blue-cloud.org/report/collections/5a9df55d-0cb7-4354-acd0-221b3d5f02fe"
  ]
}
```

## 3.2. The Blue-Cloud Common Metadata Model

A Blue-Cloud Common Metadata Model has been defined as part of the Blue-Cloud Data Discovery and Access service for describing aggregated data sets from each of the federated blue data infrastructures. As indicated earlier, from these an exchange metadata model in JSON has been derived which is presented below by means of an example:

<https://data.blue-cloud.org/api/collections/b90d7650-a0d7-409c-ba31-122a1f4f9c3f>

gives:

```
{
  "Identifier": "b90d7650-a0d7-409c-ba31-122a1f4f9c3f",
  "last_update": "2021-05-21T15:08:00",
  "Source": "SeaDataNet Products",
  "Title": "SeaDataCloud Temperature and Salinity Historical Data Collection for the Black Sea V2",
  "Abstract": "The SeaDataCloud Temperature and Salinity Historical Data Collection for the Black Sea (Version 2) includes open access in situ data on temperature and salinity of water column in the Black Sea (and a little in the Sea of Azov) for period 1868 – 2019. The data were retrieved from the SeaDataNet infrastructure at the end of 2019. The dataset format is Ocean Data View (ODV) binary collection. The quality control of the data has been performed with the help of ODV software. Data Quality Flags have been revised and set up using the elaborated by SeaDataNet2 and SeaDataCloud project QC procedures in conjunction with the visual expert check. Data duplicates have been identified and excluded from the dataset. The final number of the Temperature and Salinity profiles (stations) in the collection is 162626.",
  "Organisation": [
    "Organisation 1",
    "Organisation 2",
    "Organisation 3"
  ],
  "Keywords": [
    "Oceanographic geographical features",
    "Salinity of the water column",
    "Sea of Azov"
  ],
  "Parameters": [
    "Temperature of the water column",
    "Salinity of the water column"
  ],
  "Instruments": [
    "Salinometer",
    "CTD"
  ],
  "Platforms": [
    "Research Vessel",

```

```

    "Glider"
  ],
  "Bounding_Box_WestLongitude": "27.50",
  "Bounding_Box_EastLongitude": "42.00",
  "Bounding_Box_SouthLatitude": "40.50",
  "Bounding_Box_NorthLatitude": "47.50",
  "Temporal_Extent_Begin": "1868-06-17T08:45:00",
  "Temporal_Extent_End": "2019-06-02T12:25:15",
  "OnlineResourceURL": "https://data.blue-cloud.org/search/seadata-products/step=b90d7650-a0d7-409c-ba31-122a1f4f9c3f"
}

```

Currently, there are circa 25.000 Blue-Cloud metadata records at collection level and for each such a JSON detail file is provided through the Blue-Cloud endpoint.

This way, each JSON file gives the following content:

<b>Identifier</b>	The Identifier is unique and persistent in time. Example: b90d7650-a0d7-409c-ba31-122a1f4f9c3f
<b>Last_update</b>	2021-05-21T15:08:00" (Convention: YYYY-MM-DDThh:mm:ss)
<b>Source</b>	SeaDataNet Products
<b>Title</b>	SeaDataCloud Temperature and Salinity Historical Data Collection for the Black Sea V2
<b>Abstract</b>	Short abstract describing the data collection or data product
<b>Organisation</b>	Organisation 1 (note: can be multiple)
<b>Organisation</b>	Organisation 2
<b>Organisation</b>	Organisation 3
<b>Keywords</b>	Oceanographic geographical features (Note: can be multiple)

<b>Keywords</b>	Salinity of the water column
<b>Keywords</b>	Sea of Azov
<b>Parameters</b>	Temperature of the water column (Note: Can be NULL to multiple)
<b>Parameters</b>	Salinity of the water column
<b>Instruments</b>	Salinometer (Note: can be NULL to multiple)
<b>Instruments</b>	CTD
<b>Platform</b>	Research Vessel (Note: Can be NULL to multiple)
<b>Platform</b>	Glider
<b>Bounding_Box_WestLongitude</b>	27.50
<b>Bounding_Box_EastLongitude</b>	42.00
<b>Bounding_Box_SouthLatitude</b>	40.50
<b>Bounding_Box_NorthLatitude</b>	47.50
<b>Temporal_Extent_Begin</b>	1868-06-17T08:45:00 (Convention: YYYY-MM-DDThh:mm:ss)
<b>Temporal_Extent_End</b>	2019-06-02T12:25:15 (Convention: YYYY-MM-DDThh:mm:ss)

<b>OnlineResourceURL</b>	<p><a href="https://data.blue-cloud.org/search/seadata-products/step= b90d7650-a0d7-409c-ba31-122a1f4f9c3f">https://data.blue-cloud.org/search/seadata-products/step= b90d7650-a0d7-409c-ba31-122a1f4f9c3f</a></p> <p>With this deep link EOSC users can land in the Blue-Cloud service at level 1 for the selected record, and then can decide to go wider at level 1 or deeper at level 2.</p>
--------------------------	--

### 3.3. The EUDAT-B2FIND metadata model and mapping of BlueCloud onto B2FIND schema

To be searchable and displayable in the B2FIND service, the metadata has to be mapped to the B2FIND metadata schema. Thus, a second and final schema mapping is required, this time from the Blue-Cloud metadata schema to the B2FIND metadata schema.

B2FIND's goal is, firstly, to make research data of various disciplines with very different metadata schemas and levels of data granularity discoverable in one portal: from huge netCDF files produced by climate modeling to small audio records of Swahili syllables and phonemes; from immigrant panel data in the Netherlands to a paleoenvironment reconstruction of the Mozambique Channel and from an image of the "Maison du Chirurgien" in ancient Roman Pompeii to an Excel file giving concentrations of calcium, magnesium, potassium and sodium in throughfall, litterflow and soil in an Oriental beech forest. Secondly, B2FIND wants to support smaller research communities to open their (meta)data silos and keep a low threshold for them to publish their metadata in B2FIND by having relatively few mandatory metadata elements and flexible mapping abilities.

The B2FIND metadata schema<sup>2</sup> was developed accordingly as a very generic schema that is yet interoperable with other widely used metadata standards. It is based on the DataCite schema<sup>3</sup> top-level elements and extended with the additional elements *<Temporal Coverage>*, *<Discipline>* and *<Instrument>*, thus allowing users to search and find generic research data elements across scientific disciplines as well as searching for certain measurement tools, e.g. data produced by specific beamlines or measurement stations.

To map from the Blue-Cloud schema to the B2FIND schema, the values from the former correspond to the latter as follows:

<sup>2</sup> <http://b2find.eudat.eu/guidelines/mapping.html#b2fmdschema>

<sup>3</sup> <https://schema.datacite.org/>

**Table 1:** Mapping from Blue-Cloud to B2FIND metadata schema.

Blue-Cloud metadata item	B2FIND metadata item	Comment
IDENTIFIER	<Identifier> [may have different values: 'DOI', 'PID' or 'Source']	Link to Blue-Cloud data record
TITLE	<Title>	
KEYWORD	<Keyword>	
BOUNDING_BOX	<SpatialCoverage>	
TEMPORAL_EXTENT	<TemporalCoverage>	
PARAMETER	-	default: map it to <Description> to make the values searchable via the full- text search
INSTRUMENT	<Instrument>	
PLATFORM	<Instrument>	second value in <Instrument> field, comma separated or in parentheses; alternatively: map to <Description>
ORGANIZATION	<Contributor>	
DATESTAMP	<PublicationYear>	
DESCRIPTION	<Description>	

This crosswalk shows that a lot of the core elements of the Blue-Cloud Discovery and Access metadata schema can be semantically mapped to the B2FIND schema quite adequately. Only the PARAMETER element has no equivalent in the target schema. Here, information loss will occur, as the B2FIND CKAN portal does not allow searching for parameters. One possible - albeit inadequate - solution in this case might be to map the values of this element to the B2FIND <Description> field. However, values without their keys (“45 mph” without its key “wind speed” for example) might lose their significance in the process. This is the price of a generic and thus coarser metadata schema: it cannot carry all information of the discipline-specific source schema, but in turn enables interdisciplinary metadata exchange.

According to the B2FIND metadata schema, additional metadata elements are mandatory which can be completed as default for the Blue-Cloud schema. These are given in Table 2.



**Table 2:** Mapping B2FIND mandatory metadata elements for Blue-Cloud.

B2FIND metadata element	Default mapping for Blue-Cloud
<Community>	"Blue-Cloud"
<Publisher>	"Blue-Cloud"
<Discipline>	"Marine Science"
<OpenAccess>	"true"

### 3.4. Harvesting the Blue-Cloud metadata by B2FIND

Once the Blue-Cloud Data Discovery and Access service has been launched and is fully operational, then the harvesting of the Blue-Cloud JSON metadata by B2FIND will be done on a daily basis. This will ensure that the Blue-Cloud and B2FIND catalogues will be in sync.

B2FIND will harvest every night to retrieve any metadata records that have been modified and/or added since the last harvest. At this point, the Blue-Cloud Discovery and Access metadata will be findable and accessible in B2FIND<sup>4</sup>, which in turn is also accessible via the EOSC Portal<sup>5</sup>. As B2FIND is harvested by OpenAire Explore<sup>6</sup>, it acts as a metadata aggregator, thus further disseminating Blue-Cloud research products.

---

#### Demonstrator

For illustrating the integration of Blue-Cloud records in B2FIND, we created an example record based on given conventions and ingested this file within a test environment that offers the same GUI as the productive search portal.<sup>7</sup> This integration is twofold and consists of

- a) Community integration = representation of Blue-Cloud in B2FIND CKAN GUI

---

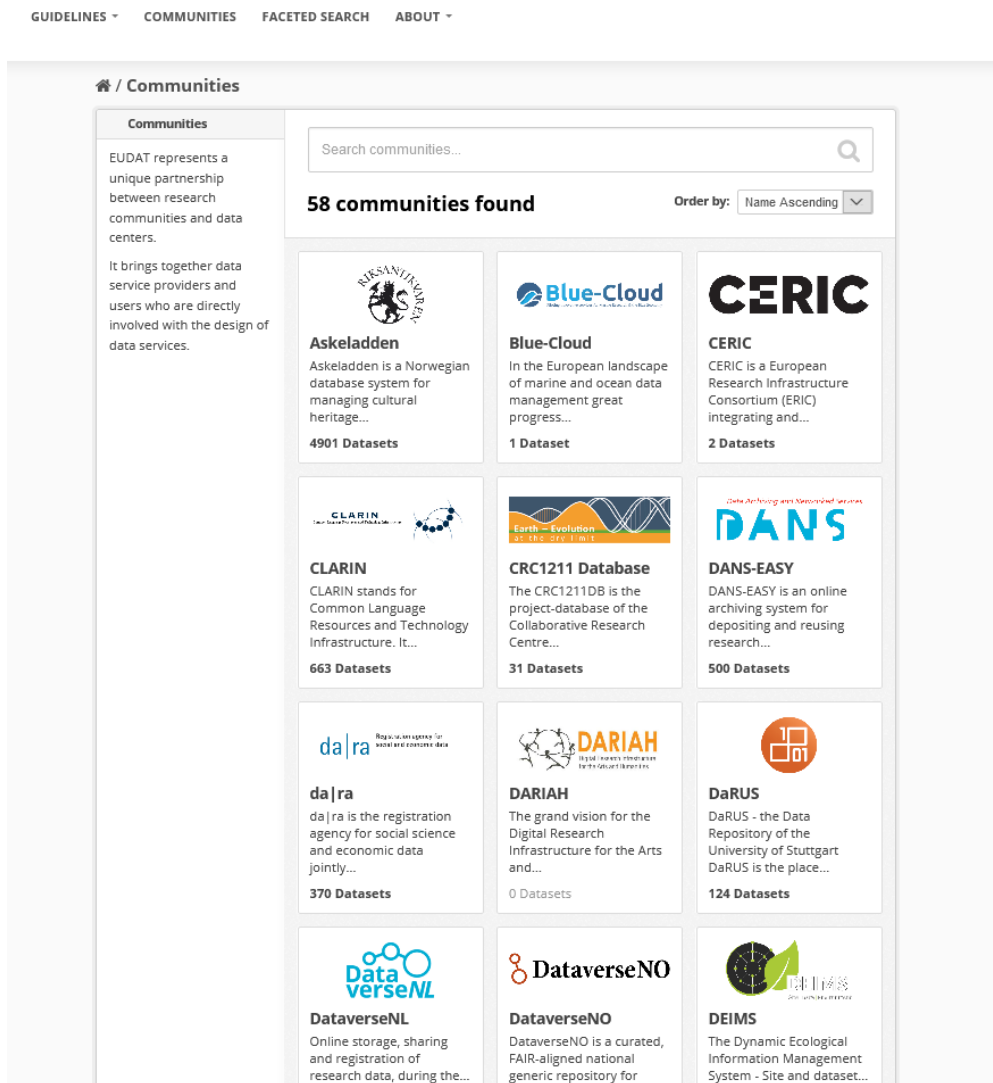
<sup>4</sup> <http://b2find.eudat.eu/>

<sup>5</sup> <https://marketplace.eosc-portal.eu/services/b2find>

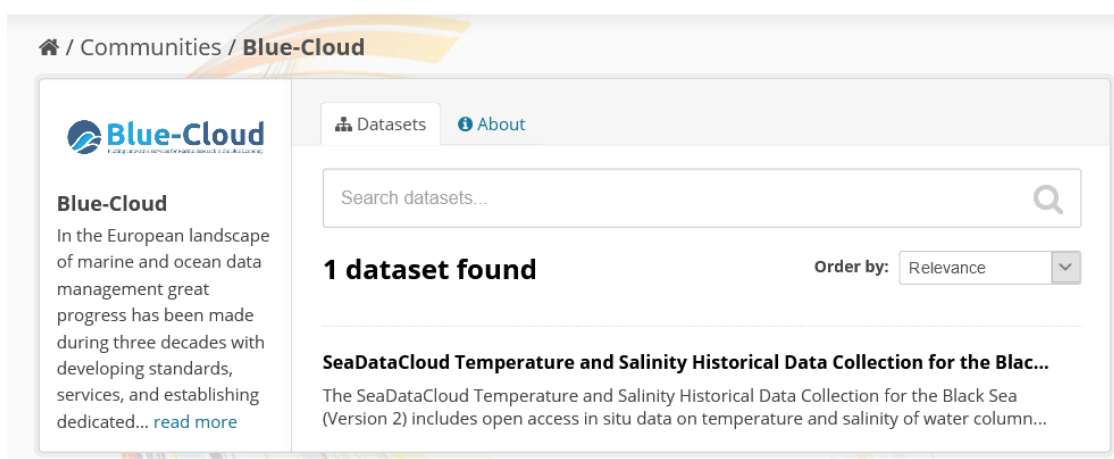
<sup>6</sup> <https://explore.openaire.eu/>

<sup>7</sup> The demonstrator for community integration is available here: <http://eudat7-devel.dkrz.de/group/bluecloud>, the single record view of Blue-Cloud example record may be seen here: <http://eudat7-devel.dkrz.de/dataset/b90d7650-a0d7-409c-ba31-122a1f4f9c3f>. Please note that this demonstrator will be outdated as soon as Blue-Cloud Data Discovery and Access service is productive and Blue-Cloud records are integrated in productive B2FIND.

For B2FIND, a ‘Community’ is defined as “The scientific community, research infrastructure, project or data provider from which B2FIND harvests the metadata”. All Communities are displayed both on a sampled view (Figure 2) and on a single view (Figure 3), displaying *Name*, *Description*, *Logo* and *Weblink* for each Community.



**Figure 2:** Sample view of all Communities on B2FIND CKAN GUI



**Figure 3:** Single Community view of Blue-Cloud on B2FIND CKAN GUI

b) Metadata Record ingestion = integration of Blue-Cloud resources within the search portal.

Records are searchable on the 'Faceted Search' website, using the facets to narrow down results (currently 13 facets are offered, including spatial/temporal search options).<sup>8</sup> From this search result page a detailed view of each record can be seen by clicking on it; the single record view for Blue-Cloud demonstrator is shown in Figure 4.

<sup>8</sup> B2FIND faceted search is available here: <http://b2find.eudat.eu/dataset>

Home / Datasets / SeaDataCloud Temperature ...

**Social**

Google+

Twitter

Facebook

**License**

Other (Open) [OPEN DATA](#)

Dataset Communities

### SeaDataCloud Temperature and Salinity Historical Data Collection for the Black Sea V2

The SeaDataCloud Temperature and Salinity Historical Data Collection for the Black Sea (Version 2) includes open access in situ data on temperature and salinity of water column in the Black Sea (and a little in the Sea of Azov) for period 1868 - 2019. The data were retrieved from the SeaDataNet infrastructure at the end of 2019. The dataset format is Ocean Data View (ODV) binary collection. The quality control of the data has been performed with the help of ODV software. Data Quality Flags have been revised and set up using the elaborated by SeaDataNet2 and SeaDataCloud project QC procedures in conjunction with the visual expert check. Data duplicates have been identified and excluded from the dataset. The final Data Number of the Temperature and Salinity profiles (stations) in the collection is 162626. Parameters: Salinity of the water, Temperature of the water column column. Platforms: Research Vessel, Glider.

Oceanographic geogr... Salinity of the wat... Sea of Azov

Identifier	
Source	<a href="https://data.blue-cloud.org/search/seadata-products/step=b90d7650-a0d7-409c-ba31-122a1f4f9c3f">https://data.blue-cloud.org/search/seadata-products/step=b90d7650-a0d7-409c-ba31-122a1f4f9c3f</a>
Metadata Access	<a href="https://data.blue-cloud.org/api/seadata-products/step=b90d7650-a0d7-409c-ba31-122a1f4f9c3f">https://data.blue-cloud.org/api/seadata-products/step=b90d7650-a0d7-409c-ba31-122a1f4f9c3f</a>

Provenance	
Instrument	Salinometer; CTD
Publisher	Blue-Cloud
Contributor	SeaDataNet
Publication Year	2021
OpenAccess	true

Representation	
Discipline	Marine Science
Spatial Coverage	27.50W, 40.50S, 42.00E, 47.50N
Temporal Coverage Begin	1868-06-17T08:45:00
Temporal Coverage End	2019-06-02T12:25:15

Figure 4: Single record view of Blue-Cloud example record.

## 6. 4. Outlook and Open Issues

As the Blue-Cloud Data Discovery and Access service is not finished at this time, the harvesting and integration into B2FIND could not take place yet. It will be carried out as soon as the Blue Cloud Data Access & Discovery service is ready to be harvested.

As shown in this deliverable, the concept of making the Blue-Cloud metadata findable in B2FIND and thus integrating it into the EOSC is finished and well established. Once the bulk of the Blue-Cloud

metadata will be ready for harvesting, which is expected to happen in the course of June 2021, the integration into B2FIND will follow shortly after, unless any of the technical preconditions described and tested in this deliverable are changed by any party involved.

The number of metadata records that are expected to be provided in B2FIND is approximately 25,000 and gradually increasing. This is of course significantly more than what we could test with the beta version, but as B2FIND has been handling providers with amounts ranging from just a few to 400,000 metadata records for years (for example Pangaea, Herbadrop and SLKS), we do not expect any problems arising from this.

## 5. References

Mark D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, et al. (2016). **The FAIR Guiding Principles for scientific data management and stewardship**. *Scientific Data*. **3**: 160018. [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

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