LifeWatch ERIC Application Profiles

The metadata schemas for the LifeWatch ERIC resources

	Name:	Title:
Authors:	Lucia Vaira Nicola Fiore Ilaria Rosati	Web Portal Officer Service Centre ICT Coordinator Research Technologist at the Italian National Research Council - Research Institute on Terrestrial Ecosystems
Reviewer:	-	-
Approver:	Christos Arvanitidis	Chief Executive Officer

With the contribution of:

Daphnis De Pooter (LifeWatch Belgium); Joana Beja (LifeWatch Belgium, VLIZ); Nikos Minadakis (LifeWatch Greece); Alessandro Oggioni (LifeWatch Italy, CNR-IREA); Caterina Bergami (LifeWatch Italy, CNR-ISMAR); Cristiano Fugazza (LifeWatch Italy, CNR- IREA); Paolo Tagliolato (LifeWatch Italy, CNR- IREA); Paul Martin (LifeWatch Netherlands); Liao Xiaofeng (LifeWatch Netherlands, University of Amsterdam); Gregor Aljancic (LifeWatch Slovenia, Karst Research Institute); Magdalena Aljancic (LifeWatch Slovenia, Karst Research Institute); José María García (LifeWatch Spain, University of Seville); Antonio José Sáenz-Albanés (LifeWatch ERIC).

How to cite: *Vaira, L., Fiore, N., & Rosati, I. (2022). LifeWatch ERIC Application Profiles (Version 1). LifeWatch ERIC. https://doi.org/10.48372/8528-9Z45*

Table of Contents

Executive Summary

The purpose of this report is to present the metadata profiles adopted by LifeWatch ERIC for managing different resources: datasets, services, Virtual Research Environments (VREs), workflows and research sites.

The work started in November 2018 on the occasion of the first "Metadata, Controlled Vocabularies & Ontologies" meeting held in Lecce, Italy, with the purpose of setting a roadmap for a common strategy to be adopted on metadata, controlled vocabularies and ontologies within the LifeWatch ERIC community and in accordance with the FAIR principles (https://www.lifewatch.eu/download/metadata-controlled-vocabularies-ontologies-1st-working-meeting-report/). A Metadata Working Group, consisting of several experts from the different LifeWatch ERIC national nodes, was established in order to define the metadata schema and profile for each resource and the requirements for the metadata management tool to be used in LifeWatch ERIC.

1. Definitions

Metadata are "machine-readable information about electronic resources or other things" (Berners-Lee, 1997) and are used to describe the features of a resource, thus making easier its management and retrieval.

A set of metadata elements combined to serve a specific purpose, constitute a **metadata schema**. The metadata elements are sometimes grouped into sections useful for different purposes (e.g., basic information, contact information, license information, etc.). For each element, the schema defines the semantics, that is the meaning intended to be conveyed, and the rules for the content, that are the data type, the domain range from which values can be used, whether the element is mandatory, optional or conditional. The schema also defines how the elements are related to each other in terms of hierarchy and cardinality.

Although the adoption of a single metadata standard would assure the findability and the reusability of resources and interoperability among applications, there exists no unique metadata schema that is appropriate to fulfil the requirements and needs of every application. Moreover, already existing schemas are not adequate enough to cover the particular requirements. That's why the use of **application profiles**, which is an

aggregation of metadata elements selected from one or more different schemas and combined into a new schema (Heery & Patel, 2000), is often suggested.

2. Introduction

To date, a huge number of resources is currently available through various catalogues, even if such resources are not always easy to find and to integrate and this can be due to the unavailability or inaccuracy of their metadata and/or to the lack of interoperability among the standards. Therefore, in order to improve the findability, accessibility, interoperability, and reuse of resources, the adoption of FAIR best practices is highly recommended.

The FAIR principles are a set of guiding principles to make data and other research outputs Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016; European Commission Expert Group Report, Turning FAIR into Reality; 2018). They emphasise machine-actionability, that is the capacity of computational systems to find, access, interoperate, and reuse data and other resources without or with a minimal human intervention.

In order to respond to this challenge by making available resources that employ FAIR best practices, a number of actions has been undertaken from LifeWatch ERIC starting from the design of its own application profiles.

The criteria followed to implement FAIR principles in the LifeWatch ERIC schema are:

- **Findability**: metadata are registered and indexed in a searchable resource in order to make resources easily and univocally discoverable for both humans and machines.
- **Accessibility**: metadata are retrievable by their identifier using a standardised communications protocol, which is open, free, and universally implementable. Moreover, when necessary, the protocol allows for authentication and authorisation procedures.
- **Interoperability**: metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- **Reusability**: metadata are richly described with a plurality of accurate and relevant attributes and are released with a clear and accessible usage license.

The metadata of the LifeWatch ERIC resources developed by following these principles and criteria, are hosted in the LifeWatch ERIC Metadata Catalogue

(https://metadatacatalogue.lifewatch.eu), in order to be searched, discovered and accessed.

The LifeWatch ERIC Metadata Catalogue is a standard-based information management system based on GeoNetwork 3.10, designed and implemented to enable access to several resources from a variety of providers through descriptive metadata, enhancing and promoting the information exchange and sharing among organisations and research infrastructures.

The main goal of the Catalogue is to increase collaboration within and among organisations for reducing duplication and enhancing information consistency and quality and to improve the accessibility of a wide variety of resources along with the associated information, organised and documented in a standard and consistent way. Moreover, the LifeWatch ERIC Metadata Catalogue allows (upon validation and verification) the creation of Digital Object Identifiers (DOIs) for resources that do not have it, by exploiting the GeoNetwork – DataCite connection.

3. The LifeWatch ERIC application profiles

The usage of the same terminology concepts is highly relevant in the LifeWatch ERIC context. Indeed, the resources hosted in the Metadata Catalogue exploit the work already done within the Working Group D of the Tesseract project where LUPO, the LifeWatch ERIC Upper Ontology (here the details: http://ecoportal.lifewatch.eu/ontologies/LUPO), has been defined so that the key concepts (dataset, service, Virtual Research Environment, workflow, etc.) are well defined and agreed within the scientific community.

The LifeWatch ERIC Metadata Catalogue allows to manage descriptive metadata related to these kinds of resources that are based on two main standards:

- ISO 19139[1] (VREs, services, workflows and research sites);
- EML 2.2.0[2] (datasets).

Each application profile is organized in sections, which reflect the main information related to each specific resource. Each section contains optional and mandatory metadata.

[1] ISO 19139 standard - https://www.iso.org/standard/32557.html

[2] EML: Ecological Metadata Language - https://eml.ecoinformatics.org

3.1 Metadata description for Datasets

According to the LUPO definitions, a dataset represents *a set or collection of data, records or information that constitutes distinct units of information in the knowledge generation process* (ID: https://kos.lifewatch.eu/ontologies/LUPO#Dataset).

The dataset metadata schema is based on a subset of the EML2.2.0 standard and includes 77 metadata elements. Figure 1 shows the different sections of the profile that are described in detail below.

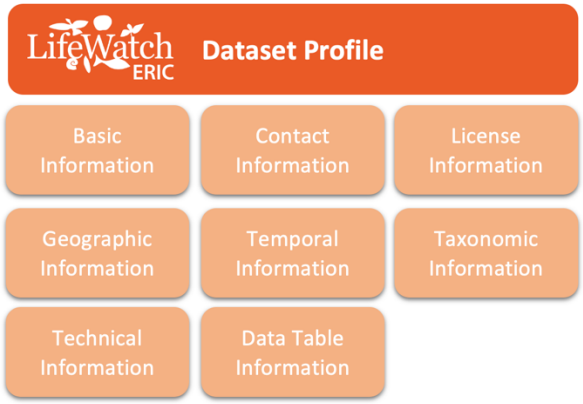

Figure 1: Dataset application profile

Basic information

This section includes the basic information related to the dataset:

- Identifier (*mandatory*), the globally unique identifier within a particular data management system. Typically, a DOI or another identifier that is both citable and resolvable is used.
- Title (*mandatory*), the title of the dataset.
- Short name (*optional*), the short name of the dataset.
- Publication date (*optional*), the date when the dataset was or will be made publicly available in the format dd/mm/YYYY.
- Dataset language (*mandatory*), the language in which the dataset is written.
- Abstract (*optional*), the brief description of the dataset, providing enough detail to differentiate it from other similar resources.
- Keyword (*mandatory and multi-value*), keywords or key phrases that concisely describe the dataset or are related to the dataset. Each keyword field should contain one and only one keyword and the use of singular form is

recommended.

- Keyword Thesaurus (*optional*), the name of the thesaurus from which the set of keywords is derived.
- Distribution information (*optional and multi-value*), metadata elements for accessing useful online resources available through the web. It is a multi-value section, hence it is possible to add one or more values. The included metadata are:
  - Online Description (*optional*), the description of the content of the URL.
  - URL (*mandatory*), the URL associated to the dataset. The DOI can be also specified here.

Contact information

This section contains information about the person(s) responsible for the dataset. It is a multi-value section, hence it is possible to add one or more persons. The relevant contacts are classified into: creator (*mandatory*), metadata provider (*optional*), associated party (*optional*), contact (*mandatory*). The included metadata for each type are:

- Organisation Name (*mandatory*), the affiliation of the person who created the dataset.
- First Name (*mandatory*).
- Surname (*mandatory*).
- Position Name (*mandatory*).
- Email (*mandatory*).
- User ID (*optional*), an identifier that links the person to a directory of individuals (e.g., ORCID).

License information

This section includes the license details related to the dataset:

- Intellectual Rights (*optional*), the intellectual property rights regarding usage and licensing of the dataset.
- License name (*mandatory*), the official name of a license that applies to the data and metadata described in this metadata record. The name should match the name of a well-known license from the SPDX license vocabulary or a similar persistent vocabulary.
- License URL (*optional*), the persistent URL for the license, typically a SPDX URL, or an official URL from another well-known license vocabulary. Users should avoid using arbitrary URLs that are not the official URL for a license.

- Acknowledgement (*optional*), the acknowledge funders and other key contributors.

Geographic information

This section includes the metadata related to the geographic description of the dataset. It is a multi-value section, hence it is possible to add one or more values. The included metadata are:

- Geographic description (*mandatory*), the short text description of the geographic areal domain of the dataset.
- Bounding Coordinates (*mandatory*):
  - West Bounding Coordinate (*mandatory*), western-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - East Bounding Coordinate (*mandatory*), eastern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - North Bounding Coordinate (*mandatory*), northern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - South Bounding Coordinate (*mandatory*), southern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - Bounding Altitudes (*optional*):
    Minimum (*mandatory*), the minimum altitude extent of coverage for the bounding box that is being described. The minimum altitude should be in reference to a known datum (e.g., Mean Sea Level), which should be part of the geographic description.
    Maximum (*mandatory*), the maximum altitude extent of coverage for the bounding box that is being described. The maximum altitude should be in reference to a known datum (e.g., Mean Sea Level), which should be part of the geographic description.
    Unit (*mandatory*), the unit that the altitude is expressed in. The value of this metadata field can be selected from the list of the available standard values.

Temporal information

This mandatory section includes the metadata related to the temporal aspect of the dataset. The included metadata are:

- Begin date (*mandatory*), the beginning of some time period in the format dd/mm/YYYY.

- End date (*mandatory*), the end of some time period in the format dd/mm/YYYY.

Taxonomic information

This section is optional, but when the user decides to specify it, all mandatory metadata have to be inserted. It is a multi-value section, hence it is possible to add one or more taxonomic coverage subsections. The included metadata are:

- Taxonomic Rank Name (*mandatory*), the name of the taxonomic rank for which the Taxon rank value is provided. This field allows to specify the name of the accepted levels of Taxa.
- Taxonomic Rank Value (*mandatory*), the taxonomic rank name being described.
- Common Name (*optional*), the specification of applicable common names.
- Taxon ID (*optional*), the identifier for this taxon from an authority.

Technical information

This section is optional, but when the user decides to specify it, all mandatory metadata have to be inserted. The included metadata are:

- Project (*optional*), it contains information on the project in which this dataset was collected. It is optional, but when the user decides to specify it, all mandatory metadata have to be inserted. The included metadata are:
  - Title (*mandatory*), the descriptive title for the research project.
  - Abstract (*optional*), the descriptive abstract that summarizes information about the research project.
  - Personnel (*mandatory and multi-value*), the details of the people involved in the research project. It is a multi-value section, hence it is possible to add one or more persons. The included metadata are:
    - Organisation Name (mandatory), the name of the personnel organization (i.e., the affiliation).
    - Given Name (mandatory), the name of the specific personnel involved in the research project.
    - Surname (mandatory), the surname of the specific personnel involved in the research project.
    - Position Name (mandatory), the title or the position of the specific personnel involved in the research project.
- Method Step
  - Description (*mandatory and multi-value*), the description of the methods employed in collecting or generating the dataset or in quality control

and assurance. It is a multi-value section, hence it is possible to add one or more method steps.

- Citation (*optional*), the literature citation related to the used methods.
- Instrumentation (*optional and multi-value*), the description of any instruments used in the data collection or quality control and quality assurance. The description should include vendor, model number, optional equipment, etc. It is a multi-value section, hence it is possible to add one or more instruments.
- Software (*optional and multi-value*), the software used in the processing of data. It is optional, but when the user decides to specify it, all mandatory metadata have to be inserted. Moreover, it is a multi-value subsection, hence it is possible to add one or more software:
  - Title (*mandatory*), the title of the software used in the processing of data.
  - Version (*mandatory*), the version of the software used in the processing of data.

- Sampling (*optional and multi-value*), the description of sampling procedures including the geographic, temporal and taxonomic coverage of the study. It is optional, but when the user decides to specify it, all mandatory metadata have to be inserted. Moreover, it is a multi-value subsection, hence it is possible to add one or more samplings:
  - Study Extent (*mandatory*), the description of the geographic area sampled (geographic coverage), the sampling frequency (temporal coverage), and living organisms sampled (taxonomic coverage).
  - Sampling Description (*mandatory*), the text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article. It is a multi-value section, hence it is possible to add one or more descriptions.

Data Table information

This section is to document the data table(s) that make up the dataset. A data table could be anything from a Comma Separated Value (CSV) file to a spreadsheet to a table in an RDBMS. It is an optional section, but if the user decides to specify it, all mandatory metadata have to be inserted:

- Name (*mandatory*), the name of the data table.

- Format Name (*mandatory*), the format of the data object.
- Attribute List (*optional and multi-value*), the list of attributes associated with the data table. It is a multi-value section, hence it is possible to specify one or more attributes.
  - Name (*mandatory*), the name of the dataset column.
  - Label (*mandatory*), the descriptive label that can be used to display the name of the attribute.
  - Definition (*mandatory*), the precise definition of the attribute in the data table being documented.
  - Standard Unit (*optional*), the unit of measurement (if available, otherwise dimensionless).
  - Missing Value Code (*optional*), the character(s) used to specify missing values in the dataset. It is repeatable to allow for multiple different codes to be present in the attribute. The included metadata are:
    Code (*mandatory*), the missing value code.
    Explanation (*mandatory*), the explanation of what the missing value code means.
  - Measurement Scale (*optional and multi-value*), the type of scale from which values are drawn for the attribute. This provides information about the scale in which the data was collected.
  - Measurement Type (*mandatory*), the type of the measurement scale. It can be nominal, ordinal, interval, or ratio.
  - Measurement Domain (*optional and multi-value*), it is a multi-value section, hence it is possible to specify one or more domains. The included metadata are:
    Code (*mandatory*), the code value that can be used in the domain.
    Definition (*mandatory*), the explanation of what the measurement domain code means.
  - Annotation (*optional and multi-value*), the precisely-defined semantic statement for the specific attribute. It is a multi-value section, hence it is possible to specify one or more annotations. The included metadata are:
    Property URI (*mandatory*), the persistent URI used to identify a property from a vocabulary.
    Value URI (*mandatory*), the persistent URI used to identify a value from a vocabulary.

3.2 Metadata description for Research Sites

The research site concept is not defined in LUPO. We adopted the definition by Wohner et al. (2019): it represents *an in-situ observation or experimentation facility, delimited in space, but varying in size and complexity of the internal organisational and observational design, for the collection of data covering e.g., biogeophysical, biotic or socio-ecological characteristics and processes*.

The research site metadata schema is based on a subset of the ISO 19139 standard and includes 16 metadata elements. Figure 2 shows the different sections of the profile that are described in detail below.

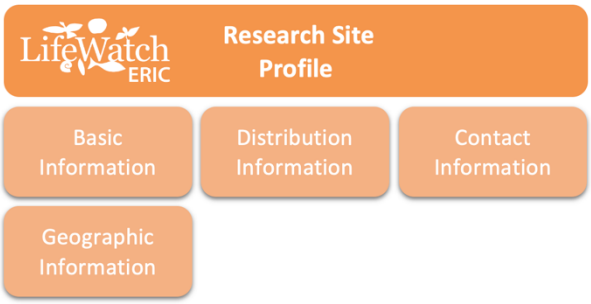

Figure 2: Research site application profile

Basic information

This section includes the basic information related to the site:

- Title (*mandatory*), the name identifying the documented observation and/or experimentation facility (the site).
- Date (*mandatory and multi-value*), the date in which the research site has been created/published/revised (Date Type) which available in the following formats: date and time, year and month, year.
- Abstract (*mandatory*), the short description of the site which includes the location, biophysical characteristics, a brief history, the main current scientific purpose and major plans for future, a brief overall description of infrastructure, and a brief overall description of available data/information.
- Status (mandatory), the current operating status of a site. It is defined among the following options: closed, inapplicable, non-reporting, operational, partly operational, planned, pre-operational, stand-by, unknown (WIGOS code list http://codes.wmo.int/wmdr/ReportingStatus).

- Keywords (*mandatory and multi-value*), the keyword(s) describing the research site, very important for discoverability purposes. It is a multi-value field, hence it is possible to add one or more keywords. Each keyword field should contain one and only one keyword and the use of singular form is recommended.

Distribution information

This section contains the information needed for accessing useful online resources related to the research site available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function. The DOI can be also specified here. It is a multi-value section; hence it is possible to add one or more values. The included metadata are:

- Protocol (*optional*), the connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).
- Linkage (*optional*), the location for online access (i.e., the URL).
- Name (*optional*), the title of the online resource (e.g., DOI, Download page, etc.).

Contact information

This section contains information about the person(s) or organisation(s) responsible for the research site, i.e., the author(s) and the point of contact(s). It is a multi-value section, hence it is possible to add one or more persons. The required metadata are:

- Organisation name (*mandatory*), the name of the person's organisation (affiliation).
- Individual name (*mandatory*), the full name of the person (name and surname).
- Email (*mandatory*), the email address of the person.
- Role (*mandatory*), the specific role of the person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User.

Geographic information

This section includes the metadata related to the geographic description of the research site. It is a multi-value section, hence it is possible to add one or more values. The included metadata are:

- Geographic description (*mandatory*), the short text description of the

geographic areal domain of the site.

- Bounding Coordinates (*mandatory*):
  - West Bounding Coordinate (*mandatory*), western-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - East Bounding Coordinate (*mandatory*), eastern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - North Bounding Coordinate (*mandatory*), northern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.
  - South Bounding Coordinate (*mandatory*), southern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.

3.3 Metadata description for Services

According to the LUPO definitions, a service represents *an intangible good (such as usage of a device or software, access of a data set, training, support for business activities, remote invocation of a software function, or in other possible forms) delivered from one party (namely service provider) to another (service consumer) based on agreement.*

It can be conceived as the continued, declared willingness and ability of an actor to execute, on demand by a client, certain activities of specific benefit to the client. An e-Service is a service automated by nature that reacts to mechanical requests through an information network, receiving inputs, manipulating, and sending back appropriate outputs (ID: https://kos.lifewatch.eu/ontologies/LUPO#Service).

The service metadata schema is based on a subset of the ISO 19139 standard and includes 34 metadata elements. Figure 3 shows the different sections of the profile that are described in detail below.

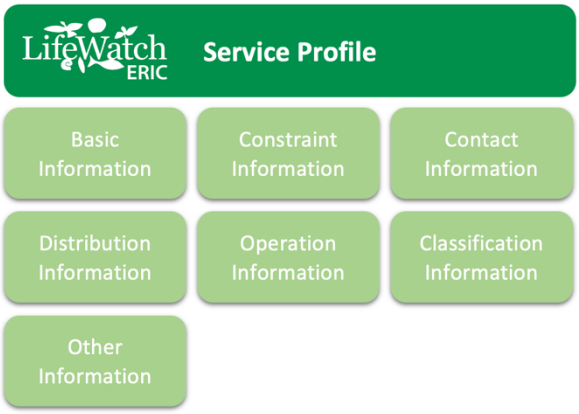

Figure 3: Service application profile

Basic information

This section includes the basic information related to the service:

- Title (*mandatory*), the title of the service.
- Date (*mandatory*), the date in which the service has been created/published/revised (Date Type), which is available in the following formats: date and time, year and month, year.
- Abstract (*mandatory*), the brief description of the service.
- Status (*mandatory*), the status of the service defined among the following options: Completed, Historical archive, Obsolete, Ongoing, Planned, Required or

Under development.

- Version (*mandatory*), the current and last version of the service.
- Keywords (*mandatory and multi-value*), the keyword(s) describing the service, very important for discoverability purposes. It is a multi-value field, hence it is possible to add one or more keywords. Each keyword field should contain one and only one keyword and the use of singular form is recommended.

Constraint information

This section includes metadata related to the license and usage condition of the service. It is a multi-value section and hence it is possible to add one or more resource constraints. The included metadata are:

- Access constraints (*mandatory*), the access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the service (possible values: Copyright, Intellectual property rights, License, Other restrictions, Patent, Pending patent, Restricted, Trademark).
- Use limitation (*mandatory*), the limitation affecting the use of the service. It is recommended to provide a link to a license type (e.g., http://creativecommons.org/licenses/by/4.0), a website or a document containing the necessary information.
- Other constraints (*optional*), other restrictions and legal prerequisites for accessing and using the service.

Contact information

This section contains information about the person(s) or organisation(s) responsible for the service, i.e., the point of contact(s) for the service. It is a multi-value section, hence it is possible to add one or more persons. The required metadata are:

- Organisation name (*mandatory*), the organisation's name of the contact person (i.e., the affiliation).
- Individual name (*mandatory*), the full name of the contact person (name and surname).
- Email (*mandatory*), the email address of the contact person.
- Role (*mandatory*), the role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-

opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-code).

Distribution information

This section contains the information needed for accessing useful online resources related to the service available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function. The DOI can be also specified here. It is a multi-value section; hence it is possible to add one or more values. The included metadata are:

- Protocol (*optional*), the connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).
- Linkage (*optional*), the location for online access (i.e., the URL).
- Name (*optional*), the title of the online resource (e.g., DOI, Download page, etc.).

Operation information

This section includes the metadata related to the operations contained in the service. The included metadata are:

- Operation name (*mandatory*), the title of the operation contained in the service.
- Website (*optional*), the link to the website of the operation.
- Description (*mandatory*), the short description of the operation.
- Function (*mandatory*), the main purpose of the operation.

For services that have specific inputs and outputs, it is possible to add them in this section in order to link the different resources.

Classification information

This section contains both mandatory and optional metadata related to the classification of the service. All mandatory metadata have to be inserted. The included metadata are:

- Related Services (*optional and multi-value*), the service(s) associated to the service. It is a multi-value metadata, hence it is possible to add one or more services.
- Required Services (*optional and multi-value*), the service(s) required to run the service. It is a multi-value metadata, hence it is possible to add one or more services.

- Topic Category (*mandatory and multi-value*), a high-level classification scheme to assist in the grouping and topic-based search of available services. A correct categorisation is very important to help users to search and find the resources they are looking for. Options are provided with a codelist (values from https://inspire-geoportal.ec.europa.eu/theme_selection.html?view=qsTheme). It is a multi-value metadata element, hence it is possible to add one or more categories.
- Service Category (*mandatory and multi-value*), the category(ies) of the service. A correct categorisation is very important to help users to search and find the resources they are looking for. It is a multi-value metadata element, hence it is possible to add one or more categories.
- Service Language (*mandatory*), the main language of the service.
- Other Languages (*optional and multi-value*), the potential other languages of the service. It is a multi-value metadata element, hence it is possible to add one or more languages.
- Service TRL (*mandatory*), the Technology Readiness Level of the service that goes from 1 (Basic principles observed) to 9 (Actual system proven in operational environment).
- Service Level Agreement (*optional*), the SLA associated to the service.

Other information

This section is optional and includes several metadata related to service that are:

- Service Funding (*optional*), the potential funding source of the service.
- Service Price (*optional*), the price of the service (if it is not free).
- Service Order (*optional*), the email address or the web page to the Order of the service.
- Service Helpdesk (*optional*), the reference to the Help Desk of the service (URL or email address).
- Service Training (*optional and multi-value*), the link(s) of the available resources related to the service for training purposes.
- Service User Manual (*optional*), the link to the user manual of the service.

3.4 Metadata description for Workflows

According to the LUPO definitions, a workflow is *the process used to produce a specific research result in a concrete domain. A workflow comprises a series of components that are arranged in a directed graph to set the order of execution of such components. Therefore, a workflow is usually a complex process and may involve both automatic and manual procedures for dealing with research objects and documenting them.*

Workflows may need instantiation in terms of input data and process parameters. A workflow can be composed by one or more workflow components, that are independent, replaceable and modular units that help to manage workflow complexity and encourage their re-use. A component can refer to an executable script or e-service.

The workflow metadata schema is based on a subset of the ISO 19139 standard and includes 24 metadata elements. Figure 4 shows the different sections of the profile that are described in detail below.

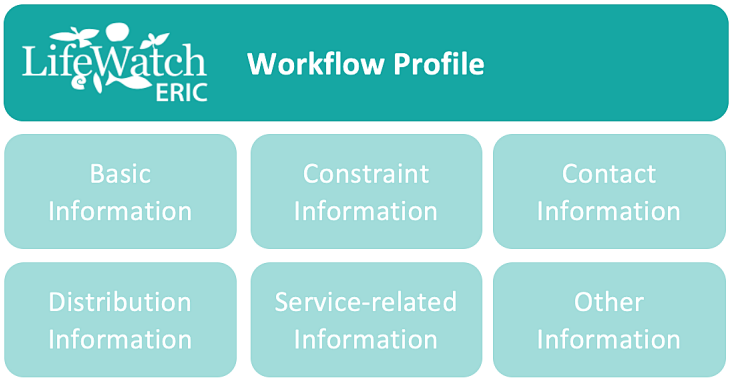

Figure 4: Workflow application profile

Basic information

This section includes the basic information related to the workflow:

- Title (*mandatory*), the title of the workflow.
- Date (*mandatory and multi-value*), the dates in which the workflow has been created/published/revised (Date Type) which available in the following formats: date and time, year and month, year.
- Abstract (*mandatory*), the brief description of the workflow.
- Status (*mandatory*), the status of the workflow defined among the following options: Completed, Historical archive, Obsolete, Ongoing, Planned, Required or

Under development.

- Version (*mandatory*), the current and last version of the workflow.
- Keywords (*mandatory and multi-value*), the keyword(s) describing the workflow, very important for discoverability purposes. It is a multi-value metadata, hence it is possible to add one or more keywords. They have to be put in the appropriate way (one per field) and with the singular form.

Constraint information

This section includes metadata related to the license and usage condition of the workflow. It is a multi-value section and hence it is possible to add more than one resource constraint. The included metadata are:

- Access constraint (*mandatory*), the access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on accessing the workflow (possible values: Copyright, Intellectual property rights, License, Other restrictions, Patent, Pending patent, Restricted, Trademark).
- Use limitation (*mandatory*), the limitation affecting the use of the workflow. It is recommended to provide a link to a license type (e.g., http://creativecommons.org/licenses/by/4.0), a website or a document containing the necessary information.
- Other constraints (*optional*), other restrictions and legal prerequisites for accessing and using the workflow.

Contact information

This section contains information about the person(s) or organisation(s) responsible for the workflow, i.e., the point of contact(s) for the workflow. It is a multi-value section; hence it is possible to add one or more persons. The required metadata are:

- Organisation name (*mandatory*), the organisation's name of the contact person.
- Individual name (*mandatory*), the full name of the contact person.
- Email (*mandatory*), the email address of the contact person.
- Role (*mandatory*), the role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-

code).

Distribution information

This section contains the information needed for accessing useful online resources related to the workflow available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function. The DOI can be also specified here. It is a multi-value section; hence it is possible to add one or more values. The included metadata are:

- Protocol (*optional*), the connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).
- Linkage (*optional*), the location for online access (i.e., the URL).
- Name (*optional*), the title of the online resource (e.g., DOI, Download page, etc.).

Service-related information

This section includes details about the service(s) included in the specific workflow. It is a multi-value section; hence it is possible to add one or more values. The included metadata are:

- Service Name (*mandatory*), the name of the service included in the workflow.
- Service Description (*mandatory*), the description of the service included in the workflow.
- Service Reference (id) (*optional*), the identifier used to refer to the service included in the workflow. It can be also a URL.

Other information

This section is optional and includes several metadata related to workflow that are:

- How to cite this workflow (*optional*), the reference of the workflow.
- Publications about this workflow (*optional and multi-value*), the list of the scientific publications related to the workflow.
- Workflow Helpdesk (*optional*), the reference to the Help Desk of the workflow (URL or email address).
- Workflow Training (*optional and multi-value*), the link(s) of the available training resources.
- Workflow User Manual (*optional*), the link to the user manual of the workflow.

3.5 Metadata description for Virtual Research Environments

According to the LUPO definitions, a Virtual Research Environment (VRE) is *a platform providing user-centric support for discovering and selecting data and software services from different sources, and composing and executing application workflows, also referred to as Virtual Laboratories or Science Gateways.*

VREs play a direct role in the lifecycle of research activities performed by scientists, for example, the planning of experiments, search and discovery of resources from different sources (notably including RIs), integration of services into cohesive workflows and collaboration with other scientists. Graphical environments, workflow management systems, and data analytics tools are typical components of such environments.

The VRE metadata schema is based on a subset of the ISO 19139 standard and includes 24 metadata elements. Figure 5 shows the different sections of the profile that are described in detail below.

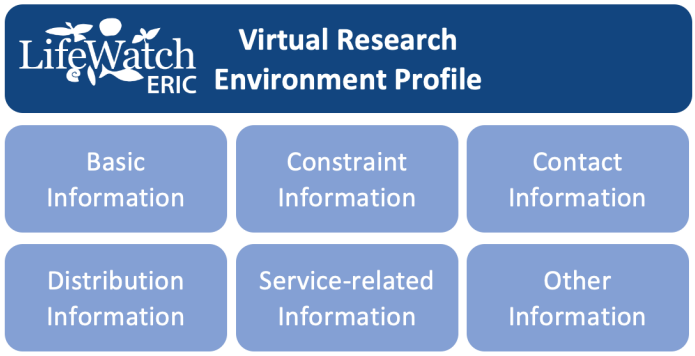

Figure 5: VRE application profile

Basic information

This section includes the basic information related to the VRE:

- Title (*mandatory*), the title of the VRE.
- Date (*mandatory and multi-value*), the dates in which the VRE has been created/published/revised (Date Type), which available in the following formats: date and time, year and month, year.
- Abstract (*mandatory*), the brief description of the VRE.
- Status (*mandatory*), the status of the VRE defined among the following options: completed, historical archive, obsolete, ongoing, planned, required or under development.

- Version (*mandatory*), the current and last version of the VRE.
- Keywords (*mandatory and multi-value*), the keyword(s) describing the VRE, very important for discoverability purposes. It is a multi-value metadata, hence it is possible to add one or more keywords. They have to be put in the appropriate way (one per field) and with the singular form.

Constraint information

This section includes metadata related to the license and usage condition of the VRE. It is a multi-value section and hence it is possible to add more than one resource constraint. The included metadata are:

- Access constraint (*mandatory*), the access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on accessing the VRE (possible values: Copyright, Intellectual property rights, License, Other restrictions, Patent, Pending patent, Restricted, Trademark).
- Use limitation (*mandatory*), the limitation affecting the use of the VRE. It is recommended to provide a link to a license type (e.g., http://creativecommons.org/licenses/by/4.0), a website or a document containing the necessary information.
- Other constraints (*optional*), other restrictions and legal prerequisites for accessing and using the VRE.

Contact information

This section contains information about the person(s) or organisation(s) responsible for the VRE, i.e., the point of contact(s) for the VRE. It is a multi-value section; hence it is possible to add one or more persons. The required metadata are:

- Organisation name (*mandatory*), the organisation's name of the contact person;
- Individual name (*mandatory*), the full name of the contact person (name and surname).
- Email (*mandatory*), the email address of the contact person.
- Role (*mandatory*), the specific role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-

code).

Distribution information

This section contains the information needed for accessing useful online resources related to the VRE available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function. The DOI can be also specified here. It is a multi-value section; hence it is possible to add one or more values. The included metadata are:

- Protocol (*optional*), the connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).
- Linkage (*optional*), the location for online access (i.e., the web address).
- Name (*optional*), the title of the online resource (e.g., DOI, Download page, etc.).

Service-related information

This section includes details about the service(s) included in the specific VRE. It is a multi-value section; hence it is possible to add one or more values. The included metadata are:

- Service Name (*mandatory*), the name of the service included in the VRE.
- Service Description (*mandatory*), the description of the service included in the VRE.
- Service Reference (id) (*mandatory*), the identifier used to refer to the service included in the VRE. It can be also a URL.

Other information

This section is optional and includes several metadata related to VRE that are:

- How to cite this VRE (*optional*), the reference of the VRE.
- Publications about this VRE (*optional and multi-value*), the list of the scientific publications related to the VRE.
- VRE Helpdesk (*optional*), the reference to the Help Desk of the VRE (URL or email address).
- VRE Training (*optional and multi-value*), the link(s) of the available training resources.
- VRE User Manual (*optional*), the link to the user manual of the VRE.

References

Berners-Lee, T. (1997). Web architecture: Metadata. Retrieved from http://www.w3.org/DesignIssues/Metadata.html.

European Commission, Directorate-General for Research and Innovation, Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data, Publications Office, 2018, https://data.europa.eu/doi/10.2777/1524.

Heery, R., & Patel, M. (2000). Application profiles: Mixing and matching metadata schemas. Ariadne, 25. Retrieved from http://www.ariadne.ac.uk/issue25/app-profiles/.

Wilkinson M.D., Dumontier M., Aalbersberg Ij.J., Appleton G., Axton M., Baak A., Mons B., The FAIR guiding principles for scientific data management and stewardship, Sci. Data, 3 (2016), p. 160018, 10.1038/sdata.2016.18.

Wohner C., Peterseil J., Poursanidis D., Kliment T., Wilson M., Mirtl M., Chrysoulakis N. (2019). DEIMS-SDR – A web portal to document research sites and their associated data, Ecological Informatics, Volume 51, 2019, Pages 15-24, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2019.01.005.

Document Revision History

Revision	Reason for and description of change	Author	Date
1.	First version	Lucia Vaira	31-08-2022

Annex A – dataset profile

Metadata Field	Description	Example	Type	Multiplicity	Required
Basic Information					
Identifier	The globally unique identifier within a particular data management system. Typically, a DOI or other identifier that is both citable and resolvable is used.	d6f40b5c-755e-4e61-a4cf-0dc0a124fa87	String	1	Mandatory
Title	The title of the dataset.	Fish community data collected in the Acquatina lagoon, Apulia, Italy, during the years 2007-2008	String	1	Mandatory
Short name	The short name of the dataset.		String	1	Optional
Publication Date	The date when the dataset was or will be made publicly available in the format dd/mm/YYYY.	21/01/2021	Date (dd/mm/yyyy)	1	Optional
Dataset Language	The language in which the dataset is written.	eng	List of controlled values	1	Mandatory
Abstract	The brief description of the dataset, providing enough detail to differentiate it from other similar resources.	The dataset contains abundance and body size data of fish community in Acquatina lagoon. Forty-five taxa were recorded, encompassing 9 orders and 20 families, all identified at the species level except two individuals, identified respectively at the genus (Pomatoschistus) and family (Labridae) level. A total amount of 99,995 individuals were collected, and the community was dominated by the species Atherina boyeri, with 95,237 individuals found.	String	1	Optional
KEYWORDSET				Multiple	Mandatory
Keyword	Keywords or key phrases that concisely describe the dataset or are related to the dataset. Each keyword field should contain one and only one keyword and the use of the singular form is recommended.	Biodiversity	String	Multiple	Mandatory
Keyword Thesaurus	The name of a thesaurus from which the set of keywords is derived.	None	String	1	Optional
DISTRIBUTION				Multiple	Optional
Online Description	The description of the content of the URL.	DOI	String	1	Optional
URL	The URL associated to the dataset. The DOI can be also specified here.	https://doi.org/10.48372/d6f40b5c-755e-4e61-a4cf-0dc0a124fa87	URL	1	Mandatory
Person / Organization				**Multiple**	**Mandatory**
CREATOR				Multiple	Mandatory
Organization Name	The affiliation of the person who created the dataset.	University of Salento	String	1	Mandatory
First Name	The first name of the person who created the dataset.	Alberto	String	1	Mandatory
Surname	The surname of the person who created the dataset.	Basset	String	1	Mandatory
Position Name	The position name of the dataset creator.	Full professor	String	1	Mandatory
Email	The emal address of the person who created the dataset.	alberto.basset@unisalento.it	Email	1	Mandatory
User ID	The identifier that links the dataset creator to a directory of individuals (e.g., ORCID).		URL	1	Optional
METADATA PROVIDER				Multiple	Optional
Organization Name	The affiliation of the dataset metadata provider.	Italian National Research Council	String	1	Mandatory
First Name	The first name of the dataset metadata provider.	Ilaria	String	1	Mandatory
Surname	The surname of the dataset metadata provider.	Rosati	String	1	Mandatory
Position Name	The position name of the dataset metadata provider.	Research Technologist	String	1	Mandatory
Email	The email address of the person who provided the dataset metadata.	ilaria.rosati@cnr.it	Email	1	Mandatory
User ID	The identifier that links the metadata provider to a directory of individuals (e.g., ORCID).		URL	1	Optional
ASSOCIATED PARTY				Multiple	Optional
Organization Name	The affiliation of other people or organizations who should be associated with the dataset.		String	1	Mandatory
First Name	The name of other people or organizations who should be associated with the dataset.		String	1	Mandatory
Surname	The surname of other people or organizations who should be associated with the dataset.		String	1	Mandatory
Position Name	The position name of the person with this role.		String	1	Mandatory
Email	The email address of other people or organizations who should be associated with the dataset.		Email	1	Mandatory
User ID	The identifier that links this party to a directory of individuals (e.g., ORCID).		URL	1	Optional
CONTACT				Multiple	Mandatory
Organization Name	The affiliation of the contact point for the dataset.	University of Salento	String	1	Mandatory
First Name	The name of the contact point for the dataset.	Alberto	String	1	Mandatory
Surname	The surname of the contact point for the dataset.	Basset	String	1	Mandatory
Email	The email address of the contact point for the dataset.	Full professor	String	1	Mandatory
Position Name	The position name of the person with this role.	alberto.basset@unisalento.it	Email	1	Mandatory
User ID	The identifier that links this contact point to a directory of individuals (e.g., ORCID).		URL	1	Optional

License Information					
Intellectual Rights	The intellectual property rights regarding usage and licensing of the dataset.	This work is licensed under a Creative Commons Attribution Non Commercial (CC-BY-NC) 4.0 License.	String	1	Optional
License Name	The official name of a license that applies to the data and metadata described in this metadata record. The name should match the name of a well-known license from the SPDX license vocabulary or a similar persistent vocabulary.	CC-BY-NC 4.0	String	1	Mandatory
License URL	The persistent URL for the license, typically a SPDX URL, or an official URL from another well-known license vocabulary. Users should avoid using arbitrary URLs that are not the official URL for a license.		URL	1	Optional
Acknowledgement	The acknowledge funders and other key contributors.		String	1	Optional
Geographic Information				Multiple	Mandatory
Geographic Description	The short text description of the geographic areal domain of the dataset.	Acquatina is a brackish coastal lagoon located in the south Apulia (Italy).	String	1	Mandatory
BOUNDING COORDINATES				1	Mandatory
West Bounding Coordinate	Western-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.	18.233028	Decimal	1	Mandatory
East Bounding Coordinate	Eastern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.	18.233028	Decimal	1	Mandatory
North Bounding Coordinate	Northern-most lilmit of a bounding box expressed in latitude, WGS84 standard.	40.448167	Decimal	1	Mandatory
South Bounding Coordinate	Southern-most limit of the bounding box expressed in latitude, WGS84 standard.	40.448167	Decimal	1	Mandatory
BOUNDING ALTITUDES				1	Optional
Minimum	The minimum altitude extent of coverage for the bounding box that is being described. The minimum altitude should be in reference to a known datum (e.g., Mean Sea Level), which should be part of the geographic description.		Decimal	1	Mandatory
Maximum	The maximum altitude extent of coverage for the bounding box that is being described. The maximum altitude should be in reference to a known datum (e.g., Mean Sea Level), which should be part of the geographic description.		Decimal	1	Mandatory
Unit	The unit that the altitude is expressed in. The value of this metadata field can be selected from the list of the available standard values.		List of controlled values	1	Mandatory
Temporal Information				1	Mandatory
Begin Date	The beginning of some time period in the format dd/mm/YYYY.	07/06/2007	Date (dd/mm/yyyy)	1	Mandatory
End Date	The end of some time period in the format dd/mm/YYYY.	020/05/2008	Date (dd/mm/yyyy)	1	Mandatory
Taxonomic Information				Multiple	Optional
Taxonomic Rank Name	The name of the taxonomic rank for which the Taxon rank value is provided. This field allows to specify the name of the accepted levels of Taxa.	Species	String	1	Mandatory
Taxonomic Rank Value	The taxonomic rank name being described.	Anguilla anguilla	String	1	Mandatory
Common Name	The specification of applicable common names.		String	1	Optional
Taxon ID	The identifier for this taxon from an authority.		URL	1	Optional
Technical Information					
PROJECT				1	Optional
Title	The descriptive title for the research project.		String	1	Mandatory
Abstract	The descriptive abstract that summarizes information about the research project.		String	1	Optional
PERSONNEL				Multiple	Optional
Organization Name	The name of the personnel organization (i.e., the affiliation).		String	1	Mandatory
Given Name	The name of the specific personnel involved in the research project.		String	1	Mandatory
Surname	The surname of the specific personnel involved in the research project.		String	1	Mandatory
Position Name	The name of the title or the position for people involved in the research project.		String	1	Mandatory

METHODS				Multiple	Optional
Description	The description of the methods employed in collecting or generating the dataset or in quality control and assurance.	Four sampling stations were located in the Acquatina lagoon at increasing distances from the inlet, following a gradient of confinement from station 1 to station 4. Fish were collected by letting down a fyke net for 24 h. The net was made up of three hoop-nets equipped with wings, set up radially at the end of a 20-m-long blocking net tethered to the shore. Each hoop net had two funnels, and a mesh size ranging from 12 mm for the wings to 6 mm for the cod end.	String	1	Mandatory
Citation	The literature citation related to the used methods.		String	1	Optional
Instrumentation	The description of any instruments used in the data collection or quality control and quality assurance. The description should include vendor, model number, optional equipment, etc.		String	Multiple	Optional
SOFTWARE				Multiple	Optional
Title	The title of the software used in the processing of data.		String	1	Mandatory
Version	The version of the software used in the processing of data.		String	1	Mandatory
SAMPLING				Multiple	Optional
Study Extent	The description of the geographic area sampled (geographic coverage), the sampling frequency (temporal coverage), and living organisms sampled (taxonomic coverage).	Four sampling stations were located in the Acquatina lagoon at increasing distances from the inlet, following a gradient of confinement from station 1 to station 4. Sampling was carried out twice per month, from June 2007 to May 2008, except August when only one sampling was performed at the end of the month, making a total of 23 sampling times. Four sampling stations were located in the Acquatina lagoon at increasing distances from the inlet, following a gradient of confinement from station 1 to station 4. Sampling was carried out twice per month, from June 2007 to May 2008, except August when only one sampling was performed at the end of the month, making a total of 23 sampling times.	String	1	Mandatory
Sampling Description	The text-based/human readable description of the adopted sampling procedures. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article.	Fish were collected by letting down a fyke net for 24 h. The net was made up of three hoop-nets equipped with wings, set up radially at the end of a 20-m-long blocking net tethered to the shore. Each hoop net had two funnels, and a mesh size ranging from 12 mm for the wings to 6 mm for the cod end.	String	1	Mandatory
Data Table Information				Multiple	Optional
Name	The name of the data table.	Fish_Acquatina_2007_2008.csv	String	1	Mandatory
Format Name	The format of the data object.	csv	String	1	Mandatory
ATTRIBUTE LIST				Multiple	Optional
Name	The name of the dataset column.	parenteventid	String	1	Mandatory
Label	The descriptive label that can be used to display the name of the attribute.	parenteventid	String	1	Mandatory
Definition	The precise definition of the attribute in the data table being documented.	An identifier for the broader Event that groups this and potentially other Events.	String	1	Mandatory
Standard Unit	The unit of measurement (if available, otherwise dimensionless).		List of controlled values	1	Optional
MISSING VALUE CODE	The character(s) used to specify missing values in the dataset.			Multiple	Optional
Code	The missing value code.		String	1	Mandatory
Explanation	The explanation of what the missing value code means.		String	1	Mandatory
MEASUREMENT SCALE	The type of scale from which values are drawn for the attribute. This provides information about the scale in which the data was collected.			1	Optional
Measurement Type	The type of the measurement scale. It can be nominal, ordinal, interval, or ratio.	nominal	List of controlled values	1	Mandatory
MEASUREMENT DOMAIN				Multiple	Optional
Code	The code value that can be used in the domain.	1	String	1	Mandatory
Definition	The explanation of what the measurement domain code means.	Station 1	String	1	Mandatory
ANNOTATION	The precisely-defined semantic statement for the specific attribute.			Multiple	Optional
Property URI	The persistent URI used to identify a property from a vocabulary.		URL	1	Mandatory
Value URI	The persistent URI used to identify a value from a vocabulary.		URL	1	Mandatory

Annex B – research site profile

Metadata Field	Description	Example	Type	Multiplicity	Required
Basic Information					
Title	The name identifying the documented observation and/or experimentation facility (the site).	Puéchabon	String	1	Mandatory
Date	The date in which the site has been created/published/revised, which available in the following formats: date and time, year and month, year.	2021-03-11	Date	Multiple	Mandatory
Abstract	The short description of the site which includes the location, biophysical characteristics, a brief history, the main current scientific purpose and major plans for future, a brief overall description of infrastructure, and a brief overall description of available data/information.	The site comprises three distincts experimental set-ups: (1) a long-term (>10 years) partial throughfall exclusion experiment replicated three times and crossed with a thinning (-30% basal area) experiment aimed at simulating long-term precipitation decrease in accordance with climate change scenario for the Mediterranean area (-30% of precipitation), (2) a total rainfall exclusion experiment using a mobile roof has been set up to simulate extreme drought events and modify precipitation seasonality, and (3) an eddy-covariance flux tower running continuously since 2001 to measure seasonal variations in ecosystem functioning and year-to-year flux responses to drought and climate.	String	1	Mandatory
Status	The current operating status of a site. It is defined among the following options: closed, inapplicable, non-reporting, operational, partly operational, planned, pre-operational, stand-by, unknown (WIGOS code list http://codes.wmo.int/wmdr/ReportingStatus).	Operational	List of controlled values	1	Mandatory
Keywords	The keyword(s) describing the site, very important for discoverability purposes. Each keyword field should contain one and only one keyword and the use of singular form is recommended.	soil water content soil temperature	String	Multiple	Mandatory
Distribution Information	It is a metadata element for accessing useful online resources available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function.			Multiple	Optional
Protocol	The connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).	Web address (URL)	String	1	Optional
Linkage	The location for online access (i.e., the URL).	https://www.anaee-france.fr/service/experimentation-in-natura/ecosystemes-forestier/ecosystemes-forestiers-mediterraneens/puechabon	URL	1	Optional
Name	The title of the online resource (e.g., DOI, Download page, etc.).	Site description page	String	1	Optional
Contact Information				Multiple	Mandatory
Organisation name	The organisation's name of the contact person (i.e., the affiliation).	CNRS	String	1	Mandatory
Individual name	The full name of the contact person (name and surname).	Jean-Marc LIMOUSIN	String	1	Mandatory
Email	The email address of the contact person.	jean-marc.limousin@cefe.cnrs.fr	Email	1	Mandatory
Role	The role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-code).	Point of contact	List of controlled values	1	Mandatory
Geographic Information				Multiple	Mandatory
Geographic Description	The short text description of the geographic areal domain of the research site.	Puéchabon	String	1	Mandatory
BOUNDING COORDINATES				1	Mandatory
West Bounding Coordinate	Western-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.	35.957	Decimal	1	Mandatory
East Bounding Coordinate	Eastern-most limit of a bounding box, expressed in degrees of longitude, WGS84 standard.	35.957	Decimal	1	Mandatory
North Bounding Coordinate	Northern-most lilmit of a bounding box expressed in latitude, WGS84 standard.	437.413	Decimal	1	Mandatory
South Bounding Coordinate	Southern-most limit of the bounding box expressed in latitude, WGS84 standard.	437.413	Decimal	1	Mandatory

Annex C – service profile

Metadata Field	Description	Example	Type	Multiplicity	Required
Basic Information					
Title	The title of the service.	WoRMS taxon match webservices (LifeWatch.be)	String	1	Mandatory
Date	The date in which the service has been created/published/revised, which available in the following formats: date and time, year and month, year.	2019-05-10	Date	Multiple	Mandatory
Abstract	The brief description of the service.	As a user or developer you can use the WoRMS webservice to feed your own application with standard WoRMS data. A non exhaustive list of applications: - get the AphiaID for your taxon - check the spelling of your taxa - get the authority for your taxa - get the full classification for your taxa - resolve your unaccepted names to accepted ones - get all synonyms for a taxon - fuzzy/near match your species list - resolve a common name/vernacular to a scientific name - get the common name(s)/vernacular(s) for a taxon - get the sources/references for a taxon - get the WoRMS citation for a taxon - get the direct children for a taxon - get all taxa modified during a time interval - get an external identifier for a taxon - get the AphiaID for an external identifier/database - get all distributions for a taxon - get all attributes for a taxon	String	1	Mandatory
Status	The status of the service defined among the following options: Completed, Historical archive, Obsolete, Ongoing, Planned, Required or Under development.	Ongoing / operational	List of controlled values	1	Mandatory
Version	The current and last version of the service.	Not available	String	1	Mandatory
Keywords	The keyword(s) describing the service, very important for discoverability purposes. Each keyword field should contain one and only one keyword and the use of singular form is recommended.	marine species register synonymy	String	Multiple	Mandatory
Constraint Information				Multiple	Mandatory
Access constraints	The access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the service (possible values: Copyright, Intellectual property rights, License, Other restrictions, Patent, Pending patent, Restricted, Trademark).	License	List of controlled values	1	Mandatory
Use limitation	The limitation affecting the use of the service. It is recommended to provide a link to a license type (e.g., http://creativecommons.org/licenses/by/4.0), a website or a document containing the necessary information.	CC-BY	String	1	Mandatory
Other constraints	Other restrictions and legal prerequisites for accessing and using the service.	The text on the WoRMS pages is open-access under the terms of the Creative Commons Attribution License (CC-BY). This License permits unrestricted use, provided it is cited as requested on the WoRMS webpages, unless stated otherwise on the individual pages. Images are by default open-access under the terms of the CC BY-NC-SA license, unless stated otherwise. Re-distribution of the entire database is not permitted, unless by prior written agreement. This is mainly to avoid circulation of (quickly) outdated copies of WoRMS, accessible through different pathways which can lead to confusion for our many users.	String	1	Optional
Contact Information				Multiple	Mandatory
Organisation name	The organisation's name of the contact person (i.e., the affiliation).	Flanders Marine Institute - LifeWatch.be	String	1	Mandatory
Individual name	The full name of the contact person (name and surname).	VLIZ	String	1	Mandatory
Electronic mail address	The email address of the contact person.	info@marinespecies.org	Email	1	Mandatory
Role	The role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-code).	Author	List of controlled values	1	Mandatory

Distribution Information	It is a metadata element for accessing useful online resources available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function.			Multiple	Optional
Protocol	The connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).	Web address (URL)	String	1	Mandatory
Linkage	The location for online access (i.e., the URL).	http://www.marinespecies.org/aphia.php?p=search	URL	1	Mandatory
Name	The title of the online resource (e.g., DOI, Download page, etc.).	Link to the service	String	1	Mandatory
Operation Information				Multiple	Mandatory
Operation name	The title of the operation contained in the service.	Aphia platform	String	1	Mandatory
Web site	The link to the website of the operation.	http://www.marinespecies.org/about.php#what_is_aphia	URL	1	Optional
Description	The short description of the operation.	The Aphia platform is an infrastructure designed to capture taxonomic and related data and information, and includes an online editing environment. It is the core platform that underpins the World Register of Marine Species (WoRMS) and all its related global, regional and thematic species databases, but it also allows the storage of non-marine data. The Aphia platform is an MS SQL database, containing over 400 fields spread over more than 80 related tables. Content-wise, the Aphia structure can roughly be divided into 10 modules: taxonomy, distribution, traits, specimen information, vernacular names, notes, links, images, identification keys and sources.	String	1	Mandatory
Function	The main purpose of the operation.	Aphia uses unique and stable identifiers for each available name in the database, through Life Science Identifiers (LSIDs). The system not only allows the storage of accepted and unaccepted names, but it also documents the relationship between names. This makes it a very powerful tool for taxonomic quality control, and also allows the linking of different pieces of information through scientific names, both within the Aphia platform and in relation to externally hosted databases.	String	1	Mandatory
Classification Information					
Related Services	The service(s) associated to the service.		String	Multiple	Optional
Required Services	The service(s) required to run the service.		String	Multiple	Optional
Topic Category	The high-level classification scheme to assist in the grouping and topic-based search of available services. A correct categorisation is very important to help users to search and find the resources they are looking for. Options are provided with a codelist (https://inspire-geoportal.ec.europa.eu/theme_selection.html?view=qsTheme) .	Species distribution	List of controlled values	Multiple	Mandatory
Service Category	The category(ies) of the service. A correct categorisation is very important to help users to search and find the resources they are looking for.	Data analysis	String	Multiple	Mandatory
Service Language	The main language of the service.	eng	List of controlled values	1	Mandatory
Other Languages	The potential other languages of the service.		List of controlled values	Multiple	Optional
Service TRL	The Technology Readiness Level of the service from 1 (Basic principles observed) to 9 (Actual system proven in operational environment).	TRL 9 – Actual system proven in operational environment	List of controlled values	1	Mandatory
Service Level Agreement	The SLA associated to the service.		String	1	Optional
Other Information				1	Optional
Service Funding	The potential funding source of the service.	The maintenance and further development of WoRMS relies on financial contributions, the time contributed by its editorial board, and support of its host institution: VLIZ. WoRMS is currently funded through the LifeWatch Belgium project.	String	1	Optional
Service Price	The price of the service (if it is not free).	Free	String	1	Optional
Service Order	The email address or the web page to the Order of the service.	webservice-subscribe@marinespecies.org	String	1	Optional
Service Helpdesk	The reference to the Help Desk of the service (URL or email address).	webservice-subscribe@marinespecies.org	String	1	Optional
Service Training	The link(s) of the available resources related to the service for training purposes.	http://www.marinespecies.org/aphia.php?p=webservice	String	Multiple	Optional
Service User Manual	The link to the user manual of the service.	http://www.marinespecies.org/aphia.php?p=manual	String	1	Optional

Annex D – workflow profile

Metadata Field	Description	Example	Type	Multiplicity	Required
Basic Information					
Title	The title of the workflow.	Functional biogeography of invaders: the case of two widely distributed omnivorous crustaceans	String	1	Mandatory
Date	The date in which the workflow has been created/published/revised (available formats: date and time, year and month, year).	2019	Date	Multiple	Mandatory
Abstract	The brief description of the workflow.	Biological invasions are to date acknowledged as significant environmental and economic threats, yet the identification of key ecological traits determining species invasiveness has remained elusive. One unappreciated source of variation concerns dietary flexibility and the ability to shift trophic position within invaded food webs. Trophic plasticity may greatly influence invasion success as it negates resource availability as a functional constraint for introduced individuals; in addition, the impact of an invader might increase with its dietary plasticity since a greater proportion of the resident assemblage of species and multiple trophic levels are affected. The validation case will focus on two invasive crustaceans widely distributed in marine and freshwater European waters, i.e., the Atlantic blue crab Callinectes sapidus and the Louisiana crayfish Procambarus clarkii. Key questions that will be addressed are: a) does the trophic position of C. sapidus vary between invaded Mediterranean and native west Atlantic habitats, testifying the occurrence of significant post-invasion shifts in dietary habits? b) does the trophic position of P. clarkii vary in invaded habitats, and which are the main ecological factors involved? c) do bioclimatic drivers influence broad-scale spatial patterns of variation in the trophic position of C. sapidus and P. clarkii?	String	1	Mandatory
Status	The status of the workflow defined among the following options: Completed, Historical archive, Obsolete, Ongoing, Planned, Required or Under development.	Under development	List of controlled values	1	Mandatory
Version	The current and last version of the workflow.	1.0	String	1	Mandatory
Keywords	The keyword(s) describing the workflow, very important for discoverability purposes. It is a multi-value metadata, hence it is possible to add one or more keywords. They have to be put in the appropriate way (one per field) and with the singular form.	crustaceans	String	Multiple	Mandatory
Constraint Information				Multiple	Mandatory
Access constraints	The access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on accessing the workflow (possible values: Copyright, Intellectual property rights, License, Other restrictions, Patent, Pending patent, Restricted, Trademark).	Copyrigth	List of controlled values	1	Mandatory
Use limitation	The limitation affecting the use of the workflow. It is recommended to provide a link to a license type (e.g., http://creativecommons.org/licenses/by/4.0), a website or a document containing the necessary information.	All rights reserved Copyright LifeWatch ERIC - © 2018	String	1	Mandatory
Other constraints	Other restrictions and legal prerequisites for accessing and using the workflow.		String	1	Optional
Contact Information				Multiple	Mandatory
Organisation name	The organisation's name of the contact person (i.e., the affiliation).	University of Salento, Lecce - Italy	String	1	Mandatory
Individual name	The full name of the contact person (name and surname).	Giorgio Mancinelli	String	1	Mandatory
Email	The email address of the contact person.	giorgio.mancinelli@unisalento.it	Email	1	Mandatory
Role	The role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-code).	Creator	List of controlled values	1	Mandatory

Distribution Information	It is a metadata element for accessing useful online resources available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function.			Multiple	Optional
Protocol	The connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).	Web address (URL)	String	1	Optional
Linkage	The location for online access (i.e., the URL).	https://www.lifewatch.eu/internal-joint-initiative/validation-cases/functional-biogeography-of-invasive-species-the-case-of-two-widely-distributed-omnivorous-crustaceans/	URL	1	Optional
Name	The title of the online resource (e.g., DOI, Download page, etc.).	Info page	String	1	Optional
Service-related Information				Multiple	Mandatory
Service name	The name of the service included in the workflow.	Crustaceans WoRMS Taxonomic Checker	String	1	Mandatory
Service Description	The description of the service included in the workflow.	A service that aims at analyzing the trophic position of the "invader" by using a Bayesian model to estimate the NA values of the trophic position that corresponds to the column 13 "TP" of the ""Tax_validated_SIA.csv" file by means of a loop that allows extracting data of single populations. It represents the Step 2.	String	1	Mandatory
Service Reference (id)	The identifier used to refer to the service included on the workflow. It can be also a URL.	https://metadatacatalogue.lifewatch.eu/srv/eng/catalog.search#/metadata/cfa5a24b-5912-48af-8244-bdf301d77b41	String	1	Mandatory
Other Information				1	Optional
How to cite this workflow	The reference of the workflow.		String	1	Optional
Publications About this workflow	The list of the scientific publications related to this workflow.		String	Multiple	Optional
Workflow Helpdesk	The reference to the Help Desk of the workflow (URL or email address).		String	1	Optional
Workflow Training	The link(s) of the available training resources.		String	Multiple	Optional
Workflow User Manual	The link to the user manual of the workflow.		String	1	Optional

Annex E – Virtual Research Environment profile

Metadata Field	Description	Example	Type	Multiplicity	Required
Basic Information					
Title	The title of the VRE.	Phytoplankton Virtual Research Environment	String	1	Mandatory
Date	The date in which the VRE has been created/published/revised (available formats: date and time, year and month, year).	2019	Date	Multiple	Mandatory
Abstract	The brief description of the VRE.	The Phytoplankton Virtual Research Environment (Phyto VRE) is a collaborative working environment supporting researchers to address basic and applied studies on phytoplankton ecology.	String	1	Mandatory
Status	The status of the VRE defined among the following options: Completed, Historical archive, Obsolete, Ongoing, Planned, Required or Under development.	Complete	List of controlled values	1	Mandatory
Version	The current and last version of the VRE.	1.0	String	1	Mandatory
Keywords	The keyword(s) describing the VRE, very important for discoverability purposes. It is a multi-value metadata, hence it is possible to add one or more keywords. They have to be put in the appropriate way (one per field) and with the singular form.	phytoplankton	String	Multiple	Mandatory
Constraint Information				Multiple	Mandatory
Access constraints	The access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on accessing the VRE (possible values: Copyright, Intellectual property rights, License, Other restrictions, Patent, Pending patent, Restricted, Trademark).	Copyrigth	List of controlled values	1	Mandatory
Use limitation	The limitation affecting the use of the VRE. It is recommended to provide a link to a license type (e.g., http://creativecommons.org/licenses/by/4.0), a website or a document containing the necessary information.	All rights reserved Copyright LifeWatch ERIC - © 2018	String	1	Mandatory
Other constraints	Other restrictions and legal prerequisites for accessing and using the VRE.	Permitted: - view, download, copy, print and save search results; - view, download, copy, print and save individual articles. Not permitted: - use e-resources for commercial gain; transmit / disseminate online content to unauthorized users.	String	1	Optional
Contact Information				Multiple	Mandatory
Organisation name	The organisation's name of the contact person (i.e., the affiliation).	University of Salento, Lecce - Italy	String	1	Mandatory
Individual name	The full name of the contact person (name and surname).	Elena Stanca	String	1	Mandatory
Email	The email address of the contact person.	elena.stanca@unisalento.it	Email	1	Mandatory
Role	The role of the contact person defined among the following options: Associated party, Author, Creator, Custodian, Distributor, Owner, Point of Contact, Principal Investigator, Processor, Publisher, Resource Provider, User (here the definition of the different roles: https://geonetwork-opensource.org/manuals/3.10.x/en/annexes/standards/iso19139.html#role-code).	Author	List of controlled values	1	Mandatory
Distribution Information	It is a metadata element for accessing useful online resources available through the web. This allows to specify the online access using an URL address or similar addressing scheme and to provide the protocol for the proper connection for accessing any types of digital documents using the download function.			Multiple	Optional
Protocol	The connection protocol to be used (e.g., DOI, Web address (URL), File for download, etc.).	Web address (URL)	String	1	Optional
Linkage	The location for online access (i.e., the URL).	http://www.servicecentrelifewatch.eu/web/phytoplankton-vre/home	URL	1	Optional
Name	The title of the online resource (e.g., DOI, Download page, etc.).	Info page	String	1	Optional

Service-related Information				Multiple	Mandatory
Service name	The name of the service included in the VRE.	Atlas of Phytoplankton	String	1	Mandatory
Service Description	The description of the service included in the VRE.	This Atlas is a reference point for marine, transitional and freshwater scientists and students involved in phytoplankton identification and classification. It includes illustrative cards with information about taxonomy, ecological characteristics and geographical distribution of species; and morphological features.	String	1	Mandatory
Service Reference (id)	The identifier used to refer to the service included on the VRE. It can be also a URL.	www.atlas.it	String	1	Mandatory
Other Information				1	Optional
How to cite this VRE	The reference of the VRE.	Elena Stanca, Leonilde Roselli, Francesco Cozzoli, Ilaria Rosati, Nicola Fiore and Alberto Basset (2017) "Phytoplankton morphological and demographic traits computation in a virtual research environment Phyto VRE", Abstract book of 52nd European Marine Biology Symposium, September 2017 Slovenia. ISBN 978-961-93486-5-9.	String	1	Optional
Publications About this VRE	The list of the scientific publications related to this VRE.	Elena Stanca, Leonilde Roselli, Francesco Cozzoli, Ilaria Rosati, Nicola Fiore and Alberto Basset (2017) "Phytoplankton morphological and demographic traits computation in a virtual research environment Phyto VRE", Abstract book of 52nd European Marine Biology Symposium, September 2017 Slovenia. ISBN 978-961-93486-5-9.	String	Multiple	Optional
VRE Helpdesk	The reference to the Help Desk of the VRE (URL or email address).	https://www.lifewatch.eu/web/guest/help-desk	String	1	Optional
VRE Training	The link(s) of the available training resources.	http://training.servicecentrelifewatch.eu/course/index.php?categoryid=28	String	Multiple	Optional
VRE User Manual	The link to the user manual of the VRE.	http://training.servicecentrelifewatch.eu/course/index.php?categoryid=28	String	1	Optional