

USEit Data Management Plan

Version 0

Description

The overarching aim of the project USEit is to propose shared and integrated approaches to improve the management of Invasive Alien Species (IAS) in Italy, in both terrestrial and aquatic habitats. This objective is achieved through the involvement of a multidisciplinary team belonging to three different institutes of the National Research Council (CNR) of Italy (*i.e.*, IRBIM [Institute for Marine Biological Resources and Biotechnology], IRET [Research Institute on Terrestrial Ecosystems], and IAS [Institute for the Study of Anthropic Impact and Sustainability in the Marine Environment]).

The USEit work plan is dedicated to the development of new capabilities for IAS management, including their use as a potential economic resource, through the adoption of shared monitoring approaches and the implementation of Open and FAIR practices for data management.

The purpose of this Data Management Plan (DMP) is to describe the data management practices adopted within the project and to facilitate the sharing and reuse of data, and other research products, produced or made available through the project. This DMP includes six datasets describing in detail how different IAS data types are managed within the project: i) IAS distribution and abundance data; ii) IAS stable isotope data; iii) IAS telemetry data; iv) IAS satellite data; v) DNA metabarcoding data of mycorrhizae associated with invasive plant roots; vi) IAS nutraceutical and elemental composition analysis.

This DMP is intended as a live document and it is regularly updated throughout the lifespan of the project.

<https://www.lifewatchitaly.eu/progetti-correlati/useit/>

Funder

National Research Council of
Italy (CNR)

Grant

Progetto SAC.AD002.173

Researchers

Giuseppe Scarcella (orcid:0000-0001-6540-1997), Carlo
Pipitone (orcid:0000-0002-7632-1228),
VINCENZOMAXIMILIANO GIACALONE (orcid:0000-0002-4316-
1723), MARCO CIOLFI (orcid:0000-0003-4831-8053), Ilaria
Rosati (orcid:0000-0003-3422-7230), LAURA LORU
(orcid:0000-0001-6443-9926), Pierluigi Strafella (orcid:0000-
0002-1947-7023), Ernesto Azzurro (orcid:0000-0002-9805-
3164), MARCO LAUTERI (orcid:0000-0003-1071-7999), Paola
Pollegioni (orcid:0000-0001-6388-1931), Cristina Di Muri
(orcid:0000-0003-4072-0662), Marina Chiappi, giorgio
mancinelli (orcid:0000-0002-5833-7322)

Organizations

National Research Council of Italy (CNR)

1. Main Info

Title of DMP: [USEit Data Management Plan](#)

Description:

The overarching aim of the project USEit is to propose shared and integrated approaches to improve the management of Invasive Alien Species (IAS) in Italy, in both terrestrial and aquatic habitats. This objective is achieved through the involvement of a multidisciplinary team belonging to three different institutes of the National Research Council (CNR) of Italy (*i.e.*, IRBIM [Institute for Marine Biological Resources and Biotechnology], IRET [Research Institute on Terrestrial Ecosystems], and IAS [Institute for the Study of Anthropic Impact and Sustainability in the Marine Environment]).

The USEit work plan is dedicated to the development of new capabilities for IAS management, including their use as a potential economic resource, through the adoption of shared monitoring approaches and the implementation of Open and FAIR practices for data management.

The purpose of this Data Management Plan (DMP) is to describe the data management practices adopted within the project and to facilitate the sharing and reuse of data, and other research products, produced or made available through the project. This DMP includes six datasets describing in detail how different IAS data types are managed within the project: i) IAS distribution and abundance data; ii) IAS stable isotope data; iii) IAS telemetry data; iv) IAS satellite data; v) DNA metabarcoding data of mycorrhizae associated with invasive plant roots; vi) IAS nutraceutical and elemental composition analysis.

This DMP is intended as a live document and it is regularly updated throughout the lifespan of the project.

<https://www.lifewatchitaly.eu/progetti-correlati/useit/>

Researchers:

Giuseppe Scarcella (orcid:0000-0001-6540-1997)

Carlo Pipitone (orcid:0000-0002-7632-1228)

VINCENZOMAXIMILIANO GIACALONE (orcid:0000-0002-4316-1723)

MARCO CIOLFI (orcid:0000-0003-4831-8053)

Ilaria Rosati (orcid:0000-0003-3422-7230)

LAURA LORU (orcid:0000-0001-6443-9926)

Pierluigi Strafella (orcid:0000-0002-1947-7023)

Ernesto Azzurro (orcid:0000-0002-9805-3164)

MARCO LAUTERI (orcid:0000-0003-1071-7999)

Paola Pollegioni (orcid:0000-0001-6388-1931)

Cristina Di Muri (orcid:0000-0003-4072-0662)

Marina Chiappi

giorgio mancinelli (orcid:0000-0002-5833-7322)

Organizations:

National Research Council of Italy (CNR)

Contact:

2. Funding

Funding organizations: National Research Council of Italy (CNR)

Grants: Progetto SAC.AD002.173

Project: PROGETTI DI RICERCA @CNR

3. License

License: CC-BY-4.0

Access Rights: Public

4. Templates

Descriptions

IAS distribution and abundance data

This DMP template describes the (meta)data schemas used to harmonise existing and novel data on IAS gathered and published within the project USEit. Specifically, this template focuses on geo-referenced occurrence and abundance data of marine and terrestrial IAS. All the data gathered are organised in CSV files available for sharing, long-term access and reuse through the LifeWatch Italy Data Portal and the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC.

Template: [Horizon Europe](#)

Type: [Dataset](#)

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

The geo-referenced distribution and abundance data include both existing and novel research data collected during other project activities as well as throughout the USEit activities.

Specifically, these research data include:

Historical occurrence and abundance data of invasive fish species in the Mediterranean Sea gathered for the [ORMEF database](#). Such data are mostly derived from bibliographic data and a large part of them is published in the [SEANOE repository](#) (DOI: 10.17882/84182);

Presence/absence and abundance data of *Anadara* spp. in the Northern Central Adriatic Sea. These data have been extracted from fishery data collected during the [CNR SOLEMON project](#);

Citizen science data of *Halyomorpha halys* in Sardinia collected within two regional projects funded by the region Sardinia for the study and the biological control of the invasive brown marmorated stink bug *H. halys* (“*Accordo di collaborazione tecnico-scientifica per lo studio della cimice asiatica Halyomorpha halys nell’ambito di un intervento finalizzato al potenziamento del Servizio fitosanitario regionale*” and “*Programma Nazionale di lotta biologica per il controllo in Italia di Halyomorpha halys mediante l’impiego del suo antagonista naturale, l’Imenottero Scelionide Trissolcus japonicus*”).

1.1.4 What is the type of the described dataset?

Observational

Research data described in this DMP template are mostly observational (citizen science data and bottom trawl data) but also compiled data derived from literature data mining.

1.1.5 What is its format?

Research data are available as Comma Separated Values (CSV).

1.1.6 What is its expected size?

The overall expected size of all research data with distribution and abundance records is 2 GB. The single dataset size ranges between 80 KB and over 2,000 KB and dataset records vary between 250 to over 5,000.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To keep on record
- To make informed decisions
- To combine with other data

One of the main aims of the project USEit is to overcome the current fragmentation of scientific information on IAS in Italy. Within the project, this is partially achieved by adopting FAIR and Open practices for data management and community-accepted standards. Research data provided by the scientists involved in USEit are shared with the whole scientific community and published using machine-actionable standards facilitating the integration and correct interpretation of research data by both humans and machines. The increased availability of Open and FAIR research data on IAS will benefit scientists and policy-makers who could integrate these data with other scientific evidence to seek more efficient solutions for IAS management. Moreover, the monitoring activities carried out within the project ensure the continuity with other research activities, research projects, and monitoring programmes that collect, directly or indirectly, information on IAS made available through the actions accomplished within USEit.

1.1.8 What is its origin / provenance?

Research data described in this DMP template derive from existing projects including the [ORMEF database](#), the CNR [SOLEMON project](#), and two regional projects funded by the region Sardinia for the study and the biological control of the invasive brown marmorated stink bug *H. halys* (“*Accordo di collaborazione tecnico-scientifica per lo studio della cimice asiatica Halyomorpha halys nell’ambito di un intervento finalizzato al potenziamento del Servizio fitosanitario regionale*” and “*Programma Nazionale di lotta biologica per il controllo in Italia di Halyomorpha halys mediante l’impiego del suo antagonista naturale, l’Imenottero Scelionide Trissolcus japonicus*”).

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers
- Education
- Economy
- The public

The research community studying biological invasion can use these data, integrate them with other evidence and contribute to an improved understanding of this global threat. These digital data published with standard format and languages enriched with semantics also allow data interoperability

and the correct exchange, interpretation and integration of information between machines. Such interoperability supports decision-making by providing the possibility to implement management solutions based on multiple scientific evidence. In addition, the project USEit explores novel management options including the use of IAS as an economic resource under the principles of a sustainable circular economy. Lastly, the collection of citizen science data within the framework of the project ensures the active participation of lays to scientific activities and the communication tasks carried out within one of the work packages of the project facilitate the involvement of the public and a better communication with society of this global issue.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

a. Yes

ORMEF: a Mediterranean database of exotic fish records

Scientific Data

b. Yes

Spatio-temporal dynamics of exotic fish species in the Mediterranean Sea: Over a century of invasion reconstructed

Global Change Biology

c. Yes

The spread of Lessepsian fish does not track native temperature conditions

ICES Journal of Marine Science

d. Yes

Lag times in Lessepsian fish invasion

Biological Invasions

e. Yes

Niche shift can impair the ability to predict invasion risk in the marine realm: an illustration using Mediterranean fish invaders.

Ecology Letters

f. Yes

Salinity, not only temperature, drives tropical fish invasions in the Mediterranean Sea, and surface-only variables explain it better

Coral Reefs

g. Yes

Climate change paves the way for a new inter-ocean fish interchange

Frontiers in Ecology and the Environment

h. Yes

Lessepsian fish invasion in Mediterranean marine protected areas: a risk assessment under climate change scenarios

ICES Journal of Marine Science

i. Yes

Integrating univariate niche dynamics in species distribution models: A step forward for marine research on biological invasions

Journal of Biogeography

j. Yes

Halyomorpha halys and its egg parasitoids Trissolcus japonicus and T. mitsukurii: the geographic dimension of the interaction

NeoBiota

k. Yes

Anadara kagoshimensis (Mollusca: Bivalvia: Arcidae) in Adriatic Sea: morphological analysis, molecular taxonomy, spatial distribution, and prediction

Mediterranean Marine Science

2.1.2 Is there a data availability statement provided along with the publication?

Yes

<https://doi.org/10.17882/84182>

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

DOI

Research data will be uploaded within a dedicated space on the LifeWatch Italy Data Portal and metadata will be harvested by and exposed through the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC. Handles are assigned to the datasets within the Data Portal and DOIs are assigned to the related metadata. Beyond USEit, the dedicated space of USEit within the LifeWatch Italy Data Portal will be maintained and extended to support the publication of all IAS research data produced within other CNR research activities and projects.

Such dedicated space will offer:

1. User-friendly tools for research (meta)data submission and curation;
2. Access to controlled vocabularies for (meta)data annotation;
3. Metrics for the evaluation of (meta)data quality and the FAIR assessment of metadata;
4. Controlled access to further services provided to a restricted group of users defined by the owner(s) of research data.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

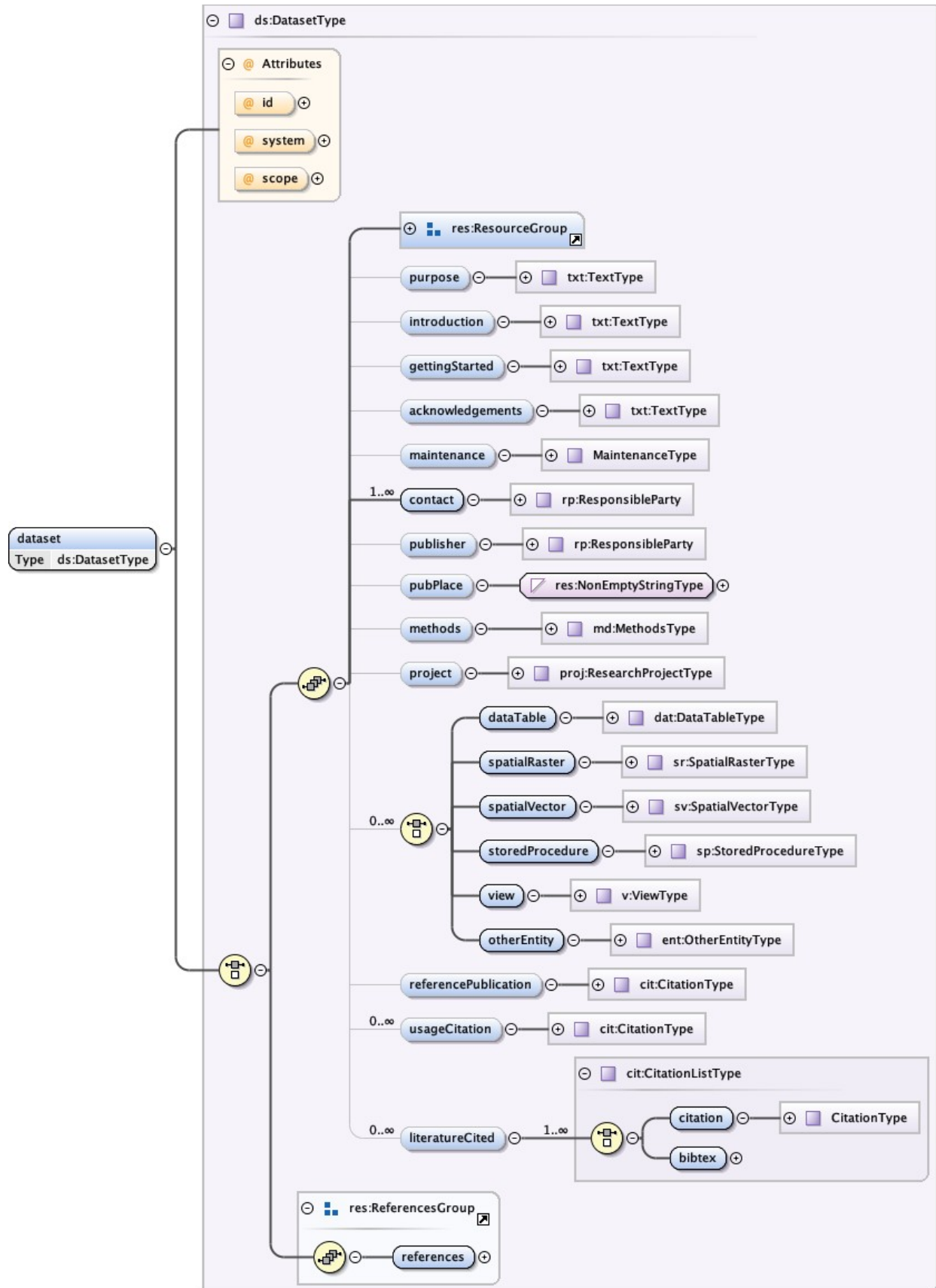


Image 1

Research data will be described using the [EML 2.2.0](#) metadata schema.

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural
- Reference
- Legal

Metadata will be created by using the *eml-dataset* and the *eml-dataTable* modules. The *eml-dataset* module contains general information describing the dataset resources. The module is intended to provide overview information about the dataset: broad information such as the title, abstract, keywords, contacts, maintenance history, purpose, and distribution of the data themselves. The *eml-dataset* module also imports many other modules that are used to describe the dataset in fine detail. Specifically, it uses the *eml-methods* module to describe methodology used in collecting or processing the dataset, the *eml-project* module to describe the overarching research context and experimental design, the *eml-access* module to define access control rules for the data and metadata, and the *eml-entity* module to provide detailed information about the logical structure of the dataset. The *eml-dataset* module is composed of a series of data entities (tables) that are linked together by particular integrity constraints and that are described within the *eml-dataTable* modules.

The *eml-dataTable* module is used to describe the logical characteristics of each tabular set of information in a dataset. A series of comma-separated text files represent the dataset, and each file would subsequently be considered a *dataTable* entity within the dataset. Since the *eml-dataTable* module extends the *eml-entity* module, it uses all of the common entity elements to describe the table, along with a few elements specific to just data table entities. The *eml-dataTable* module allows for the description of each attribute (column/field/variable) within the data table through the use of the *eml-attribute* module. Likewise, there are fields used to describe the physical distribution of the data table, its overall coverage, the methodology used in creating the data, and other logical structure information such as its orientation, case sensitivity, etc.

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

Couldn't find it? Insert it manually

3.1.1.5 Please provide URL/Description of used vocabularies

EML attributes are compatible with or refer to code lists and controlled vocabularies. In addition, some metadata attributes (e.g. keywords) can be compiled using values from controlled vocabularies.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository
- Linked Open Data
- Content Management System

The metadata will be available on the [LifeWatch ERIC](#) metadata catalogue and in the LifeWatch Italy metadata catalogue that will be soon released. Moreover, the metadata will be searchable through the [LifeWatch Italy Semantic Platform](#) by using semantically enriched queries.

3.1.1.8 Are keywords provided in the metadata?

Yes

invasion biology

Keywords are provided in the metadata of each dataset and, wherever possible, such keywords are controlled vocabularies describing terms used in invasion biology such as those of the Darwin Core ([Establishment Means](#), [Degree of Establishment](#), [Pathway](#)) and the [Alien Species Thesaurus](#) along with ecosystem, taxon or domain specific controlled vocabularies.

3.1.1.9 Are metadata harvestable?

Yes

The standard communication protocol for information exchange within the LifeWatch ERIC Metadata Catalogue is the Hypertext Transfer Protocol

Secure (HTTPS) and metadata inside the catalogue are harvestable through the [REST API](#).

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

LifeWatch Italy Data Portal

The new LifeWatch Italy Data Portal will be released shortly and the URL will be added here.

The LifeWatch Italy Data Portal is an online, open-access platform created and managed by LifeWatch Italy, the national node of the LifeWatch ERIC Research Infrastructure for Biodiversity and Ecosystem Research.

The LifeWatch Italy Data Portal supports the acquisition, curation and publication of ecological data in the research domain of biodiversity and ecosystems. Data curation is ensured by the implementation of functionalities supporting taxonomic and syntactic checks. Specifically, the taxonomic check can be performed against national (Italian taxonomic backbone) or global species databases (i.e., World Register of Marine Species [[WoRMS](#)], Catalogue of Life [[COL](#)], and World Flora Online [[WFO](#)]). The syntactic check is ensured through the use of spell checkers and algorithms for numerical, format and URL controls. The data curation workflow includes a final revision and verification carried out by data managers and domain experts who perform the quality control before data publication. Published datasets are automatically assigned with UUIDs and with DOIs can be assigned to metadata upon request.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy

- Supports retention
- Supports withdrawal
- Supports back up
- Provides Open Access content (free at the point of use)
- Assigns PIDs
- Follows metadata standards
- Uses non-proprietary formats
- Supports mid- and long-term preservation
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

The research data described within this DMP template will be available within the LifeWatch Italy Data Portal. The Research Infrastructure of LifeWatch Italy, involved in USEit, will support the project research activities by providing the ICT infrastructure, tools, services and dedicated personnel for the publication of all research data and data products on IAS produced within the project. Moreover, a dedicated space within the LifeWatch Italy Data Portal will be created to host the USEit research data.

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

DOIs are assigned to the research datasets published within the LifeWatch Italy Data Portal into the dedicated space created for the USEit project.

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

ORMEF: Occurrence Records of Mediterranean Exotic Fishes database;
Distribution and abundance of *Anadara* spp. in the Northern-Central Adriatic Sea; Citizen science records of the invasive brown marmorated stink bug *Halyomorpha halys* in Sardinia.

3.2.2.2 How is the dataset / output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Datasets will be preserved and stored in the mid- and long-term within a dedicated space created in the LifeWatch Italy Data Portal. Accessibility is guaranteed through a dedicated, user-friendly web interface that is maintained by LifeWatch Italy.

Users can freely access and download datasets and no authentication/authorisation is required.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

LifeWatch Italy is the national node of the LifeWatch ERIC Research Infrastructure and, as such, the lifetime of the infrastructure supporting the USEit research data is virtually forever.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Other

Metadata will be licenced under the Creative Commons BY 4.0 licence.

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

Metadata include the DOIs and the handles of the described datasets.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Metadata will be stored both in the LifeWatch Italy Data Portal and in the Metadata catalogues of LifeWatch Italy and LifeWatch ERIC. The lifetime of the LifeWatch ERIC infrastructure is virtually forever and, therefore, long-term preservation of metadata stored within the LifeWatch catalogues is virtually forever.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

Couldn't find it? Insert it manually

Below is provided the list of controlled vocabulary used to describe the research data of this DMP template:

1. Darwin Core List of Terms. IRI: <http://rs.tdwg.org/dwc/terms/>
2. Degree of Establishment Controlled Vocabulary List of Terms. URI: <http://rs.tdwg.org/dwcdoe/values/>
3. BODC parameter semantic model biological entity sex category. URI: <https://vocab.nerc.ac.uk/collection/S10/current/>
4. BODC parameter semantic model parameter entity names. URI: <http://vocab.nerc.ac.uk/collection/S06/current/>
5. SeaDataNet Parameter Discovery Vocabulary. URI: <http://vocab.nerc.ac.uk/collection/P02/current/>

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

Couldn't find it? Insert it manually

The [EML](#) (Ecological Metadata Language) and the [Darwin Core](#) are the standards used to describe respectively metadata and data of the described dataset.

3.3.5 What is the methodology followed?

The approach for (meta)data standardisation used by the scientific community working within LifeWatch Italy and also adopted for the research data described here includes the use of a metadata schema based on the EML standard version 2.2.0 and the use of Darwin Core vocabularies and other controlled vocabularies for data standardisation (check sections 3.1.1. and 3.3.1).

3.3.6 What community-endorsed interoperability best practices are followed?

The interoperability approach used within LifeWatch Italy is similar to/compatible with those used by international communities of ecologists and established within other global data repositories such as the Global Biodiversity Information Facility ([GBIF](#)). Specifically, the EML is a metadata specification developed for the ecology discipline and based on prior work done by the Ecological Society of America and associated efforts. The Darwin Core is a body of standards, including a glossary of terms, intended to facilitate the sharing of information about biological diversity and it is sponsored by the community of the Biodiversity Information Standards (TWDG).

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

The full legal code of the CC BY 4.0 licence is available at:
<https://creativecommons.org/licenses/by/4.0/legalcode.en>

3.4.2 What reusability and / or reproducibility methods are followed?

- Data cleaning
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

Data will be stored in the LifeWatch Italy Data Portal which will ensure the long-term accessibility to the data (check section 3.2.1).

3.4.5 Is provenance well documented?

Yes

Metadata will be enriched with provenance information. Specifically, metadata will include a detailed description of the methods, instruments and protocols used for data collection. In the case of derived research data, i.e. data compiled from one or more other 'original' datasets bibliographic resources, a detailed list of the sources is included to give credit to the data creator. Lastly, references to the projects/activities/monitoring programmes in which data were gathered will be added.

3.4.6 What documented procedures for quality assurance do you have in place?

- Use of tools for automatic checks
- Data conform to format specification

Quality assurance procedures include structural and semantic checks. Structural checks consist in the control of file formats and file names. Semantic checks include the use of community-accepted controlled vocabularies for data attributes and data values. In addition, international standards for measurement units (e.g. for variables and parameters) and for numerical values (e.g. for dates or geographical coordinates) are used to replace non-standard units or values. Obsolete values are excluded or replaced with updated information. Such quality control procedures are performed manually before uploading data into the LifeWatch Italy Data Portal.

The LifeWatch Italy Data Portal supports additional quality assurance steps through taxonomic and syntactic controls (described in section 3.2.1). In addition, research data are also reviewed and verified by data managers and domain experts who perform final quality control before data publication.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

- Storage
- Archiving

Direct cost

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Ernesto Azzurro (orcid:0000-0002-9805-3164)

Role(s): Dataset creator; Metadata provider; Contact person

Dataset Title(s): ORMEF: Occurrence Records of Mediterranean Exotic Fishes database

b. Pierluigi Strafella (orcid:0000-0002-1947-7023)

Couldn't find it? Insert it manually

Role(s): Dataset creator; Metadata provider; Contact person

Dataset Title(s): Distribution and abundance of *Anadara* spp. in the Northern-Central Adriatic Sea

c. Ilaria Rosati (orcid:0000-0003-3422-7230)

Couldn't find it? Insert it manually

Role(s): Data manager

Dataset Title(s): ORMEF: Occurrence Records of Mediterranean Exotic Fishes database; Distribution and abundance of *Anadara* spp. in the Northern-Central Adriatic Sea; Citizen science records of the invasive brown marmorated stink bug *Halyomorpha halys* in Sardinia

d. Cristina Di Muri (orcid:0000-0003-4072-0662)

Couldn't find it? Insert it manually

Role(s): Data manager

Dataset Title(s): ORMEF: Occurrence Records of Mediterranean Exotic Fishes database; Distribution and abundance of *Anadara* spp. in the Northern-Central Adriatic Sea; Citizen science records of the invasive brown marmorated stink bug *Halyomorpha halys* in Sardinia

e. LAURA LORU (orcid:0000-0001-6443-9926)

Couldn't find it? Insert it manually

Role(s): Dataset creator; Metadata provider; Contact person

Dataset Title(s): Citizen science records of the invasive brown marmorated stink bug *Halyomorpha halys* in Sardinia

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

IAS Isotope Ratio Mass Spectrometry (IRMS) data

This DMP template describes the (meta)data schemas used to harmonise stable isotope data collected and published within the project USEit. Specifically, this template focuses on geo-referenced individual stable isotope data ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) of marine and terrestrial IAS. All the data gathered are organised in CSV files available for sharing, long-term access and reuse through the LifeWatch Italy Data Portal and the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC.

Template: [Horizon Europe](#)

Type: [Dataset](#)

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

The geo-referenced stable isotope data include:

1. for marine IAS, a stable isotopes dataset with $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values of marine IAS' tissues (*Anadara kagoshimensis*, *A. transversa*, *Siganus luridus*, *S. rivulatus*, *Pterois miles*) and of those of their native competitors (*Salpa salpa* and *Sparisoma cretense*). Values of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ of prey species' tissues are also included in the dataset and used as baselines for the estimation of trophic niches;

2. for terrestrial IAS, a dataset of C and N stable isotope compositions in *Ailanthus altissima* leaves and nearby soil used to investigate the symbiotic

relationships between the invasive species and the associated fungi (different types of mycorrhizae) or bacteria belonging to the soil microbiota.

1.1.4 What is the type of the described dataset?

Experimental

Research data described in this DMP template are experimental data derived from specific sampling campaigns (either marine or terrestrial activities) and subsequent isotope ratio mass spectrometer spectrometry (IRMS) analyses. Specimen morphometric data (length and weight) are specifically included in the dataset of marine species used for IRMS analysis.

1.1.5 What is its format?

Research data are available as Comma Separated Values (CSV).

1.1.6 What is its expected size?

The overall expected size of stable isotope research data is 20 KB.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To keep on record
- To make informed decisions
- To combine with other data

One of the main aims of the project USEit is to overcome the current fragmentation of scientific information on IAS in Italy. Within the project, this is partially achieved by adopting FAIR and Open practices for data management and community-accepted standards. Research data provided by the scientists involved in USEit are shared with the whole scientific community and published using machine-actionable standards facilitating the integration and correct interpretation of research data by both humans and machines. The increased availability of Open and FAIR research data on IAS will benefit scientists and

policy-makers who could integrate these data with other scientific evidence to seek more efficient solutions for IAS management. Moreover, the monitoring activities carried out within the project ensure the continuity with other research activities, research projects, and monitoring programmes that collect, directly or indirectly, information on IAS made available through the actions accomplished within USEit.

1.1.8 What is its origin / provenance?

Research data described in this DMP template were gathered within the USEit activities and analysed using IRMS techniques. As previously mentioned, sampling procedures are widely different depending on the monitoring or experimental action carried on. Furthermore, sample collections might result largely variable due to environmental specificities (marine, freshwaters or terrestrial environments) affecting the technical execution. Once collected, specimens are usually dried and finely grounded. Small subsamples (usually 0.2 – 1 mg, depending on their C and/or N content) are then put in small tinny capsules and quantitatively burned in an elemental analyser, before admitting the produced CO₂ or N₂ peaks into an IRMS analyser for the ¹³C/¹²C or the ¹⁵N/¹⁴N isotope ratio determination.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers

The research data described within this DMP template determine an improved understanding of the trophic habits of marine IAS and of possible mutualistic relationships increasing the competition ability of plant terrestrial IAS.

Stable isotope analysis of marine IAS and of their native competitors can be used for trophic position and trophic niche estimations. Such analysis can deepen the understanding of the trophic ecology of IAS in invaded habitats and support the evaluation of the ecological impact on recipient marine food webs.

Stable isotope analysis of terrestrial IAS can be used to understand and/or to estimate the adaptedness of invasive species in terrestrial habitats by investigating variation of mutualistic relationships with associated mycorrhizal and/or bacteria, based on resource availability changes.

The research community studying biological invasion can use these data, integrate them with other evidence and contribute to an improved understanding of this global threat. These digital data published with standard format and languages enriched with semantics also allow data interoperability and the correct exchange, interpretation and integration of information between machines. Such interoperability supports decision-making by providing the possibility to implement management solutions based on multiple scientific evidence.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

DOI

Research data will be uploaded within a dedicated space on the LifeWatch Italy Data Portal and metadata will be harvested by and exposed through the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC. Handles (UUIDs) are assigned to the datasets within the Data Portal and DOIs are assigned to the related metadata. Beyond USEit, the dedicated space of USEit within the LifeWatch Italy Data Portal will be maintained and extended to support the publication of all IAS research data produced within other CNR research activities and projects.

Such dedicated space will offer:

1. User-friendly tools for research (meta)data submission and curation;
2. Access to controlled vocabularies for (meta)data annotation;
3. Metrics for the evaluation of (meta)data quality and the FAIR assessment of metadata;
4. Controlled access to further services provided to a restricted group of users defined by the owner(s) of research data.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

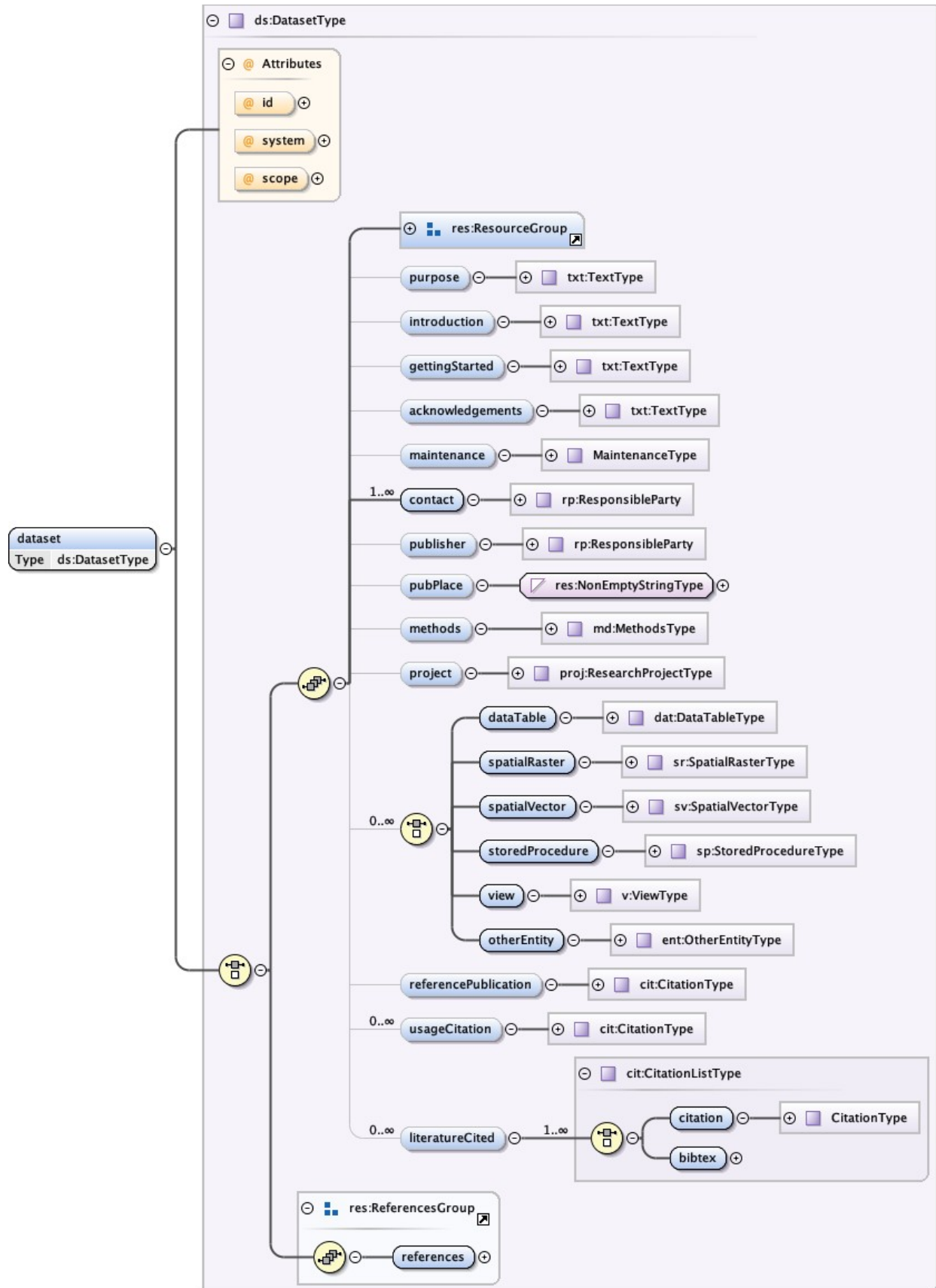


Image 2

Research data will be described using the [EML 2.2.0](#) metadata schema.

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural
- Reference
- Legal

Metadata will be created by using the *eml-dataset* and the *eml-dataTable* modules. The *eml-dataset* module contains general information describing the dataset resources. The module is intended to provide overview information about the dataset: broad information such as the title, abstract, keywords, contacts, maintenance history, purpose, and distribution of the data themselves. The *eml-dataset* module also imports many other modules that are used to describe the dataset in fine detail. Specifically, it uses the *eml-methods* module to describe methodology used in collecting or processing the dataset, the *eml-project* module to describe the overarching research context and experimental design, the *eml-access* module to define access control rules for the data and metadata, and the *eml-entity* module to provide detailed information about the logical structure of the dataset. The *eml-dataset* module is composed of a series of data entities (tables) that are linked together by particular integrity constraints and that are described within the *eml-dataTable* modules.

The *eml-dataTable* module is used to describe the logical characteristics of each tabular set of information in a dataset. A series of comma-separated text files represent the dataset, and each file would subsequently be considered a *dataTable* entity within the dataset. Since the *eml-dataTable* module extends the *eml-entity* module, it uses all of the common entity elements to describe the table, along with a few elements specific to just data table entities. The *eml-dataTable* module allows for the description of each attribute (column/field/variable) within the data table through the use of the *eml-attribute* module. Likewise, there are fields used to describe the physical distribution of the data table, its overall coverage, the methodology used in creating the data, and other logical structure information such as its orientation, case sensitivity, etc.

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

Couldn't find it? Insert it manually

3.1.1.5 Please provide URL/Description of used vocabularies

EML attributes are compatible with or refer to code lists and controlled vocabularies. In addition, some metadata attributes (e.g. keywords) can be compiled using values from controlled vocabularies.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository
- Linked Open Data
- Content Management System

The metadata will be available on the [LifeWatch ERIC](#) metadata catalogue and in the LifeWatch Italy metadata catalogue that will be soon released. Moreover, the metadata will be searchable through the [LifeWatch Italy Semantic Platform](#) by using semantically enriched queries.

3.1.1.8 Are keywords provided in the metadata?

Yes

invasion biology

Keywords are provided in the metadata of each dataset and, wherever possible, such keywords are controlled vocabularies describing terms used in invasion biology such as those of the Darwin Core ([Establishment Means](#), [Degree of Establishment](#), [Pathway](#)) and the [Alien Species Thesaurus](#) along with ecosystem, taxon or domain specific controlled vocabularies.

3.1.1.9 Are metadata harvestable?

Yes

The standard communication protocol for information exchange within the LifeWatch ERIC Metadata Catalogue is the Hypertext Transfer Protocol

Secure (HTTPS) and metadata inside the catalogue are harvestable through the [REST API](#).

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

LifeWatch Italy Data Portal

The new LifeWatch Italy Data Portal will be released shortly and the URL will be added here.

The LifeWatch Italy Data Portal is an online, open-access platform created and managed by LifeWatch Italy, the national node of the LifeWatch ERIC Research Infrastructure for Biodiversity and Ecosystem Research.

The LifeWatch Italy Data Portal supports the acquisition, curation and publication of ecological data in the research domain of biodiversity and ecosystems. Data curation is ensured by the implementation of functionalities supporting taxonomic and syntactic checks. Specifically, the taxonomic check can be performed against national (Italian taxonomic backbone) or global species databases (i.e., World Register of Marine Species [[WoRMS](#)], Catalogue of Life [[COL](#)], and World Flora Online [[WFO](#)]). The syntactic check is ensured through the use of spell checkers and algorithms for numerical, format and URL controls. The data curation workflow includes a final revision and verification carried out by data managers and domain experts who perform the quality control before data publication. Published datasets are automatically assigned with UUIDs and with DOIs can be assigned to metadata upon request.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy

- Supports retention
- Supports withdrawal
- Supports back up
- Provides Open Access content (free at the point of use)
- Assigns PIDs
- Follows metadata standards
- Uses non-proprietary formats
- Supports mid- and long-term preservation
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

The research data described within this DMP template will be available within the LifeWatch Italy Data Portal. The Research Infrastructure of LifeWatch Italy, involved in USEit, will support the project research activities by providing the ICT infrastructure, tools, services and dedicated personnel for the publication of all research data and data products on IAS produced within the project. Moreover, a dedicated space within the LifeWatch Italy Data Portal will be created to host the USEit research data.

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

DOIs are assigned to the research datasets published within the LifeWatch Italy Data Portal into the dedicated space created for the USEit project.

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Dataset titles will be identified upon publication.

3.2.2.2 How is the dataset / output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Datasets will be preserved and stored in the mid- and long-term within a dedicated space created in the LifeWatch Italy Data Portal. Accessibility is guaranteed through a dedicated, user-friendly web interface that is maintained by LifeWatch Italy.

Users can freely access and download datasets and no authentication/authorisation is required.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

LifeWatch Italy is the national node of the LifeWatch ERIC Research Infrastructure and, as such, the lifetime of the infrastructure supporting the USEit research data is virtually forever.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Other

Metadata will be licenced under the Creative Commons BY 4.0 licence.

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

Metadata include the DOIs and the handles of the described datasets.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Metadata will be stored both in the LifeWatch Italy Data Portal and in the Metadata catalogues of LifeWatch Italy and LifeWatch ERIC. The lifetime of the LifeWatch ERIC infrastructure is virtually forever and, therefore, long-term preservation of metadata stored within the LifeWatch catalogues is virtually forever.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

Couldn't find it? Insert it manually

Below is provided the list of controlled vocabulary used to describe the research data of this DMP template:

1. Darwin Core List of Terms. IRI: <http://rs.tdwg.org/dwc/terms/>
2. BODC Parameter Usage Vocabulary. URI: <http://vocab.nerc.ac.uk/collection/P01/current/>
3. Environmental Thesaurus (EnvThes). DOI: <https://doi.org/10.48373/0PWD-C575>

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

Couldn't find it? Insert it manually

The [EML](#) (Ecological Metadata Language) and the [Darwin Core](#) are the standards used to describe respectively metadata and data of the described dataset.

3.3.5 What is the methodology followed?

The approach for (meta)data standardisation used by the scientific community working within LifeWatch Italy and also adopted for the research data described here includes the use of a metadata schema based on the EML standard version 2.2.0 and the use of Darwin Core vocabularies and other controlled vocabularies for data standardisation (check sections 3.1.1. and 3.3.1).

3.3.6 What community-endorsed interoperability best practices are followed?

The interoperability approach used within LifeWatch Italy is similar to/compatible with those used by international communities of ecologists and established within other global data repositories such as the Global Biodiversity Information Facility ([GBIF](#)). Specifically, the EML is a metadata specification developed for the ecology discipline and based on prior work done by the Ecological Society of America and associated efforts. The Darwin Core is a body of standards, including a glossary of terms, intended to facilitate the sharing of information about biological diversity and it is sponsored by the community of the Biodiversity Information Standards (TWDG).

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

The full legal code of the CC BY 4.0 licence is available at: <https://creativecommons.org/licenses/by/4.0/legalcode.en>

3.4.2 What reusability and / or reproducibility methods are followed?

- Data cleaning
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

Data will be stored in the LifeWatch Italy Data Portal which will ensure the long-term accessibility to the data (check section 3.2.1).

3.4.5 Is provenance well documented?

Yes

Metadata will be enriched with provenance information. Specifically, metadata will include a detailed description of the methods, instruments and protocols used for data collection.

3.4.6 What documented procedures for quality assurance do you have in place?

- Use of tools for automatic checks
- Data conform to format specification

Quality assurance procedures include structural and semantic checks. Structural checks consist in the control of file formats and file names. Semantic checks include the use of community-accepted controlled vocabularies for data attributes and data values. In addition, international standards for measurement units (e.g. for variables and parameters) and for numerical values (e.g. for dates or geographical coordinates) are used to replace non-standard units or values. Obsolete values are excluded or replaced with updated information. Such quality control procedures are performed manually before uploading data into the LifeWatch Italy Data Portal.

The LifeWatch Italy Data Portal supports additional quality assurance steps through taxonomic and syntactic controls (described in section 3.2.1). In addition, research data are also reviewed and verified by data managers and domain experts who perform final quality control before data publication.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

- Storage
- Archiving

Direct cost

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. MARCO LAUTERI (orcid:0000-0003-1071-7999)

Role(s): Dataset creator; Metadata provider; Contact person

b. Pierluigi Strafella (orcid:0000-0002-1947-7023)

Couldn't find it? Insert it manually

Role(s): Dataset creator; Metadata provider; Contact person

c. Ilaria Rosati (orcid:0000-0003-3422-7230)

Couldn't find it? Insert it manually

Role(s): Data manager

d. Cristina Di Muri (orcid:0000-0003-4072-0662)

Couldn't find it? Insert it manually

Role(s): Data manager

e. MARCO CIOLFI (orcid:0000-0003-4831-8053)

Couldn't find it? Insert it manually

Role(s): Dataset creator; Metadata provider; Contact person

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

IAS telemetry data

This DMP template describes the (meta)data schemas used to publish IAS telemetry data gathered within the project USEit. Specifically, this template describes the data management practices used for telemetry data of *Callinectes sapidus* gathered through the use of accelerometers and ultrasonic telemetry to study the space-use behavior of the species in mesocosms and in natural invaded transitional habitats. Data will be available for sharing, long-term access and reuse through the European Tracking Network (ETN) Data Portal and metadata will be also accessible through the LifeWatch Italy and LifeWatch ERIC Metadata Catalogues.

Template: [Horizon Europe](#)

Type: [Dataset](#)

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

This template describes the data management practices used for telemetry data of *C. sapidus* generated within the research activities of the USEit project.

Research data include mesocosms experimental data and *in-situ* observational data. Mesocosms experiments are performed using accelerometers to study the feeding habits, the movements, the molting, the burying behaviour and the reproductive activity of the species. *In-situ* data are collected through the use of ultrasonic telemetry to monitor changes in space-use behavior of *C. sapidus* in relation to temperature, reproductive activity, and environmental conditions (meteorological data).

1.1.4 What is the type of the described dataset?

Observational

Research data described in this DMP template include both experimental data from mesocosms and observational data captured *in-situ*.

1.1.5 What is its format?

Research data will be available as VPS files generated by the [INNOVASEA](#) receivers used for data collection.

1.1.6 What is its expected size?

The overall expected size of research data is XX

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To combine with other data

Telemetry data collected within the research activities of USEit are used to obtain and share novel scientific information about the space-use behaviour of *C. sapidus* in invaded transitional habitats. The adoption of community-accepted standards for the data management will also allow the reuse and integration of the collected data in line with the FAIR principles and with the scope of the USEit

project that is to overcome the current fragmentation of scientific information on IAS in Italy.

1.1.8 What is its origin / provenance?

Research data described in this DMP template are gathered within the activities of the USEit project by the researchers of the CNR-IAS Capo Granitola (Trapani, Italy) and CNR-IAS Palermo (Palermo, Italy). Mesocosms experiments are carried in the laboratory facilities of the CNR-IAS in Capo Granitola whereas in-situ data are collected in the S'Ena Arrubia lagoon (Oristano, Sardinia). Instruments used for mesocosms experiments (accelerometers) and for *in-situ* observations (telemetry tags and receivers) are provided by INNOVASEA.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers

The research community studying biological invasion can use these data, integrate them with other evidence and contribute to an improved understanding of space-use behaviours of IAS. These research data can be used by the researchers of the European Tracking Network whose mission is to track aquatic animals across Europe to better understand, protect and manage them.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

URL

Research data will be uploaded within the [European Tracking Network \(ETN\) Data Portal](#) and metadata will be exposed through the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC. The ETN data management platform is the central data portal of ETN. It gives access to all types of telemetry data and provides an interface to manage, explore and download the data and metadata.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

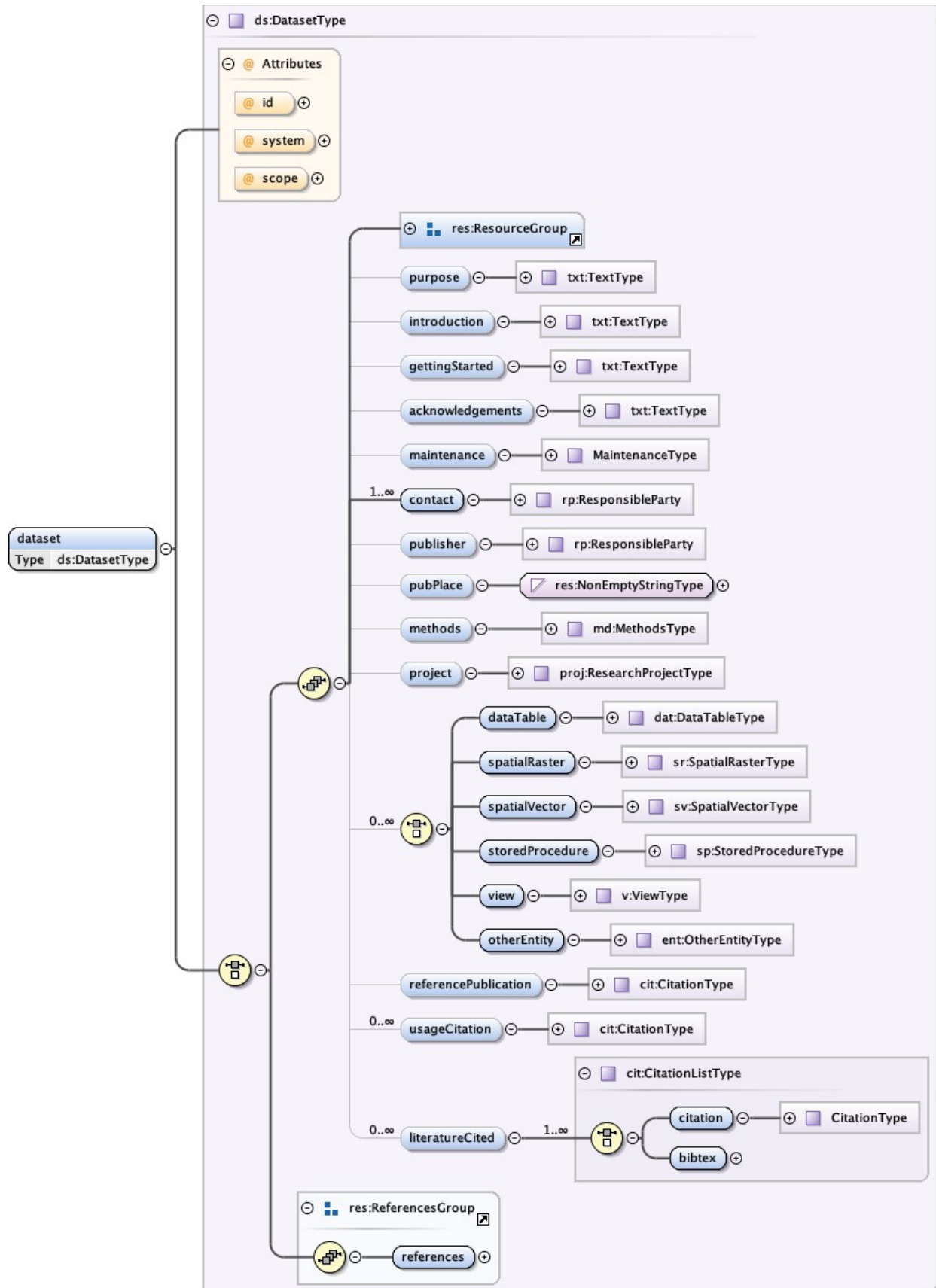


Image 3

Metadata available through the LifeWatch Italy and LifeWatch ERIC Metadata Catalogues will be described using the [EML 2.2.0](#) metadata schema.

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural
- Reference
- Legal

Metadata will be created by using the *eml-dataset* module. The *eml-dataset* module contains general information describing the dataset resources. The module is intended to provide overview information about the dataset: broad information such as the title, abstract, keywords, contacts, maintenance history, purpose, and distribution of the data themselves. The *eml-dataset* module also imports many other modules that are used to describe the dataset in fine detail. Specifically, it uses the *eml-methods* module to describe methodology used in collecting or processing the dataset, the *eml-project* module to describe the overarching research context and experimental design, the *eml-access* module to define access control rules for the data and metadata, and the *eml-entity* module to provide detailed information about the logical structure of the dataset.

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository
- Linked Open Data
- Content Management System

Metadata will be findable through the [LifeWatch ERIC](#) metadata catalogue and in the LifeWatch Italy metadata catalogue that will be soon released. Moreover, the metadata will be searchable through the [LifeWatch Italy Semantic Platform](#) by using semantically enriched queries.

3.1.1.8 Are keywords provided in the metadata?

Yes

invasion biology

Keywords are provided in the metadata of each dataset and, wherever possible, such keywords are controlled vocabularies describing terms used in invasion biology such as those of the Darwin Core ([Establishment Means](#), [Degree of Establishment](#), [Pathway](#)) and the [Alien Species Thesaurus](#) along with ecosystem, taxon or domain specific controlled vocabularies.

3.1.1.9 Are metadata harvestable?

Yes

The standard communication protocol for information exchange within the LifeWatch ERIC Metadata Catalogue is the Hypertext Transfer Protocol Secure (HTTPS) and metadata inside the catalogue are harvestable through the [REST API](#).

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

ETN Data Portal

<https://www.lifewatch.be/etn/>

The European Tracking Network (ETN) data management platform is the central data portal of ETN. It gives access to all types of aquatic animal telemetry data and provides an interface to manage, explore and download the metadata and data. Access is password protected and data moratorium rules are in place.

The ETN data portal has been developed by the Flanders Marine Institute (VLIZ) and is open to all European telemetry users. New developments and additional features are added on a continuous basis. The web application is built using PHP (using the Symfony framework) for the back-end side and Bootstrap/jQuery/datatables/... (among others) to facilitate the development of the front-end side. Raw data is uploaded to the ETN data portal and automatically stored in the [Marine Data Archive](#) (MDA). Project information is

stored in the Integrated Marine Information System (IMIS). In ETN the metadata and data can be organised and downloaded.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy
- Supports retention
- Supports withdrawal
- Supports back up
- Provides Open Access content (free at the point of use)
- Assigns PIDs
- Follows metadata standards
- Uses non-proprietary formats
- Supports mid- and long-term preservation
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

The ETN data portal is open to all European telemetry users. In addition, the involvement of the LifeWatch Italy research infrastructure in USEit will guarantee a continuous support to the project's research activities. Specifically, LifeWatch Italy will provide the ICT infrastructure, tools, services and dedicated personnel for the FAIR management of research data and data products on IAS produced within the project. Specifically, raw telemetry data collected within USEit and available in the ETN data portal will be described with rich and FAIR metadata available through the LifeWatch Italy and ERIC Metadata Catalogues.

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

URLs are assigned to the research datasets published within the ETN Data Portal.

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Dataset titles will be identified upon publication.

3.2.2.2 How is the dataset / output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset / output?

Yes

3.2.2.6 Please provide information about the method(s) needed to access the dataset / output.

Authentication to the ETN data portal is required for data access.

3.2.2.7 Please provide information about the tools needed to access the dataset / output.

Access is password protected and data moratorium rules are in place. Access to the data portal is restricted to registered members only and users can register at <http://www.lifewatch.be/etn/login>.

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Datasets will be available through the ETN Data Portal and they will be stored in the mid- and long-term within the Marine Data Archive maintained by LifeWatch Belgium.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

LifeWatch Belgium is a national node of the LifeWatch ERIC Research Infrastructure and, as such, the lifetime of the infrastructure supporting these research data is virtually forever.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Other

Metadata will be licenced under the Creative Commons BY 4.0 licence.

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

Metadata will include the URLs of the described datasets.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

The lifetime of the LifeWatch ERIC infrastructure is virtually forever and, therefore, long-term preservation of metadata stored within the LifeWatch catalogues is virtually forever.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

The full legal code of the CC BY 4.0 licence is available at:
<https://creativecommons.org/licenses/by/4.0/legalcode.en>

3.4.2 What reusability and / or reproducibility methods are followed?

- Data cleaning
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

Data will be accessible through the ETN Data Portal and will be stored in the Marine Data Archive managed by research infrastructure of LifeWatch Belgium which will ensure the long-term accessibility to the data (check section 3.2.1).

3.4.5 Is provenance well documented?

Yes

Metadata will be enriched with provenance information. Specifically, metadata will include a detailed description of the methods, instruments and protocols used for data collection.

3.4.6 What documented procedures for quality assurance do you have in place?

- Use of tools for automatic checks
- Data conform to format specification

The ETN data portal include several quality controls in place on different levels that help to improve data quality and avoid human errors. The available Quality Controls (QCs) are:

1. QC on Detections: searching for detections falling before animal release date; searching for detections falling before deployment date; searching for detections falling after deployment recovery date.
2. QC on Deployments: no overlapping deployments of a same receiver allowed.
3. QC on Tag Reuse: searching for tag reuse but no animal recapture information.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

- Storage
- Archiving

Direct cost

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Carlo Pipitone (orcid:0000-0002-7632-1228)

Role(s): Dataset creator; Metadata provider; Contact person

b. VINCENZOMAXIMILIANO GIACALONE (orcid:0000-0002-4316-1723)

Couldn't find it? Insert it manually

Role(s): Dataset creator; Metadata provider; Contact person

c. Ilaria Rosati (orcid:0000-0003-3422-7230)

Couldn't find it? Insert it manually

Role(s): Data manager

d. Cristina Di Muri (orcid:0000-0003-4072-0662)

Couldn't find it? Insert it manually

Role(s): Data manager

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

DNA metabarcoding data

This DMP template describes the (meta)data management practices used for DNA metabarcoding data gathered within the project USEit. Genetic sequences will be uploaded in NCBI and such data will also be available as DNA-derived biodiversity information organised in CSV files available for sharing, long-term access and reuse through the LifeWatch Italy Data Portal and the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC.

Template: [Horizon Europe](#)

Type: [Dataset](#)

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

DNA metabarcoding data will be generated within the research activities of USEit to characterise the bacterial and fungal communities associated to the roots of the invasive tree *Ailanthus altissima*. Such data will be used to study the competitive advantages deriving from mycorrhizae associated to the invasive species.

1.1.4 What is the type of the described dataset?

Experimental

Research data described in this DMP template are experimental data generated from DNA sequencing using metabarcoding analysis.

1.1.5 What is its format?

DNA sequences (raw data) will be available as fastq files in the Sequence Read Archive (SRA). DNA-derived biodiversity data will be available as Comma

Separated Values (CSV) files in the LifeWatch Italy Data Portal. CSV files will also include the SRA accession numbers to the original sequences available in the SRA portal of NCBI.

1.1.6 What is its expected size?

The expected size of DNA sequences submitted in the SRA is around 20 GB. The expected size of CSV files with DNA-derived biodiversity data uploaded in the LifeWatch Italy Data Portal is about 2 MB.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To combine with other data

DNA metabarcoding data collected within the research activities of USEit are used to obtain and share novel scientific information about bacterial and fungal communities associated with the roots of the invasive tree *Ailanthus altissima*. These data will be used to study the competitive advantages of the invasive species based on the characterization of mycorrhizae associated to its root apparatus. The adoption of community-accepted standards for the data management will allow the reuse and integration of the collected data in line with the FAIR principles and with the scope of the USEit project that is to overcome the current fragmentation of scientific information on IAS in Italy. In addition, the publication of DNA-derived biodiversity data along with DNA sequences will enhance the discoverability and reusability of these data. In fact, sequences information associated with geographical coordinates and timestamp represent valuable biodiversity occurrence that can be used by ecologists with no molecular background and reused in the context of broader biodiversity analysis.

1.1.8 What is its origin / provenance?

Research data described in this DMP template are gathered within the activities of the USEit project by the researchers of the CNR-IRET of Porano (Terni, Italy). Sample collection and initial sample processing is performed by the CNR-IRET Porano whereas metabarcoding analysis are carried out at Sequentia Biotech (Barcelona, Spain; <https://www.sequentiabitech.com/>). Detailed information about facilities, instruments and protocols used for DNA metabarcoding analysis will be described in the research (meta)data.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers

The DNA metabarcoding data collected through the USEit research activities will be shared as DNA sequences within the SRA portal and as biodiversity data within the LifeWatch Italy Data Portal. This management practice will enhance the reusability of these data that will be useful for molecular ecologists as well as for the broader community of researchers working with biodiversity and occurrence data.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

DOI

DNA-derived occurrence data will be uploaded within a dedicated space on the LifeWatch Italy Data Portal and metadata will be harvested by and exposed through the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC. Handles (UUIDs) are assigned to the datasets within the Data Portal and DOIs are assigned to the related metadata.

The LifeWatch Italy Data Portal will offer:

1. User-friendly tools for (meta)data submission and curation;
2. Access to controlled vocabularies for (meta)data annotation;
3. Metrics for the evaluation of (meta)data quality and the FAIR assessment of metadata;
4. Controlled access to further services provided to a restricted group of users defined by the owner(s) of research data.

In addition, sequencing data will be submitted in the SRA that will provide identifiers to: i. project (BioProject accession number); ii. samples (BioSample accession numbers); and iii. sequences (SRA accession numbers).

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

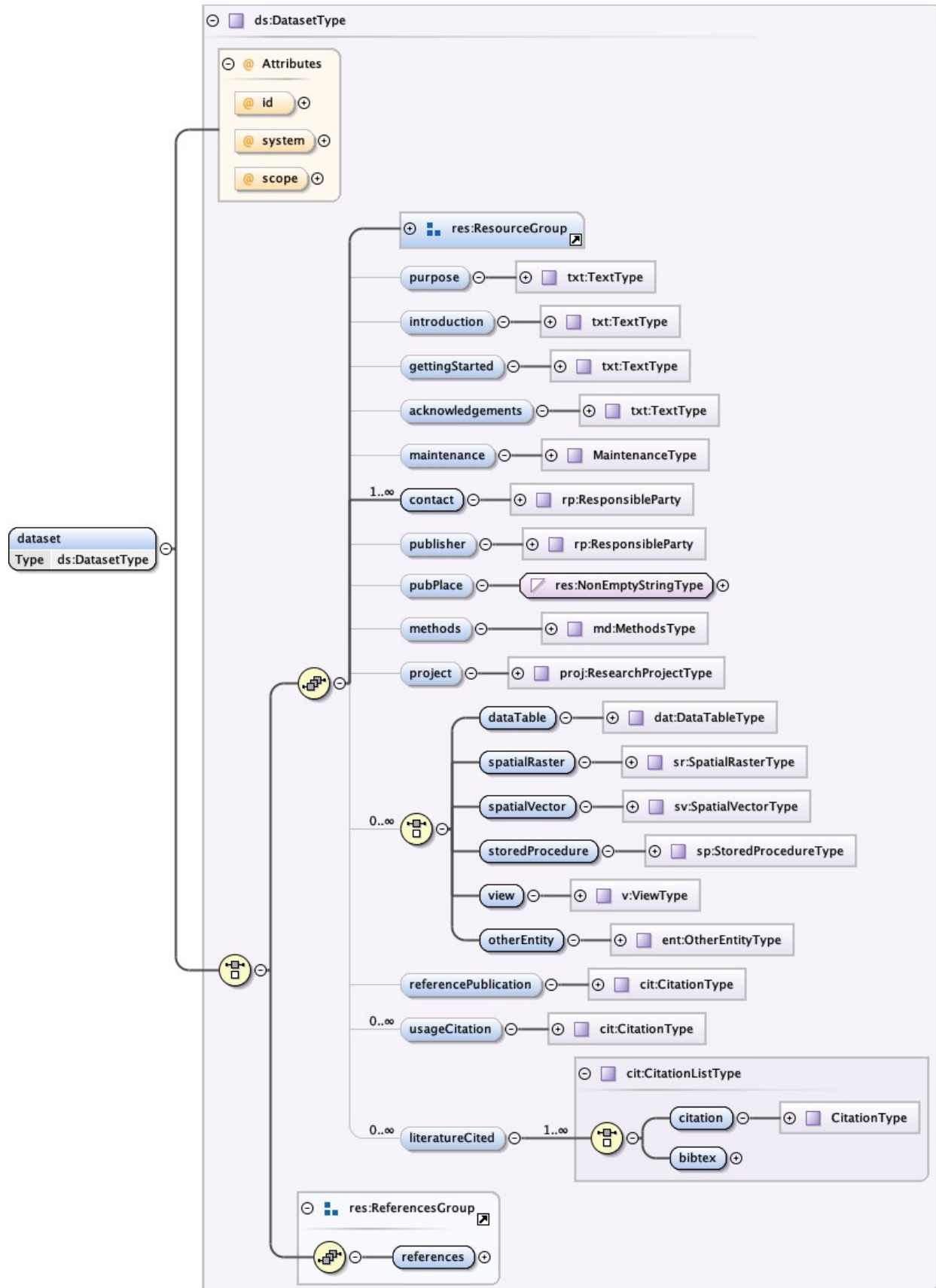


Image 4

DNA-derived occurrence data will be described using the [EML 2.2.0](#) metadata schema.

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural
- Reference
- Legal

Metadata will be created by using the *eml-dataset* and the *eml-dataTable* modules. The *eml-dataset* module contains general information describing the dataset resources. The module is intended to provide overview information about the dataset: broad information such as the title, abstract, keywords, contacts, maintenance history, purpose, and distribution of the data themselves. The *eml-dataset* module also imports many other modules that are used to describe the dataset in fine detail. Specifically, it uses the *eml-methods* module to describe methodology used in collecting or processing the dataset, the *eml-project* module to describe the overarching research context and experimental design, the *eml-access* module to define access control rules for the data and metadata, and the *eml-entity* module to provide detailed information about the logical structure of the dataset. The *eml-dataset* module is composed of a series of data entities (tables) that are linked together by particular integrity constraints and that are described within the *eml-dataTable* modules.

The *eml-dataTable* module is used to describe the logical characteristics of each tabular set of information in a dataset. A series of comma-separated text files represent the dataset, and each file would subsequently be considered a *dataTable* entity within the dataset. Since the *eml-dataTable* module extends the *eml-entity* module, it uses all of the common entity elements to describe the table, along with a few elements specific to just data table entities. The *eml-dataTable* module allows for the description of each attribute (column/field/variable) within the data table through the use of the *eml-attribute* module. Likewise, there are fields used to describe the physical distribution of the data table, its overall coverage, the methodology used in creating the data, and other logical structure information such as its orientation, case sensitivity, etc.

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

Couldn't find it? Insert it manually

3.1.1.5 Please provide URL/Description of used vocabularies

EML attributes are compatible with or refer to code lists and controlled vocabularies. In addition, some metadata attributes (e.g. keywords) can be compiled using values from controlled vocabularies.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository
- Linked Open Data
- Content Management System

The metadata will be available on the [LifeWatch ERIC](#) metadata catalogue and in the LifeWatch Italy metadata catalogue that will be soon released. Moreover, the metadata will be searchable through the [LifeWatch Italy Semantic Platform](#) by using semantically enriched queries.

3.1.1.8 Are keywords provided in the metadata?

Yes

invasion biology

Keywords are provided in the metadata of each dataset and, wherever possible, such keywords are controlled vocabularies describing terms used in invasion biology such as those of the Darwin Core ([Establishment Means](#), [Degree of Establishment](#), [Pathway](#)) and the [Alien Species Thesaurus](#) along with ecosystem, taxon or domain specific controlled vocabularies.

3.1.1.9 Are metadata harvestable?

Yes

The standard communication protocol for information exchange within the LifeWatch ERIC Metadata Catalogue is the Hypertext Transfer Protocol Secure (HTTPS) and metadata inside the catalogue are harvestable through the [REST API](#).

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

Sequence Read Archive

LifeWatch Italy Data Portal

Raw sequencing data (fastq files) will be uploaded in SRA at:
<https://submit.ncbi.nlm.nih.gov/subs/sra/>

DNA-derived occurrence data will be available within the new LifeWatch Italy Data Portal that is going to be released shortly and its URL will be added here.

The SRA is a publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis. The SRA is NIH's archive of high-throughput sequencing data and is part of the International Nucleotide Sequence Database Collaboration (INSDC) that includes the NCBI Sequence Read Archive (SRA), the European Bioinformatics Institute (EBI), and the DNA Database of Japan (DDBJ). Data submitted to any of the three organisations are shared among them.

The LifeWatch Italy Data Portal is an online, open-access platform created and managed by LifeWatch Italy, the national node of the LifeWatch ERIC Research Infrastructure for Biodiversity and Ecosystem Research.

The LifeWatch Italy Data Portal supports the acquisition, curation and publication of ecological data in the research domain of biodiversity and ecosystems. Data curation is ensured by the implementation of functionalities supporting taxonomic and syntactic checks. Specifically, the taxonomic check can be performed against national (Italian taxonomic backbone) or global species databases (i.e., World Register of Marine Species [[WoRMS](#)], Catalogue of Life [[COL](#)], and World Flora Online [[WFO](#)]). The syntactic check is ensured through the use of spell checkers and algorithms for numerical, format and URL controls. The data curation workflow includes a final revision and verification carried out by data managers and domain experts who perform the quality control before data publication. Published datasets are automatically assigned with UUIDs and with DOIs can be assigned to metadata upon request.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy
- Supports retention
- Supports withdrawal
- Supports back up

- Provides Open Access content (free at the point of use)
- Assigns PIDs
- Follows metadata standards
- Uses non-proprietary formats
- Supports mid- and long-term preservation
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

The SRA is a publicly available repository of high throughput sequencing data and registered users can freely store raw sequencing data and alignment information to enhance reproducibility of their research studies.

The research data described within this DMP template will be also available within the LifeWatch Italy Data Portal. The Research Infrastructure of LifeWatch Italy, involved in USEit, will support the project research activities by providing the ICT infrastructure, tools, services and dedicated personnel for the publication of all research data and data products on IAS produced within the project. Moreover, a dedicated space within the LifeWatch Italy Data Portal will be created to host the USEit research data.

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

DOIs are assigned to the DNA-derived occurrence datasets published within the LifeWatch Italy Data Portal into the dedicated space created for the USEit project. Moreover, raw sequencing data submitted in the SRA will be assigned with identifiers for: i. project (BioProject accession number); ii.

samples (BioSample accession numbers); and iii. sequences (SRA accession numbers).

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Dataset title will be identified upon publication.

3.2.2.2 How is the dataset / output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Raw sequencing data will be stored and available for long-term reuse in the SRA.

DNA-derived occurrence datasets will be preserved and stored in the mid- and long-term within a dedicated space created in the LifeWatch Italy Data Portal. Accessibility is guaranteed through a dedicated, user-friendly web interface that is maintained by LifeWatch Italy.

Users can freely access and download both raw sequencing data and DNA-derived occurrence datasets and no authentication/authorisation is required.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

The SRA is a NIH's archive and raw sequencing data are stored for long-term preservation and maintenance within this archive and they will be also accessible through the European Bioinformatics Institute (EBI), and the DNA Database of Japan (DDBJ).

Occurrence datasets from DNA analysis will be stored for mid- and long-term preservation within the LifeWatch Italy Data Portal. LifeWatch Italy is the national node of the LifeWatch ERIC Research Infrastructure and, as such, the lifetime of the infrastructure supporting the USEit research data is virtually forever.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Other

Metadata will be licenced under the Creative Commons BY 4.0 licence.

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

Metadata include the DOIs and the handles of the described datasets.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Metadata will be stored both in the LifeWatch Italy Data Portal and in the Metadata catalogues of LifeWatch Italy and LifeWatch ERIC. The lifetime of the LifeWatch ERIC infrastructure is virtually forever and, therefore, long-term preservation of metadata stored within the LifeWatch catalogues is virtually forever.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

Couldn't find it? Insert it manually

DNA-derived occurrence data will be published with controlled vocabularies from the [Darwin Core List of Terms](#) to describe Occurrence and Event information.

The Darwin Core Extension of GBIF for [DNA derived data](#) will be used to capture information related to DNA sequences and analysis.

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

Couldn't find it? Insert it manually

The [EML](#) (Ecological Metadata Language) and the [Darwin Core](#) are the standards used to describe respectively metadata and data of the described template.

3.3.5 What is the methodology followed?

The approach for (meta)data standardisation used by the scientific community working within LifeWatch Italy and also adopted for the research data described here includes the use of a metadata schema based on the EML standard version 2.2.0 and the use of Darwin Core vocabularies and other controlled vocabularies for data standardisation.

3.3.6 What community-endorsed interoperability best practices are followed?

The interoperability approach used within LifeWatch Italy is similar to/compatible with those used by international communities of ecologists and established within other global data repositories such as the Global Biodiversity Information Facility ([GBIF](#)). Specifically, the EML is a metadata specification developed for the ecology discipline and based on prior work done by the Ecological Society of America and associated efforts. The Darwin Core is a body of standards, including a glossary of terms, intended to facilitate the sharing of information about biological diversity and it is sponsored by the community of the Biodiversity Information Standards (TWDG).

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

The full legal code of the CC BY 4.0 licence is available at:
<https://creativecommons.org/licenses/by/4.0/legalcode.en>

3.4.2 What reusability and / or reproducibility methods are followed?

- Data cleaning
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

Raw sequencing data will be stored in the SRA of the NHI Archive and DNA-derived occurrence data in the LifeWatch Italy Data Portal which will ensure the long-term accessibility to both raw and derived data.

3.4.5 Is provenance well documented?

Yes

Metadata will be enriched with provenance information. Specifically, metadata will include a detailed description of the methods, instruments and protocols used for data collection, processing and analysis.

3.4.6 What documented procedures for quality assurance do you have in place?

- Use of tools for automatic checks
- Data conform to format specification

The LifeWatch Italy Data Portal include quality assurance procedures consisting in structural and semantic checks. Structural checks include the control of file formats and file names. Semantic checks include the use of community-accepted controlled vocabularies for data attributes and data values. The Data Portal supports also quality assurance steps through taxonomic and syntactic controls. In addition, research data are also reviewed and verified by data

managers and domain experts who perform final quality control before data publication.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

- Storage
- Archiving

Direct cost

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Paola Pollegioni (orcid:0000-0001-6388-1931)

Role(s): Dataset creator; Metadata provider; Contact person

b. Ilaria Rosati (orcid:0000-0003-3422-7230)

Couldn't find it? Insert it manually

Role(s): Data manager

c. Cristina Di Muri (orcid:0000-0003-4072-0662)

Couldn't find it? Insert it manually

Role(s): Data manager

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Remote sensing analysis

This DMP template describes the (meta)data schemas used to harmonise and publish novel data and a digital service (Alvar) used to characterise the distribution of *Ailanthus altissima*. Occurrence data are used, within Alvar, to train the machine learning algorithms in the identification of the invasive species. Occurrence data (training sample set) are organised in a shapefile available for sharing, long-term access and reuse through the LifeWatch Italy Data Portal and the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC. The digital service Alvar, available in Google Earth Engine, will be also shared through the LifeWatch metadata catalogues.

Template: [Horizon Europe](#)

Type: [Dataset](#)

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Software

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

Alvar is a Google Earth Engine cloud application used for early detection of *Ailanthus altissima*.

Alvar is based on Copernicus Sentinel-2 optical imagery, available since mid-2015, at 10m ground pixel resolution.

The rationale of the application is based on the detection of plant species signals or time-series signatures that can be detected and interpreted. Alvar supports the search and identification of *A. altissima* optical signature by exploiting the cloud computing capability of fast time-series search of the images and the vegetation-related associated optical indices.

Alvar exploits the machine learning algorithms exposed by the Google Earth Engine APIs, training them by

means of a user-provided training/control set of empirical observations. The sample set, Aalto, is provided as a training dataset for Alvar distribution.

1.1.4 What is the type of the described dataset?

Observational

1.1.5 What is its format?

Observational data (Aalto) used as training data for the Alvar application are available as shapefile.

1.1.6 What is its expected size?

The Aalto shapefile is 4 KB.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To keep on record
- To make informed decisions
- To develop a product

The Aalto sample dataset is used within the Alvar software as a training dataset to identify the optical signal of the invasive tree *A. altissima* in invaded areas by using Copernicus Sentinel-2 data. The Aalto dataset includes unpublished information on *A. altissima* distribution. The use of the Aalto dataset with the Alvar application is an efficient tool for early detection of the invasive plant in novel invaded areas and for monitoring the spread of this species.

1.1.8 What is its origin / provenance?

The data and digital service described in this DMP template were generated within the project USEit to monitor the distribution and the spread of the invasive tree *A. altissima* in terrestrial ecosystems.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Decision makers

The Alvar service can be used by researchers studying invasive alien species in terrestrial ecosystems. The service uses available satellite data to monitor the distribution and the spread of *A. altissima* but the algorithm can be adapted and/or extended to monitor the distribution and invasion areas of other invasive plants and to study the impact of invasive species on natural terrestrial ecosystems.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Data identifiers
- Researchers identifiers
- Organizations identifiers

DOI

ORCID

ROR

The Aalto dataset will be uploaded within a dedicated space on the LifeWatch Italy Data Portal and metadata will be harvested by and exposed through the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC. A

handle is assigned to the dataset within the Data Portal and a DOI to the metadata.

The Alvar service, available in Google Earth Engine, will have its metadata published through the Metadata Catalogues of LifeWatch Italy and LifeWatch ERIC, and a DOI will be assigned to it.

All metadata will include PIDs of researchers (ORCID) and organisations (ROR) involved in the creation and publication of the research products described in this template.

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

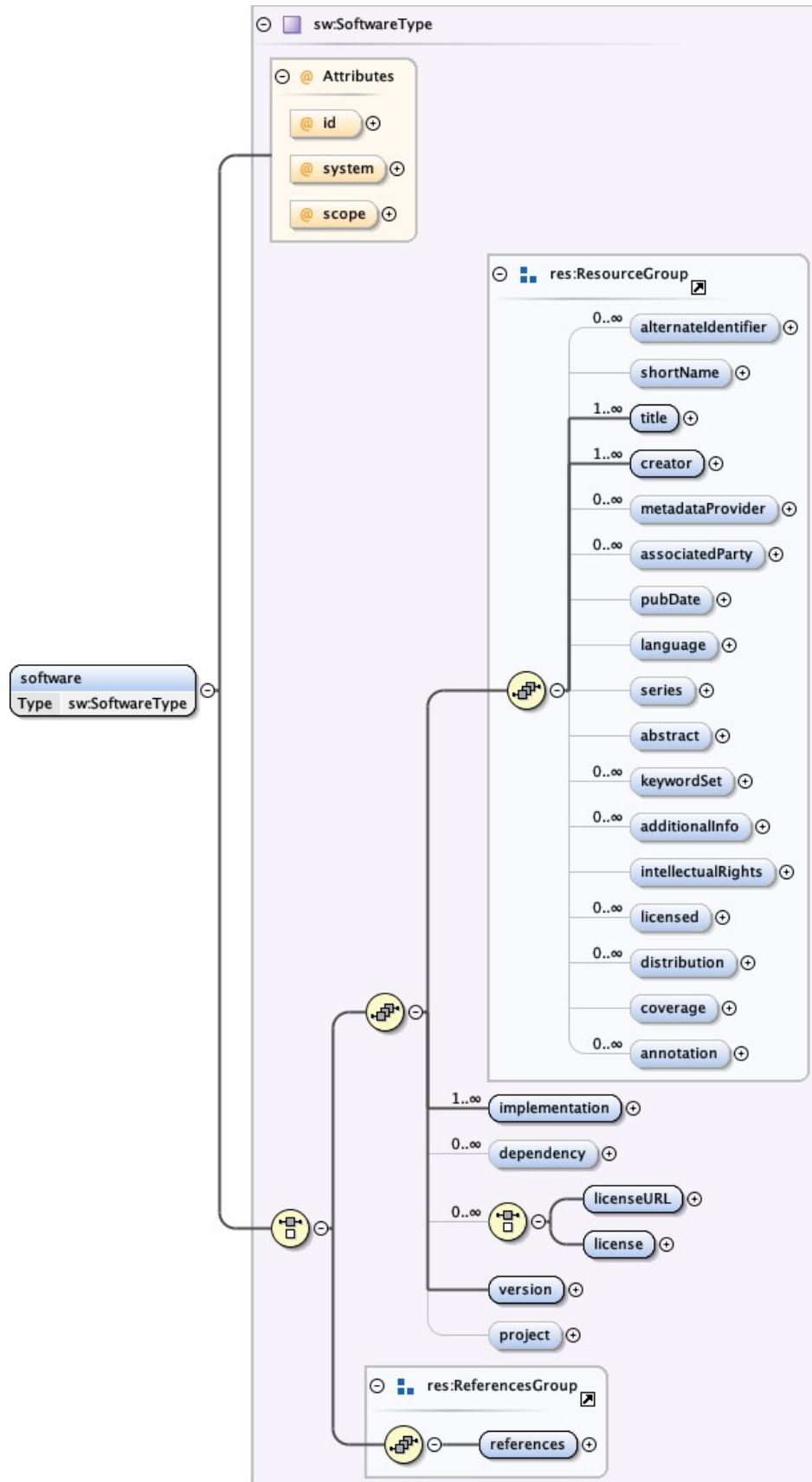


Image 5

The Alvar service will be described using the EML 2.2.0 metadata schema for software description (Image 1).

The Aalto shapefile will be described using the standard ISO 19139.

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural
- Reference
- Legal

The *eml-software* module contains general information that describes software resources. This module is intended to fully document software that is needed in order to view a resource (such as a dataset) or to process a dataset. The software module is also imported into the *eml-methods* module in order to document what software was used to process or perform quality control procedures on a dataset.

The *eml-software* module, like other modules, may be "referenced" via the <references> tag. This allows a software resource to be described once, and then used as a reference in other locations within the EML document via its ID.

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

- Registry/Catalogue
- Metadata repository
- Linked Open Data
- Content Management System

The metadata will be available on the LifeWatch ERIC metadata catalogue and in the LifeWatch Italy metadata catalogue that will be soon released. Moreover, the metadata will be searchable through the LifeWatch Italy Semantic Platform by using semantically enriched queries.

3.1.1.8 Are keywords provided in the metadata?

Yes

invasion biology, Terrestrial, monitoring, satellite, Copernicus, Sentinel-2

3.1.1.9 Are metadata harvestable?

Yes

The standard communication protocol for information exchange within the LifeWatch ERIC Metadata Catalogue is the Hypertext Transfer Protocol Secure (HTTPS) and metadata inside the catalogue are harvestable through the REST API.

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

LifeWatch Italy Data Portal

The Aalto shapefile will be available in the USEit sub-portal of the new LifeWatch Italy Data Portal that is going to be released shortly and the URL will be added here.

The Alvar application is hosted within the Google Earth Engine.

The LifeWatch Italy Data Portal is an online, open-access platform created and managed by LifeWatch Italy, the national node of the LifeWatch ERIC Research Infrastructure for Biodiversity and Ecosystem Research.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use
- Has an open access content policy
- Supports retention
- Supports withdrawal
- Supports back up
- Provides Open Access content (free at the point of use)
- Assigns PIDs
- Follows metadata standards
- Uses non-proprietary formats
- Supports mid- and long-term preservation
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

The research data and service described within this DMP template are generated within the framework of the USEit project. The Research Infrastructure of LifeWatch Italy, involved in USEit, will support the project research activities by providing the ICT infrastructure, tools, services and dedicated personnel for the publication of all research data and data products on IAS produced within the project. Moreover, a dedicated space within the LifeWatch Italy Data Portal will be created to host the USEit research data.

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

DOIs are assigned to the research datasets published within the LifeWatch Italy Data Portal into the dedicated space created for the USEit project.

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Aalto is the title of the described shapefile (sample dataset) and Alvar is the title of the application (service).

3.2.2.2 How is the dataset / output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Datasets will be preserved and stored in the mid- and long-term within a dedicated space created in the LifeWatch Italy Data Portal. Accessibility is guaranteed through a dedicated, user-friendly web interface that is maintained by LifeWatch Italy.

Users can freely access and download datasets and no authentication/authorisation is required.

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

LifeWatch Italy is the national node of the LifeWatch ERIC Research Infrastructure and, as such, the lifetime of the infrastructure supporting the USEit research data is virtually forever.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Other

Metadata will be licenced under the Creative Commons BY 4.0 licence.

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

Metadata includes the DOIs and the handles of the described data.

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes

Metadata will be stored both in the LifeWatch Italy Data Portal and in the Metadata catalogues of LifeWatch Italy and LifeWatch ERIC. The lifetime of the LifeWatch ERIC infrastructure is virtually forever and, therefore, long-term preservation of metadata stored within the LifeWatch catalogues is virtually forever.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

Couldn't find it? Insert it manually

The EML (Ecological Metadata Language) and the ISO 19139 are the standards used to describe the (meta)data of this DMP template.

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

The Aalto data (training dataset) is used within the Alvar application. The metadata of the two research products will include references and descriptions of the two products with details on how they can be used together.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

The full legal code of the CC BY 4.0 licence is available at: <https://creativecommons.org/licenses/by/4.0/legalcode.en>

3.4.2 What reusability and / or reproducibility methods are followed?

Readme files

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

Data will be stored in the LifeWatch Italy Data Portal which will ensure the long-term accessibility to the data (check section 3.2.1).

3.4.5 Is provenance well documented?

Yes

Metadata will be enriched with provenance information. Specifically, metadata will include a detailed description of the methods, instruments and protocols used for data collection and references to the projects/activities/monitoring programmes in which data were gathered will be added.

3.4.6 What documented procedures for quality assurance do you have in place?

- Use of tools for automatic checks

- Data conform to format specification

4.1 Allocation of resources

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. MARCO CIOLFI (orcid:0000-0003-4831-8053)

Role(s): Data creator; Service creator; Metadata provider; Contact person.

Research product(s) Title(s): Aalto data and Alvar (*Ailanthus altissima* visual aid for reconnaissance) application.

b. Francesca Chiocchini (orcid:0000-0002-5122-8756)

Couldn't find it? Insert it manually

Role(s): Data contributor; Service contributor.

Research product(s) Title(s): Aalto data and Alvar (*Ailanthus altissima* visual aid for reconnaissance) application.

c. Paola Pollegioni (orcid:0000-0001-6388-1931)

Couldn't find it? Insert it manually

Role(s): Data contributor; Service contributor.

Research product(s) Title(s): Aalto data and Alvar (*Ailanthus altissima* visual aid for reconnaissance) application.

d. MARCO LAUTERI (orcid:0000-0003-1071-7999)

Couldn't find it? Insert it manually

Role(s): Data contributor; Service contributor.

Research product(s) Title(s): Aalto data and Alvar (*Ailanthus altissima* visual aid for reconnaissance) application.

e. Cristina Di Muri (orcid:0000-0003-4072-0662)

Couldn't find it? Insert it manually

Role(s): Data manager.

Research product(s) Title(s): Aalto data and Alvar (Ailanthus altissima visual aid for reconnaissance) application.

f. Ilaria Rosati (orcid:0000-0003-3422-7230)

Couldn't find it? Insert it manually

Role(s): Data manager.

Research product(s) Title(s): Aalto data and Alvar (Ailanthus altissima visual aid for reconnaissance) application.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

no

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by

