GuardIAS

GUARDING EUROPEAN WATERS
FROM INVASIVE ALIEN SPECIES

Data Management Plan

Version 1 (M6, D7.1)

Planned delivery date: 30/06/2025 (M6)

Actual submission date: 28/06/2025

Deliverable leader: 10 – VLIZ

Authors: Leen Vandepitte (VLIZ), Stefanie Dekeyzer (VLIZ), Lynn Delgat (VLIZ) Jelle Rondelez (VLIZ)

Internal reviewers: Laura Verbrugge (INVASIVESNET), Laura Garzoli (INVASIVESNET), Jennie Pistevos (UAegean), Stelios Katsanevakis (UAegean)

Type of deliverable: Report

Work package Leader: WP7 – UAEGEAN

Version: 1

GA No. 101181413

Start date of the project: January 1st, 2025

www.guardias.eu

Funded by the European Union

GuardIAS receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101181413). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the EU nor REA can be held responsible for them.

DISCLAIMER

This document contains information on GuardIAS core activities. Any reference to content in this document should clearly indicate the authors, source, organisation, and publication date. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

COPYRIGHT NOTICE



This work by Parties of the GuardIAS Consortium is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

'GuardIAS receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101181413). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the EU nor REA can be held responsible for them.'

TABLE OF CONTENTS

EXECUTIVE SUMMARY 4

TERMS, ACRONYMS & ABBREVIATIONS 5

1 INTRODUCTION 6

1.1 DATA MANAGEMENT SCOPE 6

1.2 PROJECT PARTNERS 7

1.3 GUARDIAS PROJECT SUMMARY 8

2 DATA SUMMARY 9

2.1 PURPOSE OF DATA COLLECTION/GENERATION 9

2.2 TYPES AND FORMATS OF (META)DATA 10

2.3 DATA UTILITY 13

3 GUARDIAS GENERAL DATA ORGANIZATION 13

4 GUIDELINES FOR DATA PROVIDERS 14

5 OPEN AND FAIR DATA 14

5.1 APPLYING OPEN SCIENCE PRINCIPLES 14

5.2 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA 15

5.3 MAKING DATA OPENLY ACCESSIBLE 17

5.4 MAKING DATA INTEROPERABLE 17

5.5 INCREASE DATA RE-USE 18

6 OTHER RESEARCH OUTPUTS 18

7 ALLOCATION OF RESOURCES 19

8 DATA SECURITY 19

9 ETHICS 20

10 RESPONSIBILITIES 21

Executive Summary

This document is the first version (Month 6) of the Data Management Plan (DMP) for the GuardIAS project (*Guarding European Waters from Invasive Alien Species*), funded under the Horizon Europe Programme (Grant Agreement No. 101181413). It outlines the foundational principles, procedures, and infrastructure for managing the diverse datasets generated and used throughout the project. It is to be seen as a live document, which will be finetuned in future versions to reflect updates in overall project data management as well as specifics on the management of certain data types as the project progresses.

This GuardIAS Data Management Plan describes the key actions and strategies to optimize handling of data from discovery to publication, ensuring that data become and remain findable, accessible, interoperable and reusable (FAIR). It supports the project's commitment to Open Science by detailing how data will be shared responsibly, under appropriate licenses, and made available through trusted repositories to maximize reuse and long-term impact.

GuardIAS aspires to produce Five Star Open Data. The project will generate a wide array of high-value datasets, including species occurrence records, environmental and socio-economic data, molecular and acoustic libraries, and model outputs. These will be derived from both newly collected field data and harmonized existing sources such as EASIN, AquaNIS, WoRMS, GBIF, OBIS, EMODnet, Copernicus, and citizen science platforms like iNaturalist and Zooniverse. The DMP outlines the types and formats of these datasets and the metadata standards (e.g. DarwinCore, INSPIRE-compliant) that will ensure semantic and syntactic interoperability across systems.

To safeguard and disseminate these data, the Marine Data Archive (MDA) hosted by VLIZ will serve as the primary repository. Where necessary, domain-specific repositories (e.g. Zenodo, GBIF, OBIS, EMODnet) will complement MDA for archiving raw data, large file formats, or specialized outputs. All datasets will receive persistent identifiers (e.g. DOIs), and raw and derived datasets will be cross-linked to maintain data lineage and ensure reproducibility.

The plan details the procedures for quality assurance, access control, data security, and ethical considerations, including compliance with GDPR for personal and sensitive data. Robust guidelines for data providers, quality control measures, and long-term preservation strategies will ensure the integrity, usability, and impact of the data generated, thus supporting the broader scientific community, policy-making bodies, and other stakeholders engaged in aquatic biosecurity and biodiversity conservation.

Terms, Acronyms & Abbreviations

Table 1. Most commonly used terms, abbreviations and acronyms (alphabetically) used within this document

Acronym / Abbreviation	Signification
Data	Data encompasses all forms of information that have been collected, generated, or acquired—whether through purchase or other means. It may originate from diverse sources including direct observations, interviews, communications, or the use of recording technologies such as cameras and sensors. Data creation also includes outputs from programming, simulations, gaming environments, modelling processes, as well as authored content such as reports, publications, blogs, and websites. Acquired data refers to any information obtained from third parties, whether paid for or received at no cost; for the purposes of this report, all such acquisitions are treated as data purchases.
DMP	Data Management Plan
DwC	DarwinCore
EASIN	European Alien Species Information Network
FAIR principles	Make the data Findable, Accessible, Interoperable and Reuseable
GBIF	Global Biodiversity Information Facility
GDPR	General Data Protection Regulation
Guardias	Guarding European Waters from Invasive Alien Species
GuardIAS consortium, GuardIAS partners, partners	In the context of this text, the terms are used interchangeably to collectively denote all entities participating in the GuardIAS project, including both core and associated partners
IAS	Introduced Alien Species
IMIS	Integrated Marine Information System
INVASIVESNET	International Association for Open Knowledge on Invasive Alien Species
MDA	Marine Data Archive
OBIS	Ocea Biodiversity Information System
Personal data	Any information that relates to an identified or identifiable living individual (data subject). Different pieces of information, which together can lead to the identification of a particular person, may also be considered personal data. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the General Data Protection Regulation (GDPR), the EU's main data protection law. [source: https://commission.europa.eu/law/law-topic/data-protection/data-protection-explained_en]
WoRMS	World Register of Marine Species

1 Introduction

1.1 Data management scope

This Data Management Plan (DMP) of GuardIAS describes the key actions and strategies to optimize handling of data from discovery to publication, ensuring that data become and remain findable, accessible, interoperable and reusable (FAIR). This approach will maximize data discovery and exploitation by the wider community.

Data management will be performed in adherence to international accepted principles, practices and standards, defined by ICSU World Data System, Open Geospatial Consortium (OGC), International Organization for Standardization (ISO), Global Ocean Observing System (GOOS), International Oceanographic Data and Information Exchange (IODE), and Intergovernmental Oceanographic Commission of UNESCO (IOC-UNESCO).

The project ensures the compliance with platform-, data type-, and parameter-specific standards, formats and recommendations, including but not limited to those of Integrated Marine Information System (IMIS), Marine Data Archive (MDA), World Register of Marine Species (WoRMS), Global Biodiversity Information Facility (GBIF), European Marine Observation and Data Network (EMODnet), Biodiversity Information Standards (TDWG) (Darwin Core Standard (DwC)), Ocean Biodiversity Information System (OBIS), Marine Regions (MRGID) and BODC NERC Vocabulary Server. The harmonisation of metadata formats (syntactic interoperability), temporal and spatial resolution and thematic differences (semantic interoperability) will be carried out according to agreed templates including those of the INSPIRE Directive as laid down in e.g. EMODnet and ENVRI-FAIR.

GuardIAS aspires to produce Five Star Open Data (*Figure 1*). To reach the maximum five stars, data is required to:

(1) be available on the Web under an open license (OL),

(2) be available as structured data (RE),

(3) be available in a non-proprietary open format (OF),

(4) use Uniform Resource Identifiers (URIs), and

(5) link to other data (LD).

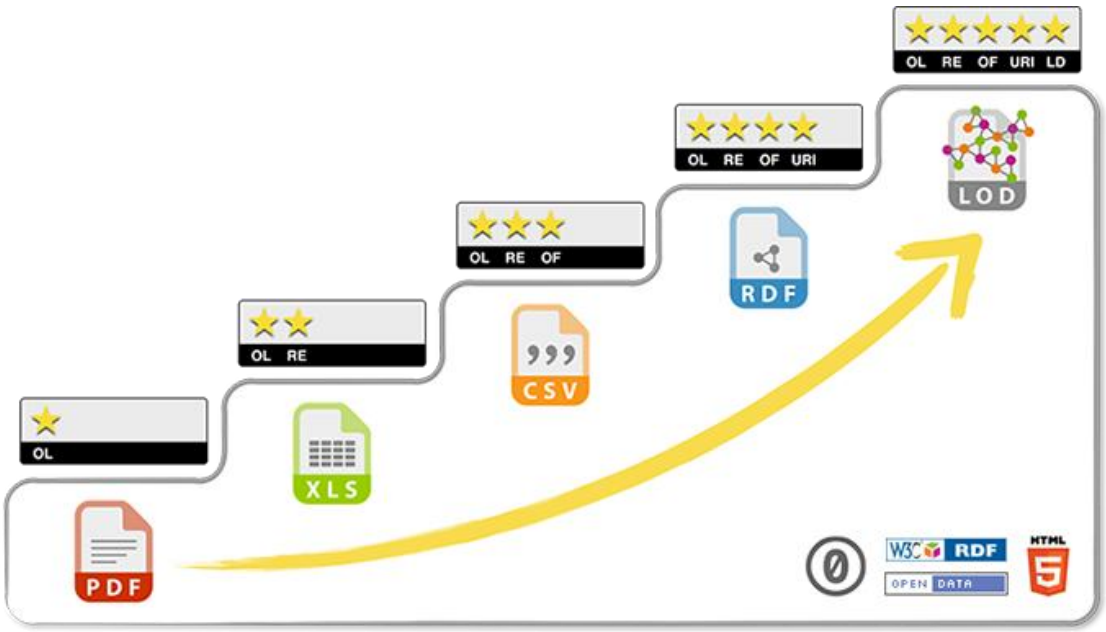


Figure 1: Five-star open data scheme

1.2 Project partners

Beneficiaries

1. Panepistimio Aigaiou (UAEGEAN), Greece
2. Leibniz Institute of Freshwater Ecology and Inland Fisheries in the FORSCHUNGSVERBUND BERLIN E.V. (FVB- FVB-IGB), Germany
3. Università Degli Studi Di Pavia (UNIPV), Italy
4. International Association for Open Knowledge on Invasive Alien Species (INVASIVESNET), Finland
5. Sequench Limited (Sequench), New Zealand
6. Klaipedos Universitetas (KU-LT), Lithuania
7. Aarhus Universitet (AU), Denmark
8. Eigen Vermogen van het Instituut voor Natuur-en Bosonderzoek (EV INBO), Belgium
9. Nord Universitet (NORD), Norway
10. Vlaams Instituut voor de Zee (VLIZ), Belgium
11. Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC), Spain
12. Università Degli Stude di Firenze (UNIFI), Italy
13. Lobster Innovations BV (Lobster), Netherlands
14. International Estuarine & Coastal Specialists (IECS), UK
15. ERINN Innovation Limited (ERINN), Ireland
16. Irish Ocean Literacy Network (IOLN), Ireland
17. Nelson Artificial Intelligence Institute Limited (NAI), New Zealand
18. NNIT SRL (NNIT), Belgium
19. Joint Research Centre – European Commission (JRC), Belgium

Associated partners

20. University of Sydney, Australia

1.3 GuardIAS project summary

GuardIAS builds on the best practices in global biosecurity and combines cutting-edge science, new technologies, innovative techniques, and information systems within a transdisciplinary framework to provide effective, technologically feasible, and socially acceptable solutions. We will apply Artificial Intelligence (AI) and new data workflows to systematically query multiple biodiversity databases for species distribution, environmental tolerances, biological traits, and genetic information, enhancing EASIN and harmonizing key datasets. This will improve IAS impact profiling and effective invasion prevention and management. The world-leading citizen science (CS) platform iNaturalist (using AI to identify species from smartphone photos) will be enhanced with expert-verified images of aquatic IAS to aid early detection at borders and track the geographic spread of IAS (cfr. WP1).

Among the most important pathways for aquatic IAS introduction and further spread are the discharge of ballast water, hull biofouling, and the movement of recreational boats. While the introduction of IAS through ballast water is expected to be effectively tackled by the IMO Ballast Water Management Convention (BWMC), biofouling remains a primary global pathway for IAS introductions, since it is regulated only by voluntary IMO guidelines. For this reason, GuardIAS focuses on innovative prevention of new introductions through hull biofouling by exploring emerging nanotechnology-based antifouling coatings. GuardIAS will also investigate the movement patterns of recreational boats along European coastlines, a significant but understudied factor in IAS secondary dispersal. An eDNA reference library and assay panel will be developed to aid effective IAS detection (cfr. WP2).

Macroecological and biogeographic trait analysis and cutting-edge models (e.g. the Nobel-Prize-winning MRIO - Multi-Region Input-Output analysis) will assess the risks of IAS introduction and spread, including risks to threatened species, critical habitats, and areas of conservation importance at present and under future invasion scenarios. Systematic conservation planning tools will be used for the first time to prioritize IAS monitoring and management actions based on their impact on Red List species and under invasion scenarios (cfr. WP3). GuardIAS will boost data collection, monitoring, early detection, and IAS literacy through innovative Citizen Science initiatives, including BioArtBlitz events (a novel type of event where figurative and performative arts will be used as a vehicle for communication and engagement), eDNA sampling, sound analysis in a Zooniverse project, and marina events for boaters. We will further engage stakeholders (policy makers, industry, scientists, aquatic managers and control officers, the public, youth, and educational institutions) through innovative tools, such as applied games (cfr. WP5 & WP6).

GuardIAS will co-design with environmental authorities, the industry, and aquatic managers, and implement numerous eradication and control efforts in both the marine and freshwater environments (cfr. WP4). By integrating Social Sciences and Humanities (SSH), GuardIAS will advance collaborative knowledge creation, facilitate comprehension of public perceptions on the significance of managing IAS, and facilitate exploitation of the project's outcomes. Particularly in

marine settings, the project aims to dispel the notion that post-establishment IAS management is futile, demonstrating tangible solutions for the implementation of Articles 17 and 19 of the IAS Regulation (cfr. WP6).

2 Data summary

2.1 Purpose of data collection/generation

GuardIAS will leverage existing data from

- EASIN (European IAS lists and georeferenced records, pathways, references)
- AquaNIS (European aquatic IAS lists, introduction events, traits, native origin, impacts)
- World Register of Introduced Marine Species (WRiMS) - global marine IAS lists, distributions, traits, links, images
- World Register of Marine Species (WoRMS) - nomenclature and classification
- GISD (IUCN's global IAS database, distribution, impacts, and management options)
- EMODnet (biology, seabed habitats, bathymetry)
- Copernicus - marine environmental data, climate change
- IUCN Red List - red-listed species, their distribution, population status, and threats
- Eionet's Reportnet infrastructure (data reported by EU MS under EU legislation, including the IAS Regulation and MSFD)
- World Database on Protected Areas (WDPA)
- Global Database on Protected Area Management Effectiveness (GD-PAME)
- Global Biodiversity Information System (GBIF)
- Ocean Biodiversity Information system (OBIS)
- FirstRecords (first occurrence of IAS per region)
- CABI-ISC (invasive species compendium)
- NCBI (molecular databases including access to GenBank)
- Global Biotic Interactions (GloBi)
- Global Library of Underwater Biological Sounds (GLUBS)
- CitizenScience platforms, including iNaturalist and Zooniverse
- previous projects
- national or institutional data centres and portals

These datasets will inform empirical analyses, e.g. updating the pan-European IAS inventory (T1.1), analysing traits (T1.2), assessing pathways (T1.3), modelling (T1.2, T3.2), curating libraries of sound emitted by aquatic IAS (T2.3) and of genetic markers and sequences (T2.4), training AI (T2.5), generating interaction networks between natives and IAS (T3.1), modelling the potential expansion of invaders under current and future scenarios (T3.2), impact and risk assessments (T3.3) and prioritizing monitoring and eradication (T3.4).

GuardIAS is prepared for the secure storage of more than 5 TB and this can be expanded if needed. Estimation of expected data sizes is ongoing.

2.2 Types and formats of (meta)data

GuardIAS will generate new datasets by harmonising and integrating data from multiple sources leading to outputs, such as:

- modelled predictions and IAS distribution maps under both current and future conditions
- a pan-European inventory of aquatic alien species and pathway assessment, sea bottom photomosaics
- GIS maps
- sound, and molecular libraries
- data mining algorithms
- potential IAS lists (AI powered Horizon scanning)
- impact and risk assessments
- questionnaires
- a wealth of reports, publications, images, videos, and communication outputs

The table below shows an exemplary overview of the different data types that could be provided in the GuardIAS project (either newly collected, or from institutional background datasets), and the different data sources that are used to collect these data. The contents of this table are subject to change during the project, and this table will be updated accordingly in future versions of the DMP.

Table 2. Exemplary overview of different domains with their specific data sources and data types

Domain	Data sources	Data types
Socio-economics	• *Interviews (spoken)* • *Surveys (written)* • *Questionnaires (written)* • *Literature study*	• *Views/opinions* • *Local Ecological Knowledge (LEK)* • *Economic data*
Biology	• *Sediment sampling* • *Observation (visual survey)* • *Molecular methods: metabarcoding / eDNA* • *Imaging (e.g. Zoo/Ecoscan)* • *Laboratory experiments* • *Drone/ satellite* • *Net tows* • *Bottom trawls* • *Citizen science* • *Molecular methods: sequencing voucher specimens* • *Acoustic survey* • *Molecular methods: telomer length*	• *Abundances* • *Occurrences* • *Traits / biotic measurements / species interactions* • *Images* • *Sequences*

	• *Ad-hoc biological traits data* • *Fishery landing data*	
Physics	• *CTD casts* • *Sediment samples* • *Seawater samples* • *Moored observatories* • *Ferrybox transects* • *Drifting buoys* • *Gliders*	• *Temperature* • *Salinity* • *Water turbidity (light attenuation)* • *Wave height and period* • *River flow* • *Underwater noise* • *Current profiles* • *Sea-ice coverage* • *Wind speed and direction*
(Bio)chemistry	• *Sediment samples* • *Seawater samples* • *Ferrybox transects* • *CTD casts* • *Gliders*	• *Organic matter ("blue carbon", sediment C/N content, etc.)* • *Chlorophyll* • *Nutrients/fertilisers* • *Heavy metals* • *Dissolved gasses* • *Marine litter* • *Acidity* • *Silicates* • *Hydrocarbons*
Data products		• *Model outputs* • *Model code* • *Statistic matrixes* • *Mind model maps* • *Graphs* • *Compiled datasets* • *Biological valuation maps* • *Compilation of rules and regulations* • *Standards document on non-economic valuation of biodiversity*

Table 3. Recommended file formats for different data types

Domain	Data types	Recommended file formats
Socio-economics	• Tabular data (tables, matrices)	• CSV / TSV
	• Networks (Fuzzy Cognitive Maps)	• NET
	• Text	• TXT • ODT • XML / HTML
Biology	• Tabular data (occurrences, abundances, traits / biotic measurements / species interactions)	• CSV / TSV
	• Images	• TIFF • JPG / PNG
	• Sequences	• FASTQ, FASTA / FAS, TXT, CSV
	• Raw acoustics data	• WAV
Physics		• CSV / TSV • NetCDF/CF
(Bio)chemistry		• CSV / TSV • NetCDF/CF
Data products	• Code	• Plain text (usually with an extension that represents the source language)
	• Geospatial data (model outputs, maps,…)	• NetCDF/CF • GeoTIFF • ESRI Shapefile (.shp, .shx, .dbf)
	• Others	Strongly depends on type

2.3 Data utility

Besides the project partner consortium, the target groups of GuardIAS are:

- Key policy stakeholders: National and statutory authorities responsible for implementing various directives and regulations (IAS Reg., MSFD, BHD, WFD, Biodiversity Strategy, Restoration Law); RSCs; IMO; European bodies (EEA, DG-ENV, DG-MARE); environmental NGOs; policy advisors.
- Industry: Industries related to the introduction and spread of IAS: shipping, aquaculture, fisheries, pet and aquarium trade.
- Scientific Community: Scientists and researchers studying IAS and their impacts on biodiversity, ecosystem services, and human health, and early career researchers.
- Aquatic Managers and Control Officers: Environmental managers, protected areas management bodies, aquatic monitoring and control agencies, consultants, and border control officers.
- General public (including citizen science): Society at large, particularly users of aquatic ecosystems.
- Youth & Educational Institutions: This includes – but is not limited to – universities and higher education trainings

3 GuardIAS general data organization

During initial data generation and quality control, the data is stored at the data provider's organisation. It is the data provider's responsibility to ensure that the data is stored securely at this stage. A data provider can be either a project partner, or someone from outside of the project consortium. Raw data should always be saved, formatted and quality controlled by the (project) partner working with the data. All published data should be stored at the Marine Data Archive (MDA). In the unlikely case the MDA is not suitable for the generated data, then the VLIZ data centre will assist in the search for a suitable archive.

As soon as possible, i.e. after initial quality control/pre-processing, the data provider will upload the data to the Marine Data Archive (MDA; https://marinedataarchive.org/), which is hosted by the Data Centre of the Flanders Marine Institute (VLIZ), where the data will be archived. A full overview of the functionalities and possibilities of the MDA is available at https://marinedataarchive.org/mdamanual.pdf. In the MDA, each dataset receives a unique identifier, to which a direct link can be provided. There are several data-sharing levels available on the MDA: (1) the data can initially be kept private; (2) the data can be shared between (subdivisions of) the GuardIAS partners and (3) the data can be made publicly available. Each dataset will have its specific, detailed metadata linked to it. When a publication about the data is published, or else at the latest 1 year after the project, datasets will be made openly accessible on the MDA and receive a DOI. Metadata may also be published in additional specialized catalogues depending on the data type.

The main purpose of the MDA is long-term preservation, and to increase findability, interoperability and re-use, data will also be published in different international repositories, such

as EMODnet, OBIS, and GBIF, depending on the data type. While most of the project data will go to the MDA, some data types may not be suitable for the MDA, specifically raw data which take up large amounts of storage space such as raw sequencing data, images, etc. These will be archived in other, data type-specific, repositories. Depending on the kind of data and their size, the VLIZ Data Centre will assist in searching for a suitable repository. Before publication, these data types can be shared upon request between partners or be shared between specified collaborators if the repository allows for that. Still, the data derived from these (e.g. occurrence data derived from images) will be archived in the MDA. For all datasets/-products generated in the project, an IMIS metadata record will be created, linking to the MDA and/or other repositories where the dataset/-product can be found. In cases where there is both a 'raw' dataset and a 'derived' dataset, these will be linked to each other.

The European Alien Species Information Network (EASIN) facilitates information on alien species occurring in Europe and supports the EC services and EU Member States competent authorities in the implementation of the EU Reg. 1143/2014 on Invasive Alien Species. As EASIN has the mandate to be the EU point for all alien invasive species data and information, ***all geospatial data and information about alien species in Europe from the GuardIAS project will need to become part of EASIN*** within the project duration, or at the latest 1 year after the project ends. From EASIN, it can then also flow to other data infrastructures, including e.g. the Global Biodiversity Information Facility (GBIF).

4 Guidelines for data providers

A document called "GuardIAS guidelines for data providers" will be compiled to help data providers ensure the data they collect in the scope of the project become available according to the FAIR principles, by providing a concise and practical overview of the needed steps for the data flow.

This document will become available by M12 and will be updated as needed. It will cover – amongst others - the following steps:

1. Submit metadata
2. Initial quality control/processing
3. Store data in trusted repository with restricted access
4. Data formatting / standardization
5. Make dataset publicly available from appropriate repositories
6. Alert data managers to link data and metadata

5 Open and FAIR data

5.1 Applying Open Science principles

GuardIAS is committed to an open science approach, ensuring the widespread dissemination and exploitation of project outcomes. The cornerstone of the GuardIAS strategy lies (1) in its

commitment to open-access (OA) publishing, also utilizing the European open Science Cloud (EOSC), (2) facilitating knowledge sharing beyond the scientific community, and (3) fostering collaborative engagement via an early-stage forum established through the network of INVASIVESNET.

The project emphasises publication of raw data by all partners. Exchange of data among partners has been formalised through the Consortium Agreement (Task 7.1) accompanied by this Data Management Plan (Task 7.3), both established at the project's outset. GuardIAS underscores the transparency and accessibility of data, ecological syntheses, and policy-level integration, aligned with the FAIR principles to ensure data's Findability, Accessibility, Interoperability, and Reusability. Data will – whenever possible – be stored in the MDA, from where they can find their way to other relevant databases, including – but not limited to e.g. EOSC, EASIN, AquaNIS, WRiMS, EMODNet, GBIF. Where possible, data will be under an open access, respecting the data limitations of e.g. personal data.

The commitment to open science extends to publication practices. Scientific articles will be published under OA licences, embracing the Green, Diamond, or Gold licensing models. Leveraging the benefits of OA, GuardIAS partners may also utilise the EC 'Open Research Europe' platform, allowing immediate, open peer-review, and cost-free OA publication. Partnering with INVASIVESNET, whose main objective is to enhance accessibility to open sources of IAS data and knowledge and many GuardIAS members being editors of renowned OA journals committed to open science, GuardIAS will actively contribute to open science practices.

Additionally, GuardIAS takes a comprehensive approach at an early stage to open sharing of knowledge and tools, encompassing multiple avenues: (1) co-creating GuardIAS methods and tools with the Advisory Board and stakeholders, a fundamental component of open science, will ensure public availability of outcomes; (2) a **Zenodo** repository will house papers, datasets, and software in open access; (3) a dedicated project in **Google Scholar** will compile references and project metrics; (4) institutional repositories are enlisted to host appropriate project materials; (5) a dedicated web page will offer a central hub for deliverables, software, and ongoing updates. Retaining the domain name and web page for >10 years after the end of the project signifies long-term commitment to knowledge sharing; (6) engaging citizens in many distinct ways (WP5), sharing knowledge and co-producing data and outputs.

5.2 Making data findable, including provisions for metadata

As many data types as possible will be stored in open formats (DarwinCore) with INSPIRE compliant metadata, ensuring enhanced findability and accessibility. Ontologies & standards include – but are not limited to – the World Register of Marine Species (WoRMS) for taxonomy, Marine Regions for geographical location names, relevant ISO & EML standards for data and metadata, BODC vocabulary.

Metadata, which will be based on standardized ontologies and vocabularies, will be made INSPIRE-compliant. This will facilitate harvesting of the metadata through the standard protocols of the several public repositories and databases that will be used to disseminate the project's output. The minimum metadata to provide are listed in the table below. Additional metadata that will be recorded is dependent on the specific data type.

Table 4. Metadata fields

Basic information	
Title in English	*Ideally the title should give an idea of the content of the dataset, the time period and the region of sampling*
Original title + language (if applicable)	*If the dataset is in another language, the title in the original language and the language.*
Citation	*The proposed dataset citation should reference the dataset, and not a paper using or describing the dataset.* *Format: Data Creator(s) Person; Data Creator(s) Institute (division); (Publication Year). Title. Identifier*
Identifier	*The globally unique identifier within a particular data management system. Typically, a DOI or another identifier that is both citable and resolvable is used.*
Abstract	*Short summary of the dataset and provides an indication of the content of the dataset, addressing the following questions: WHAT, WHERE, WHEN, HOW, WHY and WHO.*
Keywords	*Keywords or key phrases that concisely describe the dataset or are related to the dataset.*
Data theme	*General topic of the dataset (e.g. Biology, Physical, etc.).*
People involved	
Metadata provider	*Name of the person who supplies the metadata*
Contact person	*The person who should be contacted in cases of questions regarding the content of the dataset or any data restrictions. This is also the person who is most likely to stay involved in the dataset the longest.*
Contact organisation	*Organisation of the contact person*
Contact email	*Email of the contact person*
Data creator (multiple possible)	*All people and/or institute (divisions) who are responsible for the creation of the data(set).*
Data creator organization	*Organisation of data creator*

(multiple possible)	
License information	
Access constraints	*License name*
Geographic information (multiple possible)	
Geographical cover	*Most specific location to which the dataset refers.*
Temporal information (multiple possible)	
Temporal cover	*Start and end dates of data collection.*
Taxonomic information (multiple possible) (only for biological datasets)	
Taxonomic cover	*Names of the taxa that are included in the dataset.*

5.3 Making data openly accessible

Data from GuardIAS will be disseminated under a Creative Commons (CC) or Open Data Commons (ODC) license, depending on the data and/or data product. Data generated can be embargoed until analysis completion or manuscript publication, with all data accessible upon request and openly published within 12 months post-project conclusion.

The project will use established repositories and databases to ensure that the (meta)data remains available for as long as possible beyond the project's duration. Each of the public repositories and databases envisioned to be used within the project provides its own protocols and standardized access methodologies.

Data from GuardIAS will be disseminated widely through scientific publications (Supplementary material or data papers) and public repositories (EASIN, EMODnet, OBIS, GBIF, Pangaea, European Open Science Cloud, Zenodo, FigShare, …), and will be assigned unique and persistent identifiers (e.g., DOI via OpenAIRE+, VLIZ or ZENODO e-infrastructures). The first stop for alien species records and related data on traits, pathways, and impacts will be EASIN, followed by other Open Access databases (e.g., WRiMS, AquaNIS). Spatial data will be in GeoTIFF and geopackage formats, with databases in accessible formats like CSV.

Data will preferably be stored in non-proprietary file formats. These are file formats that are not controlled by a single vendor or company. They are often open standards, meaning their specifications are publicly available and can be used by anyone without restriction.

5.4 Making data interoperable

GuardIAS will adhere to the FAIR principles (Findable, Accessible, Interoperable, Reproducible). The project commits to best data management practices, ensuring data protection, sharing, and

accessibility. A comprehensive Data Management Plan (DMP) is herein delivered by month 6 and will be revised by month 20 (Deliverable 7.2) and month 35 (Deliverable 7.3), detailing strategies to maximise data usability, transparency, and reproducibility. By following this DMP, GuardIAS aims to enhance data integrity, accessibility, and collaboration, ultimately contributing to the advancement of IAS management and supporting informed decision-making in relevant sectors.

To foster data interoperability, Partners will use data mapping and transformation tools where data formats or schemas are different. Semantic interoperability will be ensured, which constitutes that the precise meaning of exchanged data is understandable and preserved across different systems of Project Partners. This is achieved by controlled vocabularies, metadata, and data dictionaries. These can include – but are not limited to - the World Register of Marine Species (WoRMS) for taxonomy, Marine Regions for geographical location names, relevant ISO & EML standards for data and metadata and BODC vocabulary. The Project Team establishes data governance policies and quality control mechanisms, crucial for ensuring the accuracy, consistency, and reliability of the data being exchanged.

Data generated in the project will adhere to standard formats and be as much as possible compliant with available (open) software applications, facilitating interoperability. All data providers are responsible for making sure their data adhere to international standards before publication to ensure interoperability. Which standards apply, depends on the data type. The project consortium implements common data models and schemas to interpret and use the data consistently. For web-based data exchange Project Partners use RESTful APIs where appropriate.

5.5 Increase data re-use

Licenses will be specified in the metadata records. Data will become open access as soon as research involving that data is published. Data not (yet) included in published research will become open access at the latest 12 months after the official end-date of the project. The data will be documented using rich metadata.

The data deposited in public repositories and databases will include clear references to README files, publications, software, vocabularies, and other additional metadata needed to facilitate re-use. Additionally, the GitHub channel of the GuardIAS project will store project scripts and output.

Initial quality control on data and metadata, as well as ensuring that the data is DarwinCore compatible, is the responsibility of the data provider. Additional quality control will be performed by the data manager before publishing. Quality of biological data will need to meet the EMODnet Biology Data Quality Criteria, which will be verified by performing checks with the LifeWatch & EMODnet Biology QC tool. Several tools to check and enhance the quality and completeness of the data will be shared with all project partners.

6 Other research outputs

GuardIAS conforms with Horizon Europe requirements to make all peer-reviewed scientific publications generated by GuardIAS open access immediately after publication. Publications will

(1) be published as Diamond, Gold, or Green open access and (2) be deposited in the GuardIAS Zenodo repository and be made open access.

The GuardIAS project will create a **GitHub** organization to store scripts and output. This developer platform allows developers to create, store, manage and share their code. This way, a virtual community for GuardIAS – consisting of project partners – can work together on software and keep track of software related issues. The GuardIAS GitHub channel is available at https://github.com/guardias-eu.

7 Allocation of resources

Task 7.3, led by VLIZ and foreseen with dedicated budget, focuses on developing, implementing, overseeing, and updating the data management plan. Each partner is responsible for their dataset(s), ensuring quality control and DarwinCore compatibility. Any time spent by partners on data management within the project, is covered by the project's budget allocation.

Budget allocation for OA papers ensures that the project's outcomes reach a broad audience. INVASIVESNET, through its allocated budget, will support publications in Regular or Special Issues in its journals (*Aquatic Invasions, Management of Biological Invasions, BioInvasions Records*) to facilitate broad open dissemination of GuardIAS outputs and broader engagement with the biological invasions community (Task 5.2). Partners will use institutional and national agreements with publishers to publish in Gold OA journals. In the absence of such agreements, recognising the high costs of many Gold OA journals, Green and Diamond OA options are prioritized, thereby supporting the best value for tax-payer funding.

There are no specific costs foreseen for the use of the envisioned public repositories and databases.

8 Data security

Non-public working datasets will be securely stored in the Marine Data Archive (MDA), an online repository specifically developed to independently archive data files in a fully documented manner, hosted by VLIZ. The MDA ensures data security through automatic backups and secure storage, supported by local institutional IT services. The MDA platform is hosted on virtual machines within the VLIZ virtualization cluster. This infrastructure allows flexibility if a system fails or would be overloaded. VLIZ foresees enough storage capacity for the processing and archival of the expected size and volume of data and documents. If VLIZ ceases to exist or terminates its data-archiving activities, VLIZ shall attempt to transfer the data files to a similar organisation that will continue the Agreement with the Data provider under similar conditions if possible. If VLIZ does not succeed in finding a similar organisation, VLIZ will organise a transfer of all files to each intellectual property right holder or data provider.

Sensitive and/or personal data subject to GDPR will be anonymized and deleted post-project. Only non-precise locations of threatened species will be disclosed. The MDA provides a secure

way to share data between the partners by making use of a shared folder. It also ensures long-term preservation of the data.

GuardIAS data will not only be stored in data archives but will also be replicated to other repositories (e.g. EASIN, OBIS, GBIF, EMODnet), as soon as they can become part of the public domain. This far horizon strategy will ensure long-term data accessibility, resilient to possible fault occurrences affecting one of the data systems where data are stored.

The project will use established repositories and databases to ensure that the (meta)data remains available for as long as possible beyond the project's duration. A dedicated web page will offer a central hub for deliverables, software, and ongoing updates. Retaining the domain name and web page for >10 years after the end of the project signifies long-term commitment to knowledge sharing.

9 Ethics

GuardIAS commits to data publishing with independent reproducibility and adherence to FAIR principles. The project will follow EU GDPR regulations in the EU countries and in the UK, which need to be adhered to by every project partner. Each organization is responsible for ensuring that their collection, processing and sharing of personal data and/or special categories of personal data are following Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (the General Data Protection Regulation (GDPR)) and other applicant regulations on personal data. GuardIAS partners will obtain consent for any personal data collected and enforce stringent privacy protections in accordance with EU law. If it is not possible for the Parties to share personal data in completely anonymized form, the Parties agree to use the EU provided templates for transfer of personal data and for processing of personal data cf. decision 2004/915/EC, decision 2001/497/EC and decision 2010/87/EU.

Data will be kept no longer than 12 months post-project unless individuals consent to continued inclusion. All GuardIAS websites will clearly communicate these policies through dedicated privacy statements.

GuardIAS primarily generates and analyses biological and environmental data. These data are neither sensitive/personal nor restricted data, however, the project intends to conduct sociological research as well. Interview and questionnaire data and transcriptions might contain names of informants and possibly also personal background and organizational or political affiliation. This information will not be made publicly accessible but will be treated with confidentiality within the project. Minimal data will be recorded (name, address, email and telephone number) for effective communication. They will be asked for written permission to store these data but all inputs from them into the discussions will be anonymised and confidential and they will not be able to be identified in any written reports.

In the publicly available analyses (e.g. reports and articles) GuardIAS will use one of the following two procedures:

- Anonymity as ‘default’ procedure for all interview data unless otherwise agreed (please see option 2 below). Anonymity will be secured in the (re)presentation of the interview persons’ perspectives, opinions etc. This is particularly relevant in interviews on the subject of fishery governance, which can be politically sensitive. When citing or re-

phrasing, there will not be reference to a specific name or title or location (by which the person could be easily identified), but rather a broad professional category like fisher, hunter, manager, industry representative. Furthermore, personal information that could easily be used to identify the interview person will be omitted or even slightly revised (e.g. details on family relations, previous career etc.) (if inconsequential for the analysis itself). If names are to be included for the sake of writing style, a pseudonym will be provided.

- The interviewees will be offered an opportunity to claim credit (full name and community) for the knowledge they offer to GuardIAS. This option will be presented in the information sheet and is to be ticked off as an additional item in the consent form.

National and EU legislation will be followed regarding sampling of species and sampling in protected areas. National and European data protection regulations will be followed. Any laboratory or field experiments will be carried out with full ethical approval of the host laboratory and all details will be recorded. The removal of any native species will be minimized where possible and as necessary.

10 Responsibilities

It is the data provider's responsibility to ensure that the data is stored securely at the first stage. A data provider can be either a project partner, or someone from outside of the project consortium.

Raw data should always be saved, formatted and quality controlled by the (project) partner working with the data. All published data should be stored at the Marine Data Archive (MDA). It is the responsibility of each project partner to upload their data onto the MDA. Instructions on how to do this will follow in M7-12. In the unlikely case the MDA is not suitable for the generated data, then the VLIZ data centre will assist in the search for a suitable archive.

Throughout the project, VLIZ – as project partner – will assist all partners who deal with newly generated data or existing data on how to properly manage these data, make sure they are FAIR, available to other project partners during the project and within the relevant data flows to become publicly available through e.g. EMODnet Biology, OBIS and GBIF during the project or – at the latest – one year after the project.

All data-related questions can always be address to data@vliz.be, with clear mention of 'GuardIAS project' in the subject line.